НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
УНИВЕРСИТЕТ ИТМО

ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

2024, ТОМ 24, НОМЕР 1 (ЯНВАРЬ–ФЕВРАЛЬ) ISSN 2226-1494 (PRINT), 2500-0373 (ONLINE)

ОПТИЧЕСКИЕ СИСТЕМЫ И ТЕХНОЛОГИИ

Структурные и спектральные свойства нанокристаллических порошков АИГ:Nd, АИГ:Ce и АИГ:Yb, синтезированных модифицированным методом Печини — *Муссауи А., Булыга Д.В., Игнатьев А.И., Евстропьев С.К., Никоноров Н.В.* — *1*

Расчетное прогнозирование в задаче идентификации стереоизображений — *Самойленко М.В., Хачикян В.А.* — *11*

Сравнение результатов применения двух спекловых методов изучения многоцикловой усталости конструкционной стали — *Владимиров А.П., Каманцев И.С., Друкаренко Н.А., Мызнов К.Е., Наумов К.В.* — *20*

Лазерно-индуцированное тепловое воздействие на электрические характеристики фоточувствительных пленок селенида свинца — *Ольхова А.А., Патрикеева А.А., Бутяева М.А., Пушкарева А.Е., Авилова Е.А., Москвин М.К., Сергеев М.М., Вейко В.П.* — *30*

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Алгоритм распознавания омографов на основе евклидовой метрики — *Израилова Э.С., Астемиров А.В., Бадаева А.С., Султанов З.А., Умархаджиев С.М., Хехаев М-С.Л., Ясаева М.Л.* — *41*

Улучшенная производительность модели RetinaNet для обнаружения огнестрельного оружия в пользовательском наборе данных и видеонаблюдения в реальном времени (на англ. яз.) — *Кхин П.П., Хтайк Н.М.* — *51*

Решение задачи предварительного разбиения разнородных данных на классы в условиях ограниченного объема — *Шарамет А.В.* — *62*

Исправление одиночных пакетов ошибок за пределами корректирующей способности кода с использованием информационных совокупностей — *Исаева М.Н., Овчинников А.А.* — *70*

Новый стратегический траекторно-базированный протокол для повышения эффективности беспроводных сенсорных сетей (на англ. яз.) — *Гопалакришнан Р., Сентил Кумар А.* — *81*

Автоматизация распознавания сложной текстовой CAPTCHA с применением условной генеративно-состязательной нейронной сети — *Задорожный А.С., Корепанова А.А., Абрамов М.В., Сабреков А.А.* — *90*

Основанное на особом интересе прогнозирование протоонкогена и обнаружение возможностей его мутации в онкоген на основе первоначального анализа последовательности аминокислот (на англ. яз.) — *Виджаялакшми М., Валлинаяги М.* — *101*

Метод хранения векторных представлений в сжатом виде с применением кластеризации — *Томилов Н.А., Туров В.П., Бабаянц А.А., Платонов А.В.* — *112*

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И КОГНИТИВНЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

Использование монокулярной оптики при оценке глубины объектов для двумерного картирования моделируемой среды (на англ. яз.) — *Бархум М., Пыркин А.А.* — *118*

Сегментация мышечной ткани на снимках компьютерной томографии на уровне позвонка L3 — *Теплякова А.Р., Шершнев Р.В., Старков С.О., Агабабян Т.А., Кукарская В.А.* — *124*

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

Обеспечение режимов функционирования кориолисовых вибрационных гироскопов с низкодобротными резонаторами — *Матвеев В.В., Лихошерст В.В., Каликанов А.В., Погорелов М.Г., Кирсанов М.Д., Телухин С.В.* — *133*

Сбор и обработка экологической информации в районах нефтегазодобычи и решение других прикладных задач методами активного поиска (обзорная статья) — *Свитнев И.В., Найданов А.Ф., Вилков А.В., Соколов Д.А., Лебедев М.Ю., Харитонова Е.А., Лукьянова Л.А.* — *144*

Использование технологий машинного обучения при решении задачи классификации сигналов мониторинга инфразвукового фона — *Фролов И.Н., Кудрявцев Н.Г., Сафонова В.Ю., Кудин Д.В.* — *156*

КРАТКИЕ СООБЩЕНИЯ

Исследование влияния формы выходного торца оптического волокна на гидроакустические процессы в жидкости, стимулируемые микросекундными импульсами излучения Yb,Er:Glass-лазера — *Нассер Р., Смирнов С.Н.* — *165*

Главный редактор — В.О. Никифоров, д.т.н., профессор

РЕДАКЦИОННАЯ КОЛЛЕГИЯ

Беззатеев С.В. — д.ф.-м.н., доцент (Россия)
Беликов А.В. — д.ф.-м.н., профессор (Россия)
Бессмертный И.А. — д.т.н., профессор (Россия)
Бобцов А.А. — д.т.н., профессор (Россия)
Бухановский А.В. — д.т.н. (Россия)
Волков К. Н. — д.ф.-м.н. (Великобритания)
Дукельский К.В. — д.т.н., доцент (Россия)
Ефимов Д.В. — PhD (Франция)
Заколдаев Д.А. — к.т.н., доцент (Россия)
Иорш И.В. — д.ф.-м.н. (Россия)
Кармановский Н.С. (заместитель главного редактора) — к.т.н. доцент (Россия)
Комаров И.И. — к.ф.-м.н., доцент (Россия)
Матвеев Ю.Н. — д.т.н. профессор (Россия)
Никоноров Н.В. — д.т.н., профессор (Россия)
Полищук Г.С. — к.т.н. (Россия)
Романов А.Е. — д.ф.-м.н., профессор (Россия)
Сидоркина И.Г. — д.т.н., профессор (Россия)
Степанов О.А. — д.т.н., профессор (Россия)
Татарникова Т.М. — д.т.н., доцент (Россия)
Тимофеев А. В. — д.т.н. (Казахстан)
Успенская М.В. — д.т.н., профессор (Россия)
Цыпкин А. Н. — д.ф.-м.н. (Россия)

Ответственный секретарь — И.В. Малькова

Учредитель: Университет ИТМО
Издание зарегистрировано в Федеральной службе по надзору в сфере связи, информационных технологий и массовых коммуникаций (свидетельство ПИ №ФС77-67990 от 6 декабря 2016 г.).
Языки журнала: русский, английский.
ISSN 2226-1494 (print version), ISSN 2500-0373 (online version)
Англоязычное название: «Scientific and Technical Journal of Information Technologies, Mechanics and Optics». Транслитерированное название: «Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki». Сокращенное название: «Sci.Tech. J. Inf. Technol. Mech. Opt.»
Журнал включен в каталог периодических изданий Ulrich's Periodical Directory.
Журнал входит в крупнейшие международные реферативные и наукометрические базы Scopus, EBSCO, ProQuest, EastView, ИВИС и другие.
Журнал входит в утвержденный Высшей аттестационной комиссией «Перечень ведущих рецензируемых научных журналов и изданий, выпускаемых в Российской Федерации, в которых должны быть опубликованы основные научные результаты диссертаций на соискание ученой степени доктора и кандидата наук».
Периодичность издания — 6 выпусков в год.
Плата за публикации и редактирование не взимается.

Редакторы: Ю.В. Михайлова, Л.Н. Точилина, Т.В. Точилина
Перевод — Н.А. Корнев
Компьютерная верстка — Е.С. Егорова
Сайт — К.В. Малькова

Оф. 2136, Университет ИТМО, ул. Ломоносова, д. 9,
Санкт-Петербург, Российская Федерация, 191002

Телефон (812) 480 02 75
http://ntv.ifmo.ru
E-mail: ntvitmo@itmo.ru

Статьи Журнала доступны по лицензии Creative Commons «Attribution-NonCommercial» 4.0 Всемирная.

Editor-in-Chief — Vladimir O. Nikiforov, D.Sc., Professor

EDITORIAL BOARD

Sergey V. Bezzateev — D.Sc., Associate Professor (Russia)
Andrey V. Belikov — D.Sc., Professor (Russia)
Igor A. Bessmertny — D.Sc., Professor (Russia)
Alexey A. Bobtsov — D.Sc., Professor (Russia)
Alexandr V. Boukhanovsky — D.Sc. (Russia)
Konstantin N. Volkov — D.Sc. (Great Britain)
Konstantin V. Dukelskii — D.Sc., Associate Professor (Russia)
Denis V. Efimov — PhD (France)
Danil A. Zakoldaev — PhD, Associate Professor (Russia)
Ivan V. Iorsh — D.Sc. (Russia)
Nikolai S. Karmanovskiy (Deputy Chief Editor) — PhD, Associate Professor (Russia)
Igor I. Komarov — PhD, Associate Professor (Russia)
Iuri N. Matveev — D.Sc., Professor (Russia)
Nikolay V. Nikonorov — D.Sc., Professor (Russia)
Grigorii S. Polishchuk — PhD (Russia)
Alexey E. Romanov — D.Sc., Professor (Russia)
Irina G. Sidorkina — D.Sc., Professor (Russia)
Oleg A. Stepanov — D.Sc., Professor (Russia)
Tatiana M. Tatarnikova — D.Sc., Associate Professor (Russia)
Andrey V. Timofeev — Dr. habil. Sc. (Kazakhstan)
Mayya V. Uspenskaya — D.Sc., Professor (Russia)
Anton N. Tcypkin — D.Sc., Professor (Russia)

Executive secretary — Irina V. Malkova

Founder: ITMO University
Publication is registered by the Federal Inspectorate Service for Communication, Information Technologies and Communication Media, certificate PI FS77-67 990 dated December, 6th, 2016.
Languages of the journal: Russian, English
ISSN 2226-1494 (print version), ISSN 2500-0373 (online version) The English title is "Scientific and Technical Journal of Information Technologies, Mechanics and Optics".
Transliterated title is "Nauchno-Tekhnicheskii Vestnik Informatsionnykh Tekhnologii, Mekhaniki i Optiki".
Abbreviated parallel title: "Sci.Tech. J. Inf. Technol. Mech. Opt."
The journal is catalogued in Ulrich's Periodical Directory.
The journal is included in the largest international abstract and scientometric databases Scopus, EBSCO, ProQuest, East View, EVIS and others.
The journal is included in the "List of leading scientific journals and periodicals under review in the Russian Federation, where the main scientific results of theses for the PhD and doctor of sciences degree must be published" approved by the Higher Attestation Commission.
Publication frequency is 6 times a year.
Publication and editing are free of charge.

Editors: Yulia V. Mikhailova, Lyubov N. Tochilina, Tatiana V. Tochilina
English Language Technical Editor — Nikolay A. Kornev
Computer layout — Ekaterina S. Egorova
Website — Kseniia V. Malkova

of. 2136, ITMO University, Lomonosova Street, 9,
St.Petersburg, 191002, Russia

Phone (812) 480 02 75
http://ntv.ifmo.ru/en
E-mail: ntvitmo@itmo.ru



Подписано к печати 01.02.2024 Тираж 350 экз. Заказ № 97 (149)
Издание Университета ИТМО
Отпечатано в ООО «Университетские телекоммуникации»
Адрес: 199034, Санкт-Петербург, Биржевая линия, д. 16
© Университет ИТМО

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

ОПТИЧЕСКИЕ СИСТЕМЫ И ТЕХНОЛОГИИ

OPTICAL ENGINEERING

doi: 10.17586/2226-1494-2024-24-1-1-10
УДК 54.057

Структурные и спектральные свойства нанокристаллических порошков АИГ:Nd, АИГ:Ce и АИГ:Yb, синтезированных модифицированным методом Печини

Амир Муссауи[1], Дмитрий Владимирович Булыга[2]✉, Александр Иванович Игнатьев[3], Сергей Константинович Евстропьев[4], Николай Валентинович Никоноров[5]

[1,2,3,4,5] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
[2,4] Акционерное общество «Научно-производственное объединение Государственный оптический институт им. С.И. Вавилова», Санкт-Петербург, 192171, Российская Федерация
[4] Санкт-Петербургский государственный технологический институт (Технический университет), Санкт-Петербург, 190013, Российская Федерация

[1] am.moussaoui92@gmail.com, https://orcid.org/0000-0002-7858-7417
[2] dmbulyga@yandex.ru✉, https://orcid.org/0000-0001-8861-2527
[3] alexandre_ignatiev@yahoo.com, https://orcid.org/0000-0002-6114-0248
[4] evstropiev@bk.ru, https://orcid.org/0000-0002-0160-8443
[5] Nikonorov@oi.ifmo.ru, https://orcid.org/0000-0002-1341-067X

Аннотация

Введение. В работе исследованы структурные и люминесцентные свойства нанокристаллических порошков состава $Y_3Al_5O_{12}:Re^{3+}$ (Re^{3+} = Nd^{3+}, Yb^{3+}, Ce^{3+}), синтезированных модифицированным методом Печини с использованием поливинилпирролидона в качестве дополнительного стабилизатора и органического «топлива» в ходе синтеза. **Метод.** Для исследования термической эволюции гелей и свойств конечного продукта применены фурье-спектроскопия, люминесцентная спектроскопия, дифференциальная сканирующая калориметрия, сканирующая электронная микроскопия, а также рентгенофазовый анализ. **Основные результаты.** Показано, что применение модифицированного метода Печини при использовании добавок поливинилпирролидона в исходные растворы обеспечивает получение высокодисперсных порошков алюмоиттриевого граната, содержащих редкоземельные ионы. Установлено, что введение поливинилпирролидона оказывает существенное влияние на эволюцию композиционных гелей при их термообработке и формирование нанокристаллов алюмоиттриевого граната происходит при температуре около 883 °C. **Обсуждение.** Полученные в работе нанокристаллические порошки могут быть использованы в качестве прекурсоров для спекания лазерной керамики и введены в оптическое волокно для создания усилителей.

Ключевые слова

алюмоиттриевый гранат, люминесценция, метод Печини, поливинилпирролидон

Ссылка для цитирования: Муссауи А., Булыга Д.В., Игнатьев А.И., Евстропьев С.К., Никоноров Н.В. Структурные и спектральные свойства нанокристаллических порошков АИГ:Nd, АИГ:Ce и АИГ:Yb, синтезированных модифицированным методом Печини // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 1–10. doi: 10.17586/2226-1494-2024-24-1-1-10

© Муссауи А., Булыга Д.В., Игнатьев А.И., Евстропьев С.К., Никоноров Н.В., 2024

Structural and spectral properties of YAG:Nd, YAG:Ce and YAG:Yb nanocrystalline powders synthesized via modified Pechini method

Amir Moussaoui[1], Dmitry V. Bulyga[2]✉, Aleksandr I. Ignatiev[3], Sergey K. Evstropiev[4], Nikolay V. Nikonorov[5]

[1,2,3,4,5] ITMO University, Saint Petersburg, Russian Federation
[2,4] JVC "RPA Vavilov State Optical Institute", Saint Petersburg, 192171, Russian Federation
[4] Saint Petersburg State Technological Institute (Technical University), Saint Petersburg, 190013, Russian Federation

[1] am.moussaoui92@gmail.com, https://orcid.org/0000-0002-7858-7417
[2] dmbulyga@yandex.ru✉, https://orcid.org/0000-0001-8861-2527
[3] alexandre_ignatiev@yahoo.com, https://orcid.org/0000-0002-6114-0248
[4] evstropiev@bk.ru, https://orcid.org/0000-0002-0160-8443
[5] Nikonorov@oi.ifmo.ru, https://orcid.org/0000-0002-1341-067X

Abstract
Synthesis of nanocrystalline yttrium-aluminum garnet doped with neodymium was performed via modified Pechini methods. Evolution of material during synthesis was studied using differential thermal analysis; the structure and morphology of synthesized nanopowders were studied using scanning electron microscopy and x-ray diffraction. It was shown that the use of an additional low-temperature stabilizer leads to formation of crystalline yttrium-aluminum garnet phase at lower temperatures. It was shown that the formation of nanocrystals occurs at the temperature of about 883 °C. Obtained powders can be used as precursors for ceramics sintering or be introduced into the optical fiber in order to fabricate optical amplifiers.

Keywords
YAG, luminescence, Pechini method, polyvinylpyrrolidone

For citation: Moussaoui A., Bulyga D.V., Ignatiev A.I., Evstropiev S.K., Nikonorov N.V. Structural and spectral properties of YAG:Nd, YAG:Ce and YAG:Yb nanocrystalline powders synthesized via modified Pechini method. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 1–10 (in Russian). doi: 10.17586/2226-1494-2024-24-1-1-10

Введение

Кристаллы алюмоиттриевого граната (АИГ, $Y_3Al_5O_{12}$), активированные редкоземельными компонентами, широко применяются в оптике и оптоэлектронике, благодаря высоким спектрально-люминесцентным свойствам [1–5]. Структура и свойства этого материала хорошо известны, и он широко используется в виде монокристаллов и поликристаллической оптической керамики в различных лазерных системах.

АИГ, легированный ионами редкоземельных металлов, широко применяется в производстве дисплеев и лазеров [6, 7]. АИГ, содержащий редкоземельные модификаторы, в виде дисперсных порошков применяется в светодиодной технике, сенсорике и нанотермометрии [1, 8, 9] и является промежуточным продуктом при формировании лазерной керамики [5, 10]. При использовании высокодисперсных порошков АИГ важную роль играют морфология материала, размер кристаллов и особенности их кристаллической структуры, в значительной мере определяющиеся методом и технологическими условиями его получения [11–13].

Нанокристаллические порошки АИГ могут быть получены при помощи твердофазного синтеза [14], золь-гель метода [11], метода сжигания органического топлива (combustion synthesis) [8, 12], гидротермального синтеза [15], пиролиза аэрозолей [13], осаждения из раствора [16], полимерно-солевого метода [9, 10]. Одним из наиболее простых и эффективных методов получения порошков АИГ является золь-гель метод Печини, в котором в качестве стабилизатора и органического топлива используется лимонная кислота. Данный метод универсален, не требует сложного технологического оборудования и обеспечивает формирование дисперсных оксидных порошков с высокими люминесцентными свойствами [17].

В работе [17] для получения высокодисперсных порошков Gd_2O_3, содержащих ионы Nd^{3+}, был использован модифицированный метод Печини, в котором, наряду с относительно низкотемпературным, традиционно используемым стабилизатором (лимонная кислота), применялись добавки в исходные растворы поливинилпирролидона (ПВП). Этот растворимый органический полимер образует в растворах металлополимерные комплексы [18], разлагается при нагревании до температур более 400 °C и широко используется в формировании различных наночастиц полимерно-солевым методом [9, 10]. Лимонная кислота образует хелатные комплексы с ионами металлов, что позволяет использовать ее в качестве стабилизатора при высушивании растворов для сохранения однородности и препятствования выпадению осадков солей металлов. Использование двух стабилизаторов (ПВП и лимонной кислоты) с различными температурами разложения позволяет продлить процесс выделения газообразных продуктов реакции и тем самым повысить дисперсность синтезируемых порошков [17].

Целью данной работы является синтез нанокристаллических порошков состава АИГ:Re^{3+} (Re: Ce^{3+}, Yb^{3+}, Nd^{3+}) модифицированным методом Печини [17] с использованием ПВП и без него, а также исследование фазового состава, структуры, морфологии поверхности и люминесцентных свойств полученного материала.

Материалы и методы

В процессе синтеза были использованы следующие химические реактивы: нитрат иттрия (Y) 6-во-

дный (химический чистый (ХЧ)), нитрат алюминия (Al) 9-водный (ХЧ), нитрат церия (Ce) 6-водный (ХЧ), нитрат неодима (Nd) 6-водный (ХЧ), хлорид иттербия (Yb) 6-водный (ХЧ), лимонная кислота 1-водная, ПВП (Sigma Aldrich, M_w = 1 300 000) и дистиллированная деионизованная вода. ПВП и лимонная кислота играют роль стабилизаторов в ходе синтеза, препятствующих выпадению осадка в ходе высушивания раствора, а также органического топлива, препятствующего дальнейшему росту нанокристаллов в ходе термообработки за счет выделения газообразных продуктов разложения.

Химический состав исходных растворов представлен в табл. 1. Было приготовлено шесть растворов с использованием ПВП и без него, соответствующих составам конечных продуктов АИГ:Ce^{3+}, АИГ:Nd^{3+}, АИГ:Yb^{3+}. После тщательного перемешивания при температуре 70 °C в течение двух часов растворы высушивались при той же температуре, затем полученные гели в течение двух часов подвергались термообработке в муфельной печи при температуре 1000 °C. Рассчитанный химический состав полученных нанокристаллических порошков (далее по тексту: нанопорошки 1–6) соответствует АИГ с 1,0 мол.% R_2O_3 (R = Ce, Nd, Yb). Поскольку Ce, Nd, Yb имеют большую атомную массу, изменение их массы в ходе термообработки гелей незначительно.

Выполним исследование термической эволюции гелей и свойства синтезированных нанокристаллических порошков следующими методами: фурье-спектроскопии, люминесцентной спектроскопии, дифференциальной сканирующей калориметрии, сканирующей электронной микроскопии и рентгенофазового анализа.

Рассмотрим кристаллическую структуру и фазовый состав материалов с помощью дифрактометра Rigaku Ultima IV. Оценку среднего размера кристаллов проведем по формуле Шеррера. Осуществим регистрацию спектров инфракрасного поглощения порошков с использованием фурье-спектрометра Bruker ALPHA с приставкой на неполное внутреннее отражение.

Проведем исследования термической эволюции гелей на дифференциальном сканирующем калориметре STA 449F1 Jupiter (Netzsch), а морфологии частиц порошков на сканирующем электронном микроскопе MIRA3 TESCAN. Спектры люминесценции в диапазоне λ = 250–800 нм изучим на спектрофлюориметре Perkin-Elmer LS 50B.

Экспериментальные результаты

Термообработка высушенных гелей привела к полному разложению органических соединений, что вызвало уменьшение массы образцов и термические эффекты. На рис. 1 представлены данные дифференциальной сканирующей калориметрии для растворов гелей 3 и 4. Небольшой эндотермический процесс, наблюдающийся в диапазоне температур 20–120 °C с потерей массы около 5 % (рис. 1, *a*), связан с дегидратацией гелей. Экзотермические процессы, происходящие в области температур 120–750 °C, связаны с разложением цитратов, нитратов и ПВП, окислением органических соединений (120–450 °C) и последующим окислением остаточного углерода кислородом воздуха с экзотермическим эффектом (450–750 °C) до полного разложения органических остатков.

Небольшой эндотермический процесс, наблюдающийся для геля 4 при температурах до 120 °C с потерей массы около 5 % связан с дегидратацией геля (рис. 1, *b*). В области температур 120–700 °C наблюдается ряд экзотермических процессов. Эффекты в области температур 200–600 °C соответствуют окислению кислородом воздуха и нитратами металлов органических соединений и остаточного углерода, а также могут быть связаны с образованием аморфных оксидов.

Согласно данным рентгенофазового анализа [17] в ходе синтеза оксида гадолиния модифицированным методом Печини, при 600 °C материал почти полностью состоит из аморфной фазы, в нем также содержатся кристаллические частицы со средним размером порядка 10 нм. При 1000 °C порошок полностью состоит из кристаллической фазы. Отсюда можно сделать вывод, что экзотермические пики с максимумами 883 °C (рис. 1, *a*) и 895 °C (рис. 1, *b*) связаны с образованием кристаллического АИГ из аморфных оксидов, образовавшихся при прокаливании гелей. В [19] образование кристаллов АИГ:Ce^{3+} из растворов гелей, полученных при использовании добавок этиленгликоля, наблюдалось при 900 °C. Формирование кристаллической фазы АИГ:Nd^{3+} при термообработке нанопорошков, полученных методом соосаждения при применении гидрокарбоната аммония [20] или мочевины [21], происходило при температурах 948 °C [20] и 900 °C [21]. В [22] образование кристаллической фазы АИГ:Eu^{3+} из растворов гелей также наблюдалось при температуре 900 °C.

Таблица 1. Химический состав исходных растворов
Table 1. Chemical composition of the initial solutions

Номер раствора	Концентрация, масс.%							
	Вода	ПВП	Лимонная кислота	$Y(NO_3)_3$	$Al(NO_3)_3$	$Ce(NO_3)_3$	$Nd(NO_3)_3$	$YbCl_3$
1	87,20	5,11	4,26	1,48	1,92	—	0,03	—
2	91,91	—	4,50	1,56	2,00	—	0,03	—
3	87,20	5,11	4,26	1,48	1,92	0,03	—	—
4	91,91	—	4,50	1,56	2,00	0,03	—	—
5	87,21	5,11	4,26	1,48	1,92	—	—	0,02
6	91,91	—	4,50	1,56	2,00	—	—	0,02

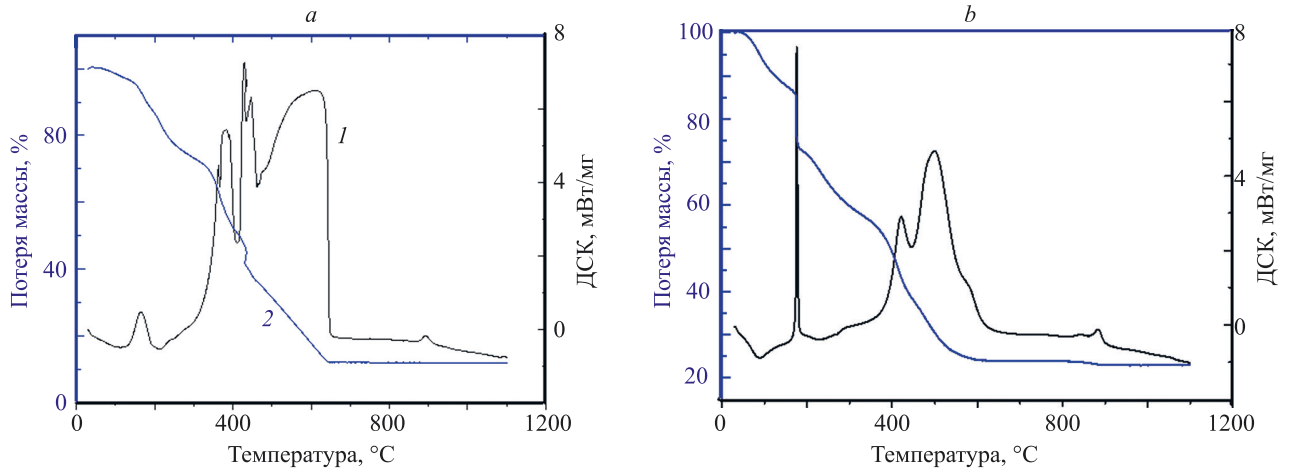


Рис. 1. Результаты дифференциальной сканирующей калориметрии (ДСК) (кривые *1*) и термогравиметрического анализа (кривые *2*) для растворов гелей 3 (*a*) и 4 (*b*)

Fig. 1. The results of differential scanning calorimetry (curves *1*) and thermogravimetric analysis (curves *2*) gel solutions 3 (*a*) and 4 (*b*)

По данным [17] введение ПВП в исходный раствор значительно ускоряет формирование кристаллической фазы, уменьшает минимальную температуру, необходимую для ее формирования, а также уменьшает средний размер нанокристаллов.

На рис. 2, *a* представлены спектры инфракрасного поглощения высушенных растворов гелей 1, 3 и 5. Интенсивная полоса с максимумом около 2945 $см^{-1}$ соответствует колебаниям группы O–H воды, содержащейся в образцах [22], полоса с максимумом в области 1620 $см^{-1}$ — колебаниям карбонильной группы –COO–, содержащейся в молекуле лимонной кислоты [23]. К колебаниям C–H относится пик с максимумом около 1290 $см^{-1}$, а полоса поглощения с максимумом 1415 $см^{-1}$ — нитратам [24]. Следует отметить, что в спектре растворов гелей наблюдаются полосы колебаний связей M–O (M = Y и Al) в области спектра 550–850 $см^{-1}$ [22].

На рис. 2, *b* показаны инфракрасные спектры нанопорошков, полученных путем термообработки растворов гелей при температуре 1000 °C. В спектрах наблюдаются полосы с максимумами около 685, 720 и 787 $см^{-1}$. Колебаниям связи Al-O соответствуют полосы поглощения с максимумами около 787 и 685 $см^{-1}$, а Y–O — полоса с максимумом 720 $см^{-1}$ [25]. Полученные инфракрасные спектры полностью соответствуют спектру кристалла АИГ.

На рис. 3 приведены электронно-микроскопические снимки нанокристаллических порошков 4 и 3. Как видно из рис. 3, *a*, *c*, полученные материалы состоят из частиц микронного размера, имеющих сложную структуру. Наблюдается заметное различие в морфо-

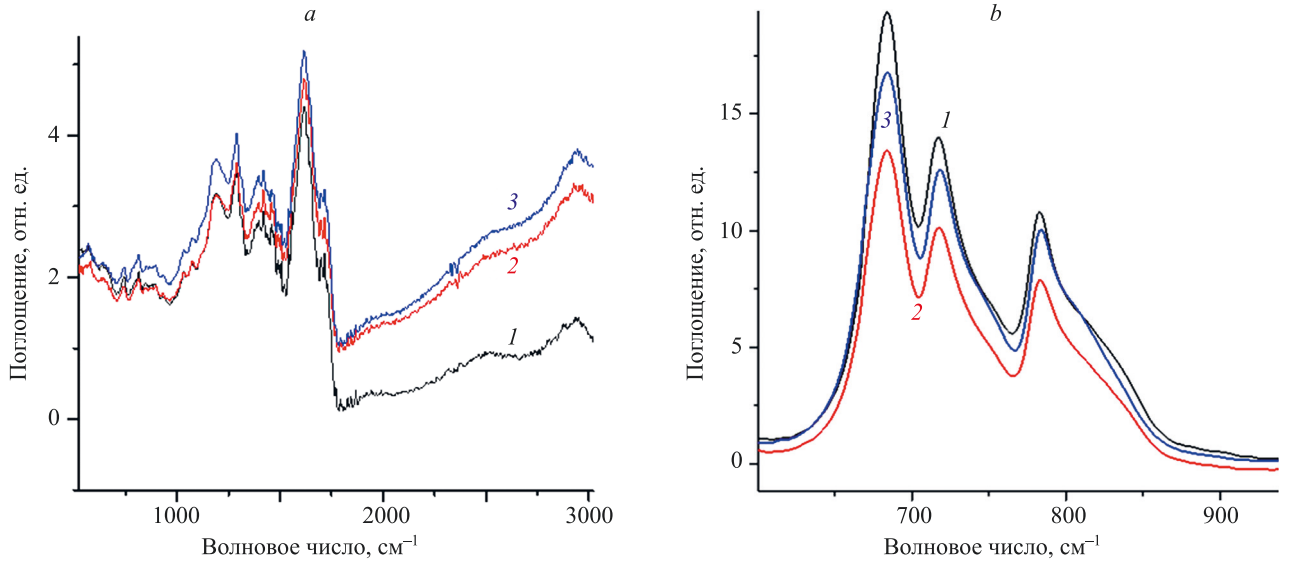


Рис. 2. Инфракрасные спектры высушенных растворов гелей 1 (кривая *1*), 3 (кривая *2*) и 5 (кривая *3*) (*a*) и нанокристаллических порошков, полученных из растворов гелей 1 (кривая *1*), 3 (кривая *2*) и 5 (кривая *3*), синтезированных при температуре 1000 °C (*b*)

Fig. 2. IR spectra of dried gel solutions 1 (curve *1*), 3 (curve *2*) and 5 (curve *3*) (*a*) and nanocrystalline powder obtained from the gel solution 1 (curve *1*) synthesized at a temperature of 1000 °C (*b*)

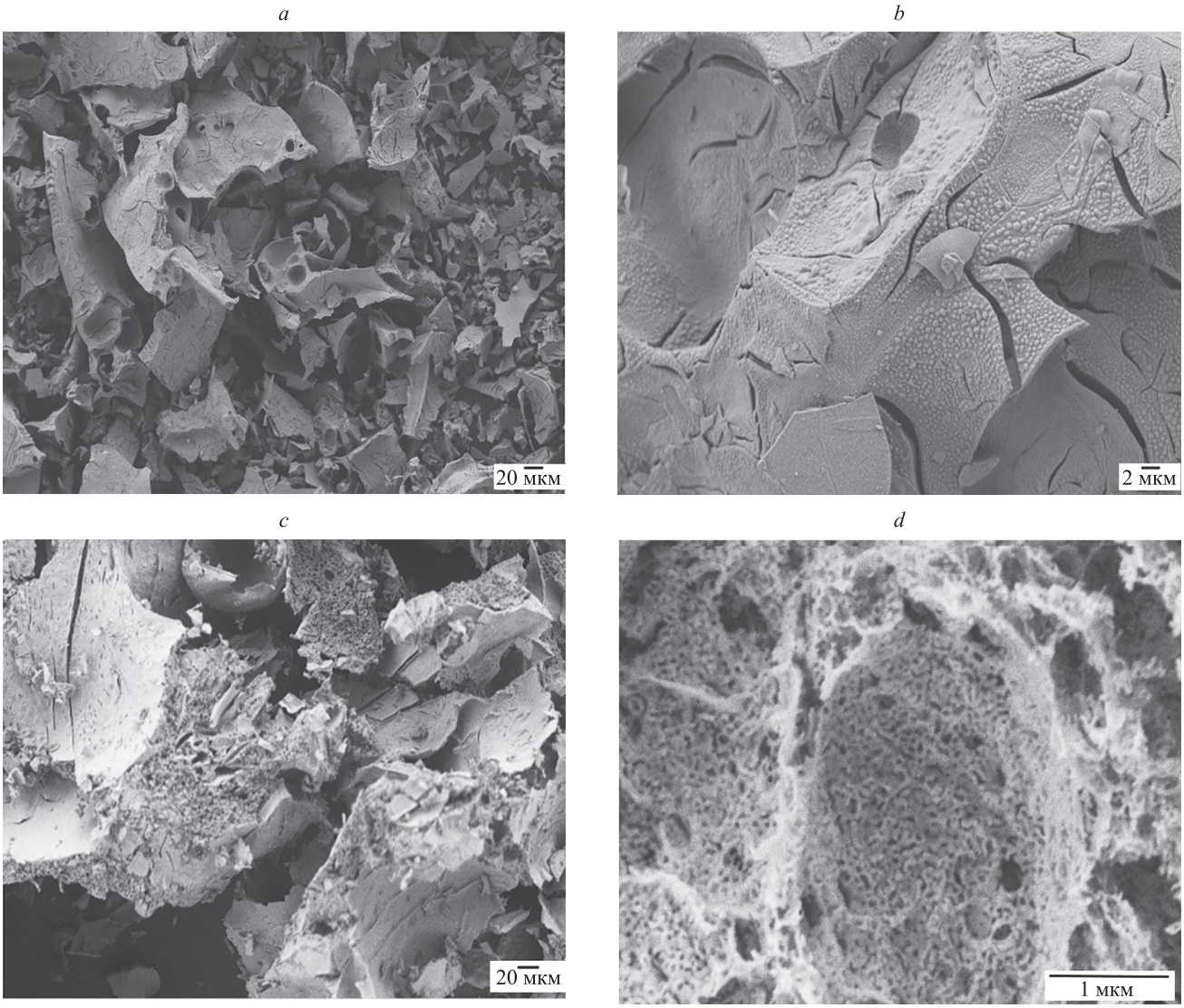


Рис. 3. Электронно-микроскопические снимки нанопорошков 4 (*a*, *b*) и 3 (*c*, *d*)
Fig. 3. Electron microscopic images of nanopowders 4 (*a*, *b*) and 3 (*c*, *d*)

логии частиц порошка, полученного с использованием ПВП и без него. Нанопорошок 4, полученный без применения ПВП, состоит из плотных частиц микронного размера (рис. 3, *a*), на поверхности которых видны отдельные наночастицы (рис. 3, *b*). В структуре нанопорошка 3, синтезированного при использовании добавки ПВП, наблюдаются крупные высокопористые агрегаты (рис. 3, *c*), состоящие из наночастиц (рис. 3, *d*). Наличие пористой структуры в материалах, полученных при применении ПВП, связано с сильным газовыделением при разложении нитратов и ПВП при температурах 400–600 °C (рис. 1, *a*). Такие особенности морфологии материалов, синтезированных с использованием ПВП, отмечены в работах [17, 24].

На рис. 4 приведены рентгенограммы нанопорошков 3 (кривая *1*) и 4 (кривая *2*). Аналогичные дифрактограммы были получены для всех порошков АИГ, содержащих различные редкоземельные ионы. Видно, что полученные материалы целиком состоят из кубической фазы АИГ и обладают высокой кристалличностью, что является важным при создании светодиодов, поскольку степень кристалличности оказывает существенное влияние на эффективность люминесценции [26].

Расчет среднего размера нанокристаллов осуществим по формуле Шеррера [27]:

$$d = k\lambda/(\beta\cos\theta),$$

где k — константа Шеррера (k = 0,9 для кубической фазы); λ — длина волны рентгеновского излучения (λ = 0,15406 нм для Cu Kα); β — ширина на полувысоте наиболее интенсивного пика в дифрактограмме; θ — угол Брэгга.

В результате получено, что средний размер нанокристаллов в синтезированных порошках варьируется от 40 до 68 нм. Наибольший средний размер нанокристаллов (60, 68 нм) достигается для растворов гелей 5 и 6, активированных Yb. Все эти явления, вероятно, связаны с особенностями процесса кристаллизации АИГ:Yb^{3+}.

Введение ПВП в исходный раствор приводит к заметному уменьшению средних размеров наночастиц порошка. Подобный эффект описан в работе [17] для

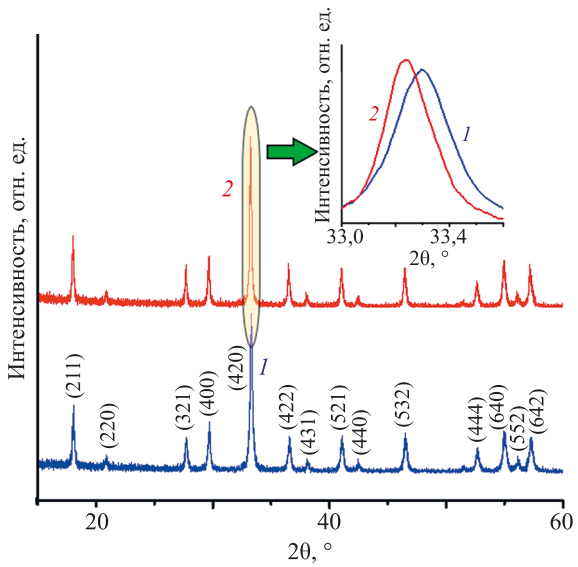


Рис. 4. Дифрактограммы нанопорошков 3 (кривая *1*) и 4 (кривая *2*).

На вставке показан сдвиг пиков в сторону больших углов при синтезе с использованием поливинилпирролидона

Fig. 4. Diffractograms of nanopowders 3 (curve *1*) and 4 (curve *2*).

The inset shows the shift of peaks towards higher angles during synthesis using polyvinylpyrrolidone

наночастиц Gd_2O_3:Nd^{3+}. Это явление может быть объяснено сильным выделением газов при разложении ПВП и нитратов металлов, происходящем в температурном интервале 400–600 °C (рис. 1, *а*). Выделяющиеся в больших объемах газы обеспечивают пространственное разделение образующихся из нитратов частиц оксидов металлов, что при дальнейшей термообработке, при более высоких температурах, приводит к формированию более мелких нанокристаллов.

Ионные радиусы редкоземельных активаторов, внедренных в кристаллы АИГ, заметно отличаются от замещаемого ими иона Y (ионные радиусы Y^{3+}, Yb^{3+}, Nd^{3+} и Ce^{3+} составляют 0,1019; 0,0985; 0,1109 и 0,1143 нм соответственно [28]). Это определяет деформацию и изменение размеров элементарной ячейки АИГ (табл. 2). Отметим, что деформация элементарной ячейки кристаллов (отклонение значения параметра кристаллической решетки от значения для монокристаллического АИГ) при синтезе наночастиц АИГ:Ce модифицированным методом Печини и применении добавки ПВП меньше, чем наблюдалась в нанокристаллах, синтезированных в [19] аналогичным методом при использовании добавки этиленгликоля (табл. 2). Уменьшение межплоскостного расстояния также свидетельствует о сдвиге пиков в сторону больших углов при синтезе с использованием ПВП (рис. 4). Полученный результат может быть связан с тем, что в ходе синтеза с использованием ПВП нанокристаллы растут из зародышей с меньшим размером, чем при использовании добавок с меньшей температурой разложения.

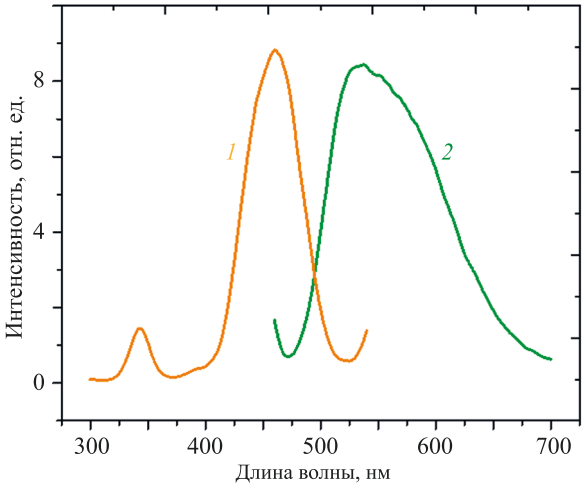


Рис. 5. Спектры возбуждения люминесценции (кривая *1*) и люминесценции (кривая *2*) порошка АИГ:Ce^{3+}, синтезированного с использованием поливинилпирролидона

Fig. 5. Luminescence excitation spectrum (curve *1*) and luminescence spectrum (curve *2*) of YAG:Ce^{3+} powder synthesized using PVP

Таблица 2. Параметры решетки и средний размер нанокристаллов в порошках АИГ:R^{3+} (R^{3+} = Ce^{3+}, Nd^{3+}, Yb^{3+})

Table 2. Lattice parameters and average size of nanocrystals in YAG powders: R^{3+} (R^{3+} = Ce^{3+}, Nd^{3+}, Yb^{3+})

Образец	Средний размер нанокристаллов, нм	Параметры кристаллической решетки	
		a, нм	*V*, нм^3
АИГ:Nd с ПВП (раствор 1)	40	1,20363	1,74371
АИГ:Nd без ПВП (раствор 2)	42	1,20629	1,75532
АИГ:Ce с ПВП (раствор 3)	44	1,20348	1,74306
АИГ:Ce без ПВП (раствор 4)	63	1,20504	1,74987
АИГ:Yb с ПВП (раствор 5)	60	1,20358	1,74352
АИГ:Yb без ПВП (раствор 6)	68	1,20512	1,75001
Монокристаллический АИГ [15]	—	1,20020	1,72886
Монокристаллический АИГ, (JCPDS № 01-073-3184)	—	1,20100	1,73232
Нанопорошок АИГ:Ce (1 мол.%) [19]	29	1,20390	1,74490
Нанопорошок АИГ:Nd (1,2 масс.%) [10]	40	1,20610	1,75449

Известно, что повышение температуры термообработки нанокристаллов АИГ, сформированных низкотемпературным методом, повышает степень кристалличности материала и уменьшает размер элементарной ячейки нанокристаллов [19]. Введение ПВП в состав исходной смеси приводит к существенному повышению температуры материала в процессе термообработки (рис. 1), что определяет уменьшение параметров элементарных ячеек формирующихся нанокристаллов АИГ (табл. 2).

На рис. 5 представлены спектры возбуждения (кривая *1*) и люминесценции (кривая *2*) нанокристаллического порошка АИГ:Ce^{3+}. Полоса возбуждения с максимумом около 454 нм (электронный переход $4f^1 \rightarrow 4f^05d^1$ ионов Ce^{3+}) хорошо соответствует полосе излучения светодиодов на основе InGaN, таким образом, АИГ:Ce^{3+} может использоваться в качестве переизлучателя для данного типа светодиодов [29]. Асимметричная полоса люминесценции с максимумом около 540 нм связана с электронными переходами $5d^1 \rightarrow {}^2F_{5/2}$ и $5d^1 \rightarrow {}^2F_{7/2}$ ионов Ce^{3+} [19].

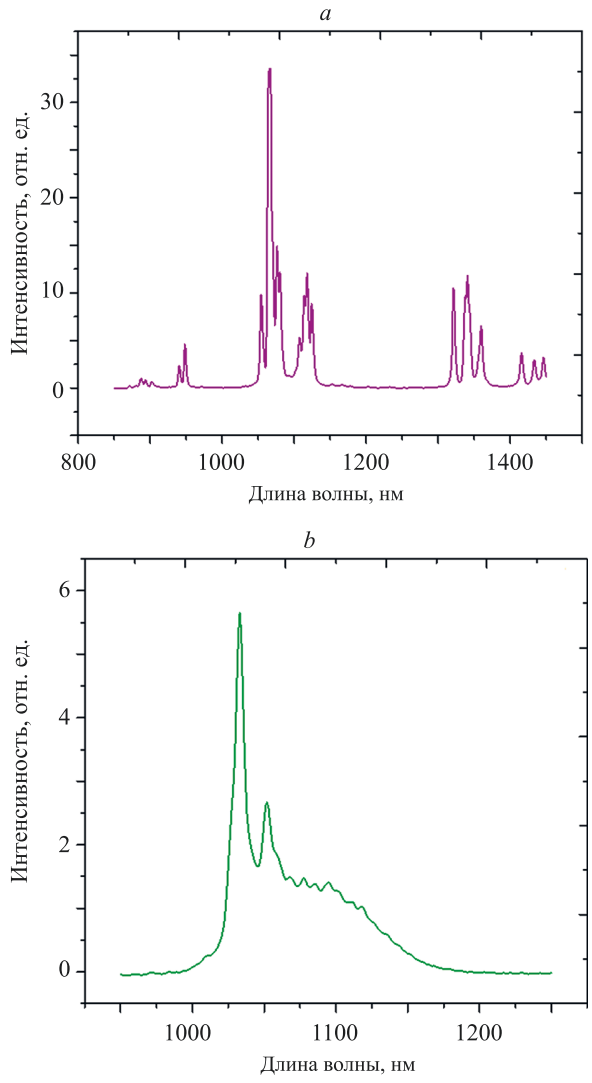


Рис. 6. Спектры люминесценции нанопорошков АИГ:Nd^{3+} (*a*) и АИГ:Yb^{3+} (*b*), синтезированных с использованием поливинилпирролидона

Fig. 6. Luminescence spectra of YAG nanopowders: YAG:Nd^{3+} (*a*), YAG:Yb^{3+} (*b*) synthesized using PVP

Спектры люминесценции образцов АИГ:Nd^{3+} и АИГ:Yb^{3+} представлены на рис. 6. В спектре люминесценции образца, легированного Nd, наблюдаются характерные полосы люминесценции в инфракрасном диапазоне 1025–1150 нм, соответствующие электронному переходу ${}^4F_{3/2} \rightarrow {}^4I_{11/2}$ в ионах Nd [30]. Две менее интенсивные группы полос в области около 950 и 1340 нм определяются переходами ${}^4F_{3/2} \rightarrow {}^4I_{9/2}$ и ${}^4F_{3/2} \rightarrow {}^4I_{13/2}$ (рис. 6, *a*).

По полученным спектрам люминесценции нанопорошка АИГ:Yb^{3+} (рис. 6, *b*) видна наиболее интенсивная полоса люминесценции с максимумом около 1030 нм. Остальные полосы с меньшей интенсивностью соответствуют электронному переходу ионов Yb ${}^2F_{5/2} \rightarrow {}^2F_{7/2}$ [10].

Полученные спектры люминесценции полностью соответствуют спектрам материалов аналогичного химического состава (АИГ:Ce^{3+}, АИГ:Nd^{3+}, АИГ:Yb^{3+}), синтезированных при помощи модифицированного метода Печини [31], цитратного золь-гель метода [32], а также полимерно-солевого метода [10].

Заключение

В работе для синтеза нанокристаллических порошков алюмоиттриевого граната, активированного ионами редкоземельных элементов (Ce, Nd, Yb) применен модифицированный метод Печини. Метод основан на использовании двух стабилизирующих добавок: водорастворимого полимера (поливинилпирролидона), способного образовывать комплексы с ионами металлов, а также лимонной кислоты, играющей роль хелатирующего агента, т. е. структуры, которая способна одновременно присоединять два или более донорных атома к одному и тому же иону металла и образовывать одно или несколько колец.

Исследования, проведенные методами дифференциальной сканирующей калориметрии и термогравиметрии, показали, что процессы кристаллизации материалов происходят при температурах более 880–900 °C. Все материалы, синтезированные при 1000 °C, полностью состоят только из кристаллов алюмоиттриевого граната. Средний размер полученных кристаллов АИГ:R^{3+} варьируется от 40 до 68 нм. Добавка поливинилпирролидона в исходные растворы сформировала кристаллы АИГ:R^{3+} меньшего размера. Внедрение редкоземельных ионо-активаторов в кристаллическую решетку алюмоиттриевого граната привело к деформации элементарных ячеек кристаллов. Применение добавок поливинилпирролидона обеспечило формирование кристаллов с более компактной кристаллической структурой (с меньшим объемом кристаллической решетки).

Согласно данным сканирующей электронной микроскопии, добавка поливинилпирролидона оказывает существенное влияние на морфологию частиц порошка. Структура порошков, полученных без добавок поливинилпирролидона, состоит из микроскопических агрегатов, содержащих плотноупакованные нанокристаллы

алюмоиттриевого граната. Добавка поливинилпирролидона в исходные смеси позволяет сформировать в результате синтеза пористые частицы. Подобное изменение морфологии частиц не оказывает существенного влияния на люминесцентные свойства порошков, однако играет важную роль в задаче введения частиц в волокно для создания волоконных усилителей, а также при спекании прозрачной керамики, что является одной из наиболее перспективных областей применения полученного материала.

Спектры люминесценции нанопорошков в целом соответствуют спектрам монокристаллического алюмоиттриевого граната, содержащего редкоземельные активаторы. Добавление поливинилпирролидона в исходный раствор не оказывает существенного влияния на спектры люминесценции синтезируемого материала. Синтезированные порошки алюмоиттриевого граната могут служить в качестве прекурсоров для спекания керамики, а также как материалы для создания люминесцентных датчиков температуры.

Литература

1. Chen L., Luo Y., Xia Y., Kang B., Yu S. Densification, microstructure and optical properties of YAG transparent ceramics prepared by dry-pressing and gelcasting // Optical Materials. 2021. V. 121. P. 111509. https://doi.org/10.1016/j.optmat.2021.111509
2. Li J., Liu J., Liu B., Liu W., Zeng Y., Ba X., Xie T., Jiang B., Liu Q., Pan Y., Feng X., Guo J. Influence of heat treatment of powder mixture on the microstructure and optical transmission of Nd:YAG transparent ceramics // Journal of the European Ceramic Society. 2014. V. 34. N 10. P. 2497–2507. https://doi.org/10.1016/j.jeurceramsoc.2014.03.004
3. Wang H.M., Huang Z.Y., Jiang J.S., Liu K., Duan M.Y., Lu Z.W., Cedelle J., Guan Z.W., Lu T.C., Wang Q.Y. Unique mechanical properties of nano-grained YAG transparent ceramics compared with coarse-grained partners // Materials & Design. 2016. V. 105. P. 9–15. https://doi.org/10.1016/j.matdes.2016.04.094
4. Sim S.-M., Keller K.A., Mah T.-I. Phase formation in yttrium aluminum garnet powders synthesized by chemical methods // Journal of Materials Science. 2000. V. 35. N 3. P. 713–717. https://doi.org/10.1023/A:1004709401795
5. Tachiwaki T., Yoshinaka M., Hirota K., Ikegami T., Yamaguchi O. Novel synthesis of $Y_3Al_5O_{12}$ (YAG) leading to transparent ceramics // Solid State Communications. 2001. V. 119. N 10-11. P. 603–606. https://doi.org/10.1016/s0038-1098(01)00293-9
6. Sluzky E., Lemoine M., Hesse K. Phosphor development for α-silicon liquid crystal light valve projection display // Journal of the Electrochemical Society. 1994. V. 141. N 11. P. 3172. https://doi.org/10.1149/1.2059297
7. Lu C.H., Huang C.H. Sensitized photoluminescence of Eu^{3+} and Gd^{3+}-doped $Y_3Al_5O_{12}$ phosphors prepared via a reverse microemulsion process // Chemistry Letters. 2004. V. 33. N 12. P. 1568–1569. https://doi.org/10.1246/cl.2004.1568
8. Benayas A., del Rosal B., Pérez-Delgado A., Santacruz-Gómez K., Jaque D., Hirata G.A., Vetrone F. Nd:YAG near-infrared luminescent nanothermometers // Advanced Optical Materials. 2015. V. 3. N 5. P. 687–694. https://doi.org/10.1002/adom.201400484
9. Евстропьев С.К., Демидов В.В., Булыга Д.В., Садовничий Р.В., Пчелкин Г.А., Шурупов Д.Н., Подрухин Ю.Ф., Матросова А.С., Никоноров Н.В., Дукельский К.В. Люминесцентные волоконно-оптические датчики на основе нанолюминофоров YAG : R^{3+} (R = Ce, Dy, Yb) для измерения температуры в диапазоне 20–500 °C // Квантовая электроника. 2022. Т. 52. № 1. С. 94–99.
10. Bulyga D.V., Evstropiev S.K. Intermediate products of Yb:YAG laser ceramics fabrication: structural features, morphology, and luminescent properties // Research on Chemical Intermediates. 2021. V. 47. N 8. P. 3501–3514. https://doi.org/10.1007/s11164-021-04484-w
11. Saladino M.L., Nasillo G., Martino D.C., Caponetti E. Synthesis of Nd:YAG nanopowder using the citrate method with microwave irradiation // Journal of Alloys and Compounds. 2010. V. 491. N 1-2. P. 737–741. https://doi.org/10.1016/j.jallcom.2009.11.054
12. Laishram K., Mann R., Malhan N. Single step synthesis of yttrium aluminum garnet ($Y_3Al_5O_{12}$) nanopowders by mixed fuel solution combustion approach // Ceramics International. 2011. V. 37. N 8. P. 3743–3746. https://doi.org/10.1016/j.ceramint.2011.05.052
13. Li J.G., Ikegami T., Lee J.-H., Mori T., Yajima Y. Co-precipitation synthesis and sintering of yttrium aluminum garnet (YAG) powders: the effect of precipitant // Journal of the European Ceramic Society. 2000. V. 20. N 14-15. P. 2395–2405. https://doi.org/10.1016/s0955-2219(00)00116-3

References

1. Chen L., Luo Y., Xia Y., Kang B., Yu S. Densification, microstructure and optical properties of YAG transparent ceramics prepared by dry-pressing and gelcasting. *Optical Materials*, 2021, vol. 121, pp. 111509. https://doi.org/10.1016/j.optmat.2021.111509
2. Li J., Liu J., Liu B., Liu W., Zeng Y., Ba X., Xie T., Jiang B., Liu Q., Pan Y., Feng X., Guo J. Influence of heat treatment of powder mixture on the microstructure and optical transmission of Nd:YAG transparent ceramics. *Journal of the European Ceramic Society*, 2014, vol. 34, no. 10, pp. 2497–2507. https://doi.org/10.1016/j.jeurceramsoc.2014.03.004
3. Wang H.M., Huang Z.Y., Jiang J.S., Liu K., Duan M.Y., Lu Z.W., Cedelle J., Guan Z.W., Lu T.C., Wang Q.Y. Unique mechanical properties of nano-grained YAG transparent ceramics compared with coarse-grained partners. *Materials & Design*, 2016, vol. 105, pp. 9–15. https://doi.org/10.1016/j.matdes.2016.04.094
4. Sim S.-M., Keller K.A., Mah T.-I. Phase formation in yttrium aluminum garnet powders synthesized by chemical methods. *Journal of Materials Science*, 2000, vol. 35, no. 3, pp. 713–717. https://doi.org/10.1023/A:1004709401795
5. Tachiwaki T., Yoshinaka M., Hirota K., Ikegami T., Yamaguchi O. Novel synthesis of $Y_3Al_5O_{12}$ (YAG) leading to transparent ceramics. *Solid State Communications*, 2001, vol. 119, no. 10-11, pp. 603–606. https://doi.org/10.1016/s0038-1098(01)00293-9
6. Sluzky E., Lemoine M., Hesse K. Phosphor development for α-silicon liquid crystal light valve projection display. *Journal of the Electrochemical Society*, 1994, vol. 141, no. 11, pp. 3172. https://doi.org/10.1149/1.2059297
7. Lu C.H., Huang C.H. Sensitized photoluminescence of Eu^{3+} and Gd^{3+}-doped $Y_3Al_5O_{12}$ phosphors prepared via a reverse microemulsion process. *Chemistry Letters*, 2004, vol. 33, no. 12, pp. 1568–1569. https://doi.org/10.1246/cl.2004.1568
8. Benayas A., del Rosal B., Pérez-Delgado A., Santacruz-Gómez K., Jaque D., Hirata G.A., Vetrone F. Nd:YAG near-infrared luminescent nanothermometers. *Advanced Optical Materials*, 2015, vol. 3, no. 5, pp. 687–694. https://doi.org/10.1002/adom.201400484
9. Evstropiev S.K., Demidov V.V., Sadovnichii R.V., Pchelkin G.A., Shurupov D.N., Matrosova A.S., Dukelskii K.V., Bulyga D.V., Nikonorov N.V., Podrukhin Y.F. YAG : R^{3+} (R = CE, DY, YB) nanophosphor-based luminescent fibre-optic sensors for temperature measurements in the range 20-500 C. *Quantum Electronics*, 2022, vol. 52, no. 1, pp. 94–99. https://doi.org/10.1070/QEL17971
10. Bulyga D.V., Evstropiev S.K. Intermediate products of Yb:YAG laser ceramics fabrication: structural features, morphology, and luminescent properties. *Research on Chemical Intermediates*, 2021, vol. 47, no. 8, pp. 3501–3514. https://doi.org/10.1007/s11164-021-04484-w
11. Saladino M.L., Nasillo G., Martino D.C., Caponetti E. Synthesis of Nd:YAG nanopowder using the citrate method with microwave irradiation. *Journal of Alloys and Compounds*, 2010, vol. 491, no. 1-2, pp. 737–741. https://doi.org/10.1016/j.jallcom.2009.11.054
12. Laishram K., Mann R., Malhan N. Single step synthesis of yttrium aluminum garnet ($Y_3Al_5O_{12}$) nanopowders by mixed fuel solution combustion approach. *Ceramics International*, 2011, vol. 37, no. 8, pp. 3743–3746. https://doi.org/10.1016/j.ceramint.2011.05.052
13. Li J.G., Ikegami T., Lee J.-H., Mori T., Yajima Y. Co-precipitation synthesis and sintering of yttrium aluminum garnet (YAG) powders: the effect of precipitant. *Journal of the European Ceramic Society*, 2000, vol. 20, no. 14-15, pp. 2395–2405. https://doi.org/10.1016/s0955-2219(00)00116-3

14. Ikesue A., Kinoshita T., Kamata K., Yoshida K. Fabrication and optical properties of high-performance polycrystalline Nd:YAG ceramics for solid=state lasers // Journal of the American Ceramic Society. 1995. V. 78. N 4. P. 1033–1040. https://doi.org/10.1111/j.1151-2916.1995.tb08433.x
15. Li X., Liu H., Wang J., Cui H., Han F., Boughton R.I. Production of nanosized YAG powders with spherical morphology and nonaggregation via a solvothermal method // Journal of the American Ceramic Society. 2004. V. 87. N 12. P. 2288–2290. https://doi.org/10.1111/j.1151-2916.2004.tb07507.x
16. Caponetti E., Martino D.C., Saladino M.L., Leonelli C. Preparation of Nd:YAG nanopowder in a confined environment // Langmuir. 2007. V. 23. N 7. P. 3947–3952. https://doi.org/10.1021/la0625906
17. Moussaoui A., Bulyga D.V., Evstropiev S.K., Ignatiev A.I., Nikonorov N.V., Podruhin Y.F., Sadovnichii R.V. Modified Pechini method by PVP addition for Nd:Gd_2O_3 nanophosphors fabrication // Ceramics International. 2021. V. 47. N 24. P. 34307–34313. https://doi.org/10.1016/j.ceramint.2021.08.341
18. Волкова Н.А., Евстропьев С.К., Никоноров Н.В., Евстропьев К.С. Особенности взаимодействия в водных растворах молекул поливинилпирролидона с ионами цинка и серебра по данным ИК спектроскопии // Оптика и спектроскопия. 2019. Т. 127. № 10. С. 687–690. https://doi.org/10.21883/OS.2019.10.48377.128-19
19. Guerbous L., Boukerika A. Nanomaterial host bands effect on the photoluminescence properties of Ce-doped YAG nanophosphor synthesized by sol-gel method // Journal of Nanomaterials. 2015. V. 2015. P. 617130. https://doi.org/10.1155/2015/617130
20. Ji X., Kang B., Deng J., Huang H., Wang X. Thermal decomposition and evolved gas analysis of neodymium-doped yttrium aluminum garnet precursor prepared by co-precipitation // Thermochimica Acta. 2013. V. 552. P. 23–27. https://doi.org/10.1016/j.tca.2012.11.021
21. AitMellal O., Oufni L., Messous M.Y., Tahri M., Neatu Ş., Florea M., Neatu F., Secu M. Structural properties and near-infrared light from Ce^{3+}/Nd^{3+}-co-doped $LaPO_4$ nanophosphors for solar cell applications // Journal of Materials Science: Materials in Electronics. 2022. V. 33. N 7. P. 4197–4210. https://doi.org/10.1007/s10854-021-07615-6
22. Boyer D., Bertrand-Chadeyron G., Mahiou R. Structural and optical characterizations of YAG:Eu^{3+} elaborated by the sol–gel process // Optical Materials. 2004. V. 26. N 2. P. 101–105. https://doi.org/10.1016/j.optmat.2003.11.005
23. Xia G., Zhou S., Zhang J., Xu J. Structural and optical properties of YAG:Ce^{3+} phosphors by sol–gel combustion method // Journal of Crystal Growth. 2005. V. 279. N 3-4. P. 357–362. https://doi.org/10.1016/j.jcrysgro.2005.01.072
24. Miller F.A., Wilkins C.H. Infrared spectra and characteristic frequencies of inorganic ions // Analytical Chemistry. 1952. V. 24. N 8. P. 1253–1294. https://doi.org/10.1021/ac60068a007
25. He X., Liu X., Li R., Yang B., Yu K., Zeng M., Yu R. Effects of local structure of Ce^{3+} ions on luminescent properties of $Y_3Al_5O_{12}$:Ce nanoparticles // Scientific Reports. 2016. V. 6. P. 22238. https://doi.org/10.1038/srep22238
26. Yang H., Lee D.K., Kim Y.S. Spectral variations of nano-sized $Y_3Al_5O_{12}$:Ce phosphors via codoping/substitution and their white LED characteristics // Materials Chemistry and Physics. 2009. V. 114. N 2-3. P. 665–669. https://doi.org/10.1016/j.matchemphys.2008.10.019
27. Klug H.P., Alexander L.E. Quantitative analysis of powder mixtures // X-ray Diffraction Procedures. 1954.
28. Shannon R.D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides // Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography. 1976. V. 32. N 5. P. 751–767. https://doi.org/10.1107/s0567739476001551
29. Zhu Q.Q., Li S., Yuan Q., Zhang H., Wang L. Transparent YAG:Ce ceramic with designed low light scattering for high-power blue LED and LD applications // Journal of the European Ceramic Society. 2021. V. 41. N 1. P. 735–740. https://doi.org/10.1016/j.jeurceramsoc.2020.09.006
30. Zhydachevskii Y., Syvorotka I.I., Vasylechko L., Sugak D., Borshchyshyn I.D., Luchechko A.P., Vakhula Ya.I., Ubizskii S.B., Vakiv M.M., Suchocki A. Crystal structure and luminescent properties of nanocrystalline YAG and YAG:Nd synthesized by sol–gel method // Optical Materials. 2012. V. 34. N 12. P. 1984–1989. https://doi.org/10.1016/j.optmat.2011.12.023
31. Hassanzadeh-Tabrizi S.A. Synthesis and luminescence properties of YAG:Ce nanopowder prepared by the Pechini method // Advanced

14. Ikesue A., Kinoshita T., Kamata K., Yoshida K. Fabrication and optical properties of high-performance polycrystalline Nd:YAG ceramics for solid-state lasers. *Journal of the American Ceramic Society*, 1995, vol. 78, no. 4, pp. 1033–1040. https://doi.org/10.1111/j.1151-2916.1995.tb08433.x
15. Li X., Liu H., Wang J., Cui H., Han F., Boughton R.I. Production of nanosized YAG powders with spherical morphology and nonaggregation via a solvothermal method. *Journal of the American Ceramic Society*, 2004, vol. 87, no. 12, pp. 2288–2290. https://doi.org/10.1111/j.1151-2916.2004.tb07507.x
16. Caponetti E., Martino D.C., Saladino M.L., Leonelli C. Preparation of Nd:YAG nanopowder in a confined environment. *Langmuir*, 2007, vol. 23, no. 7, pp. 3947–3952. https://doi.org/10.1021/la0625906
17. Moussaoui A., Bulyga D.V., Evstropiev S.K., Ignatiev A.I., Nikonorov N.V., Podruhin Y.F., Sadovnichii R.V. Modified Pechini method by PVP addition for Nd:Gd_2O_3 nanophosphors fabrication. *Ceramics International*, 2021, vol. 47, no. 24, pp. 34307–34313. https://doi.org/10.1016/j.ceramint.2021.08.341
18. Volkova N.A., Evsropiev S.K., Nikonorov N.V., Evstropyev K.S. Specific Features of interactions of polyvinylpyrrolidone molecules with zinc and silver ions in aqueous solutions according to IR spectroscopy data. *Optics and Spectroscopy*, 2019, vol. 127, no. 4, pp. 738–741. https://doi.org/10.1134/s0030400x19100308
19. Guerbous L., Boukerika A. Nanomaterial host bands effect on the photoluminescence properties of Ce-doped YAG nanophosphor synthesized by sol-gel method. *Journal of Nanomaterials*, 2015, vol. 2015, pp. 617130. https://doi.org/10.1155/2015/617130
20. Ji X., Kang B., Deng J., Huang H., Wang X. Thermal decomposition and evolved gas analysis of neodymium-doped yttrium aluminum garnet precursor prepared by co-precipitation. *Thermochimica Acta*, 2013, vol. 552, pp. 23–27. https://doi.org/10.1016/j.tca.2012.11.021
21. AitMellal O., Oufni L., Messous M.Y., Tahri M., Neatu Ş., Florea M., Neatu F., Secu M. Structural properties and near-infrared light from Ce^{3+}/Nd^{3+}-co-doped $LaPO_4$ nanophosphors for solar cell applications. *Journal of Materials Science: Materials in Electronics*, 2022, vol. 33, no. 7, pp. 4197–4210. https://doi.org/10.1007/s10854-021-07615-6
22. Boyer D., Bertrand-Chadeyron G., Mahiou R. Structural and optical characterizations of YAG:Eu^{3+} elaborated by the sol–gel process. *Optical Materials*, 2004, vol. 26, no. 2, pp. 101–105. https://doi.org/10.1016/j.optmat.2003.11.005
23. Xia G., Zhou S., Zhang J., Xu J. Structural and optical properties of YAG:Ce^{3+} phosphors by sol–gel combustion method. *Journal of Crystal Growth*, 2005, vol. 279, no. 3-4, pp. 357–362. https://doi.org/10.1016/j.jcrysgro.2005.01.072
24. Miller F.A., Wilkins C.H. Infrared spectra and characteristic frequencies of inorganic ions. *Analytical Chemistry*, 1952, vol. 24, no. 8, pp. 1253–1294. https://doi.org/10.1021/ac60068a007
25. He X., Liu X., Li R., Yang B., Yu K., Zeng M., Yu R. Effects of local structure of Ce^{3+} ions on luminescent properties of $Y_3Al_5O_{12}$:Ce nanoparticles. *Scientific Reports*, 2016, vol. 6, pp. 22238. https://doi.org/10.1038/srep22238
26. Yang H., Lee D.K., Kim Y.S. Spectral variations of nano-sized $Y_3Al_5O_{12}$:Ce phosphors via codoping/substitution and their white LED characteristics. *Materials Chemistry and Physics*, 2009, vol. 114, no. 2-3, pp. 665–669. https://doi.org/10.1016/j.matchemphys.2008.10.019
27. Klug H.P., Alexander L.E. Quantitative analysis of powder mixtures. *X-ray Diffraction Procedures*, 1954.
28. Shannon R.D. Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Crystallographica Section A: Crystal Physics, Diffraction, Theoretical and General Crystallography*, 1976, vol. 32, no. 5, pp. 751–767. https://doi.org/10.1107/s0567739476001551
29. Zhu Q.Q., Li S., Yuan Q., Zhang H., Wang L. Transparent YAG:Ce ceramic with designed low light scattering for high-power blue LED and LD applications. *Journal of the European Ceramic Society*, 2021, vol. 41, no. 1, pp. 735–740. https://doi.org/10.1016/j.jeurceramsoc.2020.09.006
30. Zhydachevskii Y., Syvorotka I.I., Vasylechko L., Sugak D., Borshchyshyn I.D., Luchechko A.P., Vakhula Ya.I., Ubizskii S.B., Vakiv M.M., Suchocki A. Crystal structure and luminescent properties of nanocrystalline YAG and YAG:Nd synthesized by sol–gel method. *Optical Materials*, 2012, vol. 34, no. 12, pp. 1984–1989. https://doi.org/10.1016/j.optmat.2011.12.023
31. Hassanzadeh-Tabrizi S.A. Synthesis and luminescence properties of YAG:Ce nanopowder prepared by the Pechini method. *Advanced*

Powder Technology. 2012. V. 23. N 3. P. 324–327. https://doi.org/10.1016/j.apt.2011.04.006
32. Yu S., Jing W., Tang M., Xu T., Yin W., Kang B. Fabrication of Nd:YAG transparent ceramics using powders synthesized by citrate sol-gel method // Journal of Alloys and Compounds. 2019. V. 772. P. 751–759. https://doi.org/10.1016/j.jallcom.2018.09.184

Powder Technology, 2012, vol. 23, no. 3, pp. 324–327. https://doi.org/10.1016/j.apt.2011.04.006
32. Yu S., Jing W., Tang M., Xu T., Yin W., Kang B. Fabrication of Nd:YAG transparent ceramics using powders synthesized by citrate sol-gel method. *Journal of Alloys and Compounds*, 2019, vol. 772, pp. 751–759. https://doi.org/10.1016/j.jallcom.2018.09.184

Авторы

Муссауи Амир — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0002-7858-7417, am.moussaoui92@gmail.com
Булыга Дмитрий Владимирович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация; младший научный сотрудник, Акционерное общество «Научно-производственное объединение Государственный оптический институт им. С.И. Вавилова», Санкт-Петербург, 192171, Российская Федерация, sc 57217158694, https://orcid.org/0000-0001-8861-2527, dmbulyga@yandex.ru
Игнатьев Александр Иванович — ведущий инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57189902013, https://orcid.org/0000-0002-6114-0248, alexandre_ignatiev@yahoo.com
Евстропьев Сергей Константинович — доктор химических наук, ведущий научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация; начальник отдела, Акционерное общество «Научно-производственное объединение Государственный оптический институт им. С.И. Вавилова», Санкт-Петербург, 192171, Российская Федерация; профессор, Санкт-Петербургский государственный технологический институт (Технический университет), Санкт-Петербург, 190013, Российская Федерация, sc 6507317768, https://orcid.org/0000-0002-0160-8443, evstropiev@bk.ru
Никоноров Николай Валентинович — доктор физико-математических наук, профессор, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 7003772604, https://orcid.org/0000-0002-1341-067X, Nikonorov@oi.ifmo.ru

Authors

Amir Moussaoui — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0002-7858-7417, am.moussaoui92@gmail.com
Dmitry V. Bulyga — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation; Senior Researcher, JVC “RPA Vavilov State Optical Institute”, Saint Petersburg, 192171, Russian Federation, sc 57217158694, https://orcid.org/0000-0001-8861-2527, dmbulyga@yandex.ru

Aleksandr I. Ignatiev — Leading Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 57189902013, https://orcid.org/0000-0002-6114-0248, alexandre_ignatiev@yahoo.com

Sergey K. Evstropiev — D.Sc. (Chemistry), Leading Researcher, ITMO University, Saint Petersburg, 197101, Russian Federation; Head of Department, JVC “RPA Vavilov State Optical Institute”, Saint Petersburg, 192171, Russian Federation; Professor, Saint Petersburg State Institute of Technology, Saint Petersburg, 190013, Russian Federation, sc 6507317768, https://orcid.org/0000-0002-0160-8443, evstropiev@bk.ru

Nikolay V. Nikonorov — D.Sc. (Physics & Mathematics), Full Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 7003772604, https://orcid.org/0000-0002-1341-067X, Nikonorov@oi.ifmo.ru

Статья поступила в редакцию 28.11.2023
Одобрена после рецензирования 11.12.2023
Принята к печати 13.01.2024

Received 28.11.2023
Approved after reviewing 11.12.2023
Accepted 13.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-11-19
УДК 528.72: 514.88: 004.02

Расчетное прогнозирование в задаче идентификации стереоизображений

Марина Витальевна Самойленко[1]✉, Владимир Александрович Хачикян[2]

[1] Независимый исследователь, Москва, Российская Федерация
[2] Московский авиационный институт (национальный исследовательский университет), Москва, 125993, Российская Федерация

[1] samoi.mar@mail.ru✉, https://orcid.org/0000-0002-4798-2414
[2] zaktol@yandex.ru, https://orcid.org/0009-0006-1279-7661

Аннотация

Введение. Рассмотрены вопросы повышения эффективности и достоверности идентификации стереоизображений за счет расчетного прогнозирования положения и размеров зоны неопределенности, в которой заведомо находится искомая точка соответствия. **Метод.** На одном из стереоизображений выбирается контрольная точка, для которой требуется найти точку соответствия на втором стереоизображении. По известным параметрам стереоскопической телевизионной системы и координатам контрольной точки с помощью предложенного в работе математического аппарата рассчитываются координаты границ зоны неопределенности на втором стереоизображении. Вторая точка соответствия находится с использованием поисковой процедуры путем сопоставления одинаковых малых областей с центрами в контрольной точке на первом стереоизображении и в точках зоны неопределенности на втором. Сопоставление производится по критерию минимума квадратического рассогласования интенсивностей. Необходимой априорной информацией для реализации метода являются предельные высоты отображаемого на стереоизображениях рельефа. **Основные результаты.** Получены соотношения линейных размеров на плоском рельефе и на изображении, формируемом по принципу центрального проецирования. Выведены соотношения, позволяющие расчетным путем получить координаты точек соответствия и стереоскопического рассогласования для стереоизображений плоского рельефа. Для стереоизображений объемного рельефа приведены расчетные формулы определения границ зоны неопределенности на втором стереоизображении, в пределах которой осуществляется поиск точки соответствия. Корректность и работоспособность полученных соотношений подтверждены компьютерным моделированием. **Обсуждение.** Ограничение размеров области поиска посредством расчетного прогнозирования зоны неопределенности позволяет уменьшить вычислительные и временные затраты на поисковую процедуру. За счет этого повышается эффективность идентификации точек стереоизображений и уменьшается вероятность ложной идентификации.

Ключевые слова

стереоизображения, идентификация, зона неопределенности, точки соответствия, прогнозирование

Ссылка для цитирования: Самойленко М.В., Хачикян В.А. Расчетное прогнозирование в задаче идентификации стереоизображений // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 11–19. doi: 10.17586/2226-1494-2024-24-1-11-19

Computational prediction in the problem of stereo image identification

Marina V. Samoilenko[1]✉, Vladimir A. Hachikian[2]

[1] Independent Researcher, Moscow, Russian Federation
[2] Moscow Aviation Institute (National Research University), Moscow, 125993, Russian Federation

[1] samoi.mar@mail.ru✉, https://orcid.org/0000-0002-4798-2414
[2] zaktol@yandex.ru, https://orcid.org/0009-0006-1279-7661

Abstract

The paper examines the issues of increasing the efficiency and reliability of stereo image identification through computational prediction of the position and size of the uncertainty zone in which the desired correspondence point

© Самойленко М.В., Хачикян В.А., 2024

is known to be located. A control point is selected on one of the stereo images, for which it is necessary to find a correspondence point on the second stereo image. Based on the known parameters of the stereoscopic television system and the coordinates of the control point, using the mathematical apparatus proposed in the work, the coordinates of the boundaries of the uncertainty zone on the second stereo image are calculated. The second point of correspondence is found by the search procedure by comparing identical small areas with centers in the control point on the first stereo image and in the points of the uncertainty zone on the second; the comparison is made according to the criterion of minimum quadratic mismatch of intensities. The necessary a priori information for implementing the method is the maximum heights of the relief displayed on stereo images. The ratios of linear dimensions on a flat relief and on an image formed according to the principle of central projection were obtained. Relationships have been obtained that make it possible to obtain, by calculation, the coordinates of the correspondence points and the stereoscopic mismatch for stereo images of a flat relief. For stereo images of a volumetric relief, calculation formulas are obtained for determining the boundaries of the zone of uncertainty in the second stereo image within which the search for the point of correspondence is carried out. The correctness and performance of the obtained relationships are confirmed by computer modeling. Limiting the size of the search area by means of calculated prediction of the uncertainty zone makes it possible to reduce the computational and time costs of the search procedure. Due to this, the efficiency of identifying stereo image points increases and the likelihood of false identification decreases.

Keywords

stereo images, identification, uncertainty zone, correspondence points, prognostication

For citation: Samoilenko M.V., Hachikian V.A. Computational prediction in the problem of stereo image identification. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 11–19 (in Russian). doi: 10.17586/2226-1494-2024-24-1-11-19

Введение

Обработка стереоизображений имеет самые разные применения и, соответственно, различается по способам обработки [1, 2]. При решении целевых задач используется специфика тех или иных объектов, на которые ориентировано применение стереосистем. Однако в любом случае применение стереосистем и стереоизображений приводит к получению пространственной информации, будь то расположение городских зданий [3], контроль с беспилотных летательных аппаратов состояния изоляторов энергосистем [4] или распознавание и локализация движущихся целей в помещении [5]. Для решения любых задач с применением стереосистем необходима идентификация стереоизображений, т. е. выявление на них геометрических или точечных соответствий.

Идентификация стереоизображений включает множество подходов — от работы оператора со стереоизображениями в интерактивном режиме до полностью автоматических систем. В настоящее время широкое распространение получили методы глубокого обучения [3, 4, 6] и другие вариации применения нейронных сетей [7]. Однако успешность этих широко разрабатываемых методов определяется обучающими данными, которые закономерно связаны с конкретным применением систем, от чего зависят используемые в обучении модели. При этом, чем более обобщенно сформулирована задача для обучаемой нейронной сети, тем будет сложнее сеть и тем больший объем памяти потребуется для хранения необходимых параметров, как отмечается, например, в [8].

Другой подход связан с процедурой выявления на стереоизображениях признаков, характерных для анализируемых сцен [9–12]. Собственно, на этом принципе базируется и применение нейронных систем. При этом делаются попытки, например, в [13], перехода к таким принципам идентификации стереоизображений, при которых решение не ограничивалось бы конкретными структурами данных.

В [14] отмечено, что при разработке методов идентификации стереоизображений важным фактором является обеспечение баланса между точностью и эффективностью. Желательна и универсальность алгоритма, его независимость от содержания и структуры наблюдаемой сцены. В этом плане актуальной является поточечная идентификация изображений, когда сопоставляются не структурные фрагменты, а точки стереоизображений [15–19]. Однако применение этого подхода ограничено из-за большого объема вычислений при идентификации каждой точки. В настоящей работе предложено решение, позволяющее существенно, на порядок, сократить вычислительные процедуры поиска, для чего разработан метод расчетного прогнозирования *зоны неопределенности*, в которой осуществляется поиск.

Первые методы поточечной идентификации стереоизображений, применяемые и в настоящее время — *площадные методы* [20, 21]. Данные методы реализуются следующим образом: на одном из стереоизображений выбирается контрольная точка соответствия и вблизи нее выделяется окно идентификации; на втором стереоизображении таким же по форме и размерам «окном» сканируется зона неопределенности, в которой ищется вторая точка соответствия. В настоящей работе рассматривается площадной метод идентификации стереоизображений.

Важными параметрами для площадных методов являются окна идентификации и зона неопределенности. От них зависит как точность, так и вычислительные затраты и время идентификации. В [14] отмечается, что выбор габаритов «окна» происходит «эвристически, уменьшаясь на более информативных участках и увеличиваясь на менее информативных». Очевидно, так же эвристически выбирается и зона неопределенности.

В данной работе предложен аналитический, расчетный подход к определению размеров и локализации зоны неопределенности, в пределах которой на втором стереоизображении осуществляется поиск точки соответствия.

В работе рассматривается площадной метод идентификации, однако это не ограничивает значимость разработанного математического аппарата, поскольку прогнозирование размеров и локализации зоны неопределенности полезны при любом подходе к идентификации стереоизображений, будь то фрагментация или применение нейронных сетей.

Цель работы — повышение достоверности и эффективности идентификации стереоизображений за счет сокращения вычислительных и временных затрат.

Для достижения поставленной цели разработан математический аппарат, позволяющий расчетным путем прогнозировать параметры идентификации стереоизображений, включая положение границ зоны неопределенности, в которой расположена вторая точка соответствия. Расчетное уточнение границ зоны неопределенности приводит к сокращению как временных, так и вычислительных затрат на поисковую процедуру. Кроме этого, минимизируется вероятность ложной идентификации, под которой понимается ошибочное определение положения точки соответствия на втором изображении. Последнее преимущество показано в компьютерных экспериментах.

В основе работы лежат идеи и разработки доктора технических наук профессора В.И. Самойленко.

Постановка задачи

Рассмотрим стереоскопическую телевизионную систему (СТС), состоящую из двух камер. Для определенности изображение левой камеры обозначим как изображение 1, а правой — как изображение 2. Назовем точку соответствия на изображении 1 и «окно» вблизи этой точки *контрольными*. Контрольную точку соответствия обозначим П1. На изображении 2 *сканирующим* «окном» осуществляется поисковая процедура для обнаружения второй точки соответствия П2.

Показателем качества при поиске выберем квадратическое отклонение

$$\eta(\boldsymbol{\rho}_{o2}) = (\mathbf{f}_1(\boldsymbol{\rho}_{П1}) - \mathbf{f}_2(\boldsymbol{\rho}_{o2}))^T(\mathbf{f}_1(\boldsymbol{\rho}_{П1}) - \mathbf{f}_2(\boldsymbol{\rho}_{o2})), \quad (1)$$

где $\mathbf{f}_1$ — вектор интенсивности контрольного «окна»; $\boldsymbol{\rho}_{П1}$ — координаты контрольной точки соответствия; $\mathbf{f}_2(\boldsymbol{\rho}_{o2})$ — вектор интенсивности сканирующего «окна» при положении его центра $\boldsymbol{\rho}_{o2}$; векторы «окон» получим разверткой их изображений.

При точном положении сканирующего «окна» центр его совпадет с истинным положением второй точки соответствия: $\boldsymbol{\rho}_{o2} = \boldsymbol{\rho}_{П2}$. В результате поисковой процедуры по критерию минимума (1) получим оценку истинного положения второй точки соответствия $\widehat{\boldsymbol{\rho}}_{П2}$.

Изображения левого и правого «окон» в общем случае не совпадают ни при каком положении сканирующего «окна» $\boldsymbol{\rho}_{o2}$ вследствие стереоскопичности. Тем не менее, в идеальном случае в результате поисковой процедуры получим $\widehat{\boldsymbol{\rho}}_{П2} = \boldsymbol{\rho}_{П2}$, а в неидеальном случае $\widehat{\boldsymbol{\rho}}_{П2} \neq \boldsymbol{\rho}_{П2}$ и возникает ошибка идентификации

$$\Delta_{П2} = \|\widehat{\boldsymbol{\rho}}_{П2} - \boldsymbol{\rho}_{П2}\|. \quad (2)$$

Ставится задача разработать математический аппарат, позволяющий расчетным путем прогнозировать параметры идентификации стереоизображений: положение второй точки соответствия $\boldsymbol{\rho}_{П2}$ (для плоского рельефа); стереоскопическое рассогласование $\boldsymbol{\delta} = \boldsymbol{\rho}_{П2} - \boldsymbol{\rho}_{П1}$ и зону неопределенности на изображении 2, в которой находится точка соответствия (для объемного рельефа).

Соотношения размеров базовой плоскости рельефа и изображения

Найдем соотношения, описывающие соответствие горизонтального $\varepsilon_г$ и вертикального $\varepsilon_в$ размеров на изображении и соответствующих им размеров на опорной плоскости рельефа, которая совпадает с горизонтальной плоскостью базовой системы координат.

В соответствии с рис. 1 найдем соотношение $\varepsilon_г$ на изображении и соответствующего размера отрезка a на опорной плоскости.

С учетом малости $\varepsilon_г$, получим

$$\frac{a}{\varepsilon_г} = \frac{\|\mathbf{r}_P - \mathbf{r}_К\|}{\|\mathbf{r}_o + \mathbf{r}_П\|}, \quad (3)$$

где $\mathbf{r}_К$ — координатный вектор центра проецирования камеры; $\mathbf{r}_o$ — оптический вектор, направление которого совпадает с направлением оптической оси камеры, а длина определяется фокусным расстоянием; $\mathbf{r}_P$ — координатный вектор порождающей точки P; $\mathbf{r}_П$ — координатный вектор точки П (отображения порождающей точки).

Найдем аналогичное соотношение для вертикального отрезка $\varepsilon_в$ на плоскости изображения и соответствующего ему отрезка b на опорной плоскости (рис. 2).

Из рис. 2 с учетом малости $\varepsilon_в$ найдем отношение $\frac{t}{\varepsilon_в} = \frac{\|\mathbf{r}_P - \mathbf{r}_К\|}{\|\mathbf{r}_o + \mathbf{r}_П\|}$, где отрезок t параллелен плоскости изо-

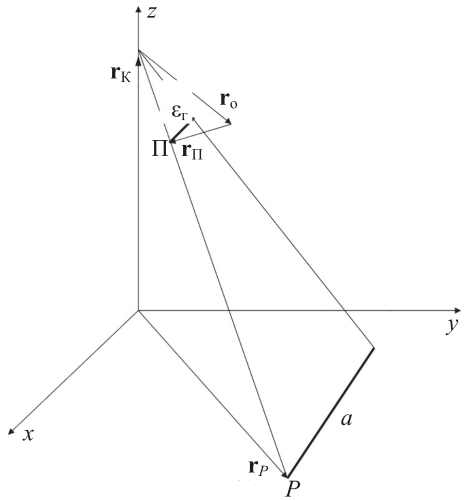


Рис. 1. К определению соотношения размеров горизонтального отрезка на изображении $\varepsilon_г$ и соответствующего отрезка a на опорной плоскости

Fig. 1. For the ratio determination of the horizontal segment dimensions in the image $\varepsilon_г$ and the corresponding segment a on the reference plane

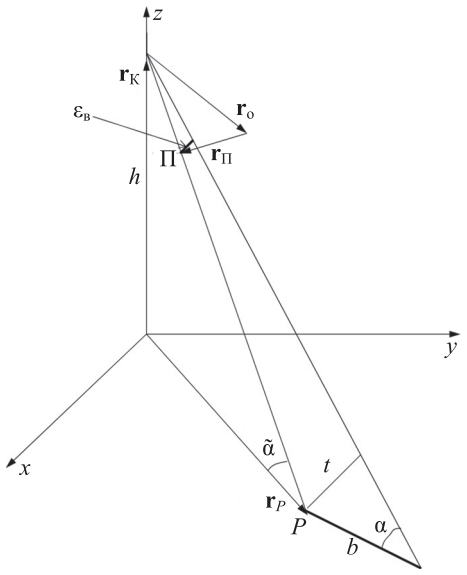


Рис. 2. К определению соотношения размеров вертикального отрезка на изображении $\varepsilon_в$ и соответствующего отрезка на опорной плоскости b

Fig. 2. For the ratio determination of the vertical segment dimensions in the image $\varepsilon_в$ and the corresponding segment on the reference plane b

бражения. Из этого же рисунка явствует, что $b \approx \frac{t}{\sin\alpha} \approx \approx \frac{t}{\sin\tilde{\alpha}} = \frac{t\|\mathbf{r}_P - \mathbf{r}_К\|}{h}$, где h — высота камеры над опорной плоскостью.

Из полученных соотношений находим:

$$\frac{b}{\varepsilon_в} = \frac{\|\mathbf{r}_P - \mathbf{r}_К\|^2}{h\|\mathbf{r}_о + \mathbf{r}_П\|}. \quad (4)$$

Отметим, что выражения (3) и (4) неудобны тем, что, помимо параметров СТС, требуют знания наклонной дальности порождающей точки $\|\mathbf{r}_P - \mathbf{r}_К\|$. Сделаем эти формулы более удобными, избавившись от $\|\mathbf{r}_P - \mathbf{r}_К\|$.

С помощью несложных геометрических построений можно показать, что

$$\|\mathbf{r}_P - \mathbf{r}_К\| = \frac{h\|\mathbf{r}_о + \mathbf{r}_П\|}{|r_{оz} + r_{Пz}|}, \quad (5)$$

где $r_{оz}$ и $r_{Пz}$ — z-компоненты соответствующих векторов.

В результате подстановки (5) в выражения (3) и (4) получим

$$\frac{a}{\varepsilon_г} = \frac{h}{|r_{оz} + r_{Пz}|}, \quad (6)$$

$$\frac{b}{\varepsilon_в} = \frac{h\|\mathbf{r}_о + \mathbf{r}_П\|}{(r_{оz} + r_{Пz})^2}. \quad (7)$$

Выражения (6) и (7) определяют соотношения размеров на опорной плоскости (a, b) и их отображений в плоскости изображения ($\varepsilon_г$, $\varepsilon_в$) только через параметры СТС, что делает удобным их использование.

Расчетное прогнозирование в задаче идентификации стереоизображений плоского рельефа

Приведем формулы, позволяющие расчетным путем определить стереоскопическое рассогласование и координаты точки соответствия на правом изображении по заданным координатам контрольной точки соответствия на левом изображении. Решим задачу для плоского рельефа, лежащего в горизонтальной плоскости базовой системы координат.

Точку на левом изображении, расположенную в центре изображения, назовем первой эталонной точкой соответствия:

$$\boldsymbol{\rho}_{1эт} = \boldsymbol{\rho}_C = \begin{bmatrix} x_{Cиз} \\ y_{Cиз} \end{bmatrix},$$

где $x_{Cиз}$, $y_{Cиз}$ — координаты центра изображения в системе координат изображения с началом в левом нижнем углу.

Заметим, что векторы $\boldsymbol{\rho}$ — двухмерные, обозначают координаты в системе координат изображения; векторы $\mathbf{r}$ — трехмерные, обозначают координаты в базовой системе координат. Преобразование векторов выполняется с помощью матрицы перехода [21] $\mathbf{M}$ по формулам $\boldsymbol{\rho} = \mathbf{M}\mathbf{r}$, $\mathbf{r} = \mathbf{M}^T\boldsymbol{\rho}$.

Координаты второй эталонной точки соответствия (на правом изображении) рассчитаем по формулам из работы [22]. Для этого вычислим координаты порождающей точки, отображающейся в центре изображения 1:

$$\mathbf{r}_P = r_{К1} + \xi\mathbf{r}_0, \quad (8)$$

$$\xi = -\frac{h}{r_{oz}}. \quad (9)$$

Координаты второй эталонной точки соответствия вычислим как отображение на правое изображение порождающей точки (8):

$$\mathbf{r}_{2эт} = \varsigma(\mathbf{r}_P - \mathbf{r}_{К2}) - \mathbf{r}_0, \quad (10)$$

где

$$\varsigma = \frac{\mathbf{r}_0^T\mathbf{r}_0}{\mathbf{r}_0^T\mathbf{r}_P - \mathbf{r}_0^T\mathbf{r}_{К2}}. \quad (11)$$

В системе координат изображения с началом в левом нижнем углу

$$\boldsymbol{\rho}_{2эт} = \mathbf{M}\mathbf{r}_{2эт} + \boldsymbol{\rho}_C,$$

$$\boldsymbol{\rho}_{2эт} = \begin{bmatrix} x_{2эт} \\ y_{2эт} \end{bmatrix}, \quad (12)$$

где $x_{2эт}$, $y_{2эт}$ — координаты второй эталонной точки соответствия на изображении.

Вектор координат второй эталонной точки (12) позволяет вычислить векторное стереоскопическое рассогласование эталонных точек соответствия

$$\boldsymbol{\delta}_0 = \boldsymbol{\rho}_{2эт} - \boldsymbol{\rho}_C = \begin{bmatrix} \delta_{0г} \\ \delta_{0в} \end{bmatrix}, \quad (13)$$

где $\delta_{0г}$ и $\delta_{0в}$ — горизонтальная и вертикальная составляющие стереоскопического рассогласования.

Рассчитаем эталонные линейные размеры на опорной плоскости, которые отображаются в центральной части изображения компонентами вектора (13) $\delta_{0г}$ и $\delta_{0в}$. Для этого преобразуем выражения (6) и (7), подставив в них $\mathbf{r}_{П} = 0$ и $r_{Пz} = 0$:

$$a_0 = \delta_{0г}\frac{h}{|r_{oz}|}, \qquad b_0 = \delta_{0в}\frac{h\|\mathbf{r}_o\|}{r_{oz}^2}. \tag{14}$$

В соответствии с принципом взаимности полагаем, что независимо от места на изображении, стереоскопическое рассогласование точек соответствия определяется одними и теми же эталонными размерами на рельефе (14). Тогда для любой заданной контрольной точки с координатами $\mathbf{r}_{П1}$, согласно (6) и (7), рассчитаем стереоскопическое рассогласование:

$$\delta_{г\ расч} = \frac{a_0|r_{oz} + r_{П1z}|}{h}, \qquad \delta_{в\ расч} = \frac{b_0(r_{oz} + r_{П1z})^2}{h\|\mathbf{r}_o + \mathbf{r}_{П1}\|}, \qquad \boldsymbol{\delta}_{расч} = \begin{bmatrix} \delta_{г\ расч} \\ \delta_{в\ расч} \end{bmatrix}. \tag{15}$$

Расчетные координаты второй точки соответствия найдем по формулам:

$$x_{П2\ расч} = \rho_{П1x} - \delta_{г\ расч}, \qquad y_{П2\ расч} = \rho_{П1y} - \delta_{в\ расч}, \tag{16}$$

где $\boldsymbol{\rho}_{П1} = \begin{bmatrix} \rho_{П1x} \\ \rho_{П1y} \end{bmatrix} = \mathbf{M}\mathbf{r}_{П1}$; $\mathbf{M}$ — матрица перехода.

Выражения (15) и (16) позволяют рассчитать стереоскопическое рассогласование и координаты второй точки соответствия при заданных координатах контрольной точки соответствия.

Для плоского рельефа представленный математический аппарат позволяет идентифицировать стереоизображения расчетным путем. Достаточно выбрать контрольные точки соответствия на изображении 1 и вычислить для них стереоскопическое рассогласование (15) и координаты второй точки соответствия (16). Отметим, что данный расчет может быть применен только для плоского рельефа при отсутствии ошибок измерений и юстировки СТС. Более естественным, однако, является разновысотный рельеф. В этом случае полученные соотношения позволяют расчетным путем определить зону неопределенности на правом изображении, в которой заведомо находится вторая точка соответствия.

Расчетное прогнозирование в задаче идентификации стереоизображений объемного рельефа

Приведем решение задачи расчетного прогнозирования зоны неопределенности на правом изображении, в которой находится вторая точка соответствия.

Задачу решаем для одномерного случая, что обусловлено следующим.

В работе [12] показано, что асимметрия интенсивности в пределах «окна» приводит к ошибке идентификации. Профессор В.И. Самойленко показал, что на конечный результат определения пространственных координат точек рельефа основное влияние оказывает асимметрия по дальности. Это значит, что чем меньше высота «окна», тем меньше влияние асимметрии интенсивности на точность восстановления рельефа. В этом плане приоритет имеет СТС с одинаковыми камерами, оптические оси которых ортогональны базе стереоскопичности и параллельны между собой, плоскости изображений ортогональны оптическим осям. В такой системе, при отсутствии ошибок юстировки, любая порождающая точка отображается на одну и ту же строку обоих стереоизображений. Рассмотрим такую СТС. Минимальную высоту «окна» примем равной одному элементу разрешения изображения; такое «окно» имеет размеры $w_x \times dy$, где w_x — ширина «окна»; dy — размер элемента разрешения изображения по вертикальной оси.

В результате задача становится одномерной: поиск точки соответствия на правом изображении осуществим сканированием по строке «окном», представляющим собой фрагмент этой строки.

Если опорная плоскость, в которой лежит плоский рельеф, не совпадает с горизонтальной плоскостью базовой системы координат, а параллельна ей и находится на высоте H, то в формулы (9), (14) и (15) вместо h подставим $(h - H)$. Это приведет к изменению координат искомой точки соответствия (16). С учетом принятой конструкции СТС требует определения только абсциссы второй точки соответствия — $x_{2\ расч}$; ордината этой точки та же, что и у контрольной точки соответствия. В зависимости от значения H положение второй точки соответствия на строке будет меняться: чем больше H, тем больше стереоскопическое рассогласование и тем левее расположена точка соответствия на правом изображении.

Пусть известны максимальная H_{max} и минимальная H_{min} высоты рельефа, отображаемого на стереоизображениях. Подставим вместо h в формулы (16) разность $(h - H_{max})$, найдем по первой формуле значение $x_{2\ расч}$, которое обозначим как $x_{2л}$. Аналогично, подставив H_{min}, найдем $x_{2пр}$. Переменные $x_{2л}$ и $x_{2пр}$ обозначают левую и правую границы зоны неопределенности. В этой зоне отображается порождающая точка контрольной точки соответствия, при условии, что она лежит в диапазоне высот рельефа $H_{min} - H_{max}$.

Описанное расчетное прогнозирование зоны неопределенности осуществляется по параметрам СТС и априорной информации о предельных высотах рельефа. Такое прогнозирование позволяет не только сократить время и вычислительные затраты на поисковую процедуру идентификации точек соответствия, но и уменьшить вероятность ложной идентификации за счет максимального сокращения области поиска.

Компьютерные эксперименты

С помощью компьютерных экспериментов продемонстрируем корректность и работоспособность разработанного математического аппарата, а также продемонстрируем возможность с помощью расчетного прогнозирования избежать ложной идентификации точек.

Параметры СТС в компьютерных экспериментах следующие: положение левой камеры в базовой системе координат $\mathbf{r}_{K1} = [7{,}5\ -5\ \ 5]^T$ м, положение правой камеры $\mathbf{r}_{K2} = [8\ -5\ \ 5]^T$ м. Угол наклона оптических осей камер по отношению к вертикали составляет 68°, размер изображений — 18 × 18 см^2, элемент разрешения изображения — 0,5 × 0,5 мм^2. Изображения оцифровываются 128-разрядными аналого-цифровыми преобразователями.

Рельеф представляет собой случайным образом расположенные на плоскости объекты (полусфера, конус и пирамида; радиус полусфер и высоты конусов и пирамид моделируются случайным образом в диапазоне 0,2–0,5 м). Максимальная высота рельефа составляет $H_{max} = 0{,}5$ м, а минимальная — $H_{min} = 0$. По всей поверхности рельефа распределена интенсивность, представляющая собой коррелированный шум.

В первом эксперименте зона неопределенности не прогнозировалась, а была принята из естественного понимания того, что на правом изображении точка соответствия лежит левее контрольной точки соответствия. Зона неопределенности в этом эксперименте лежит между $x_{2л} = 20{,}5$ мм и $x_{2пр} = 155$ мм; ширина ее составляет $d_{зн} = 134{,}5$ мм.

На рис. 3 показаны результаты идентификации точек соответствия. Точки соответствия являются центрами «окон» — контрольного на левом изображении и сканирующего на правом.

Координаты контрольной точки соответствия $\boldsymbol{\rho}_{П1} = (155\ \ 170)^T$ мм (центр контрольного «окна» на рис. 3, *a*). Координаты ее порождающей точки при объемном рельефе нельзя определить расчетным путем, как это делалось для плоского рельефа (см. раздел «Расчетное прогнозирование в задаче идентификации стереоизображений плоского рельефа») из-за неизвестной высоты порождающей точки над горизонтальной плоскостью. Эти координаты определены программным методом при формировании

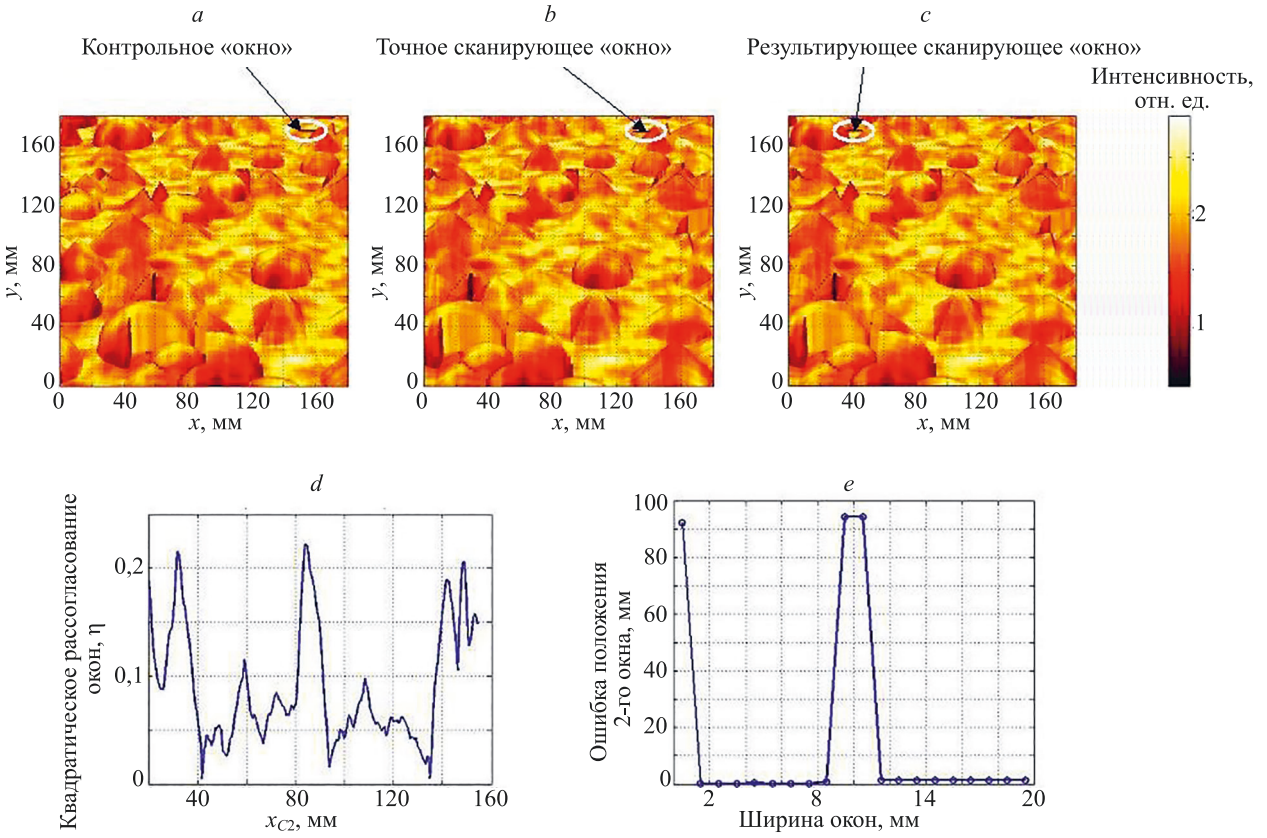


Рис. 3. Идентификация без использования расчетного прогнозирования: левое изображение с контрольным «окном» (*a*); правое изображение с точным положением сканирующего «окна» (*b*); правое изображение с результирующим положением сканирующего «окна» при ширине «окна» $w_x = 10{,}5$ мм (*c*); зависимость квадратического отклонения η от положения сканирующего «окна» при ширине «окна» $w_x = 10{,}5$ мм (*d*); зависимость ошибки идентификации $\Delta_{П2}$ от размера «окна» w_x (*e*)

Fig. 3. Identification without the use of computational prediction: left image with a control "window" (*a*); right image with the exact position of the scanning "window" (*b*); right image with the resulting position of the scanning "window" with the width of the "window" $w_x = 10.5$ mm (*c*); dependence of the square deviation η vs. the position of the scanning "window" with the width of the "window" $w_x = 10.5$ mm (*d*); dependence of the identification error $\Delta_{П2}$ vs. the size of the "window" w_x (*e*)

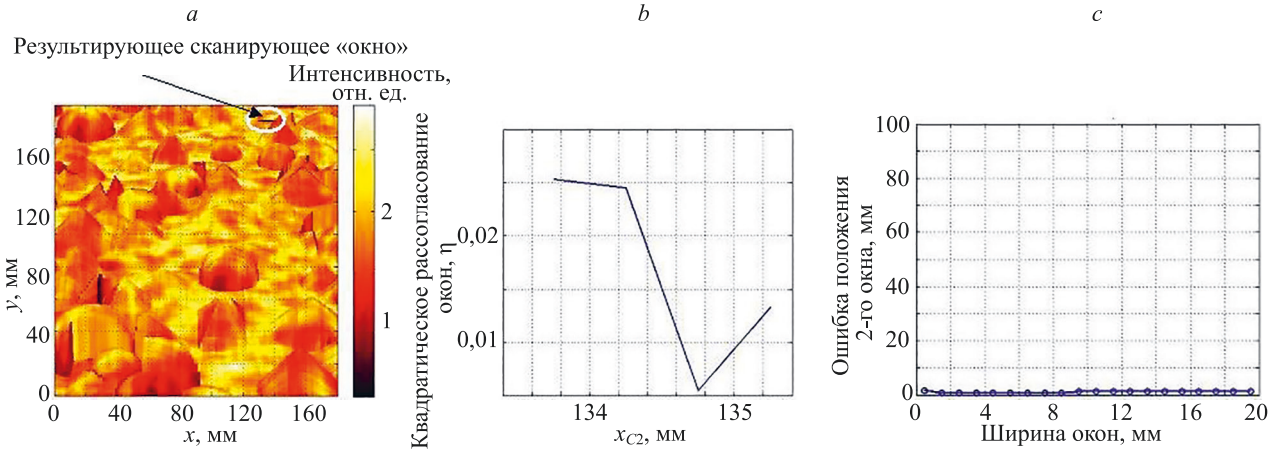


Рис. 4. Идентификация с расчетным прогнозированием зоны неопределенности: правое изображение с результирующим положением сканирующего «окна» (*a*); зависимость показателя η от положения сканирующего «окна» при ширине «окна» w_x = 10,5 мм (*b*); зависимость ошибки идентификации $\Delta_{П2}$ от размера «окна» w_x (*c*)

Fig. 4. Identification with calculated prediction of the uncertainty zone: right image with the resulting position of the scanning "window" with the width of the "window" (*a*); dependence of the indicator η vs. the position of the scanning "window" with the width of the "window" w_x = 10.5 mm (*b*); dependence of the identification error $\Delta_{П2}$ vs. the size of the "window" w_x (*c*)

изображений. По координатам порождающей точки рассчитано по формулам (10) и (11) точное положение второй точки соответствия, которое составило $\boldsymbol{\rho}_{П2\,точн}$ = = (135,9 170)^T мм (рис. 3, *b*). Сравнив с этим точным значением найденное результирующее положение сканирующего «окна», определим ошибку идентификации (2).

Результат поиска сканирующим «окном» на правом изображении зависит от ширины «окна» (рис. 3, *e*). На рис. 3, *e* видно, что при ширине «окна» w_x = 10,5 мм результат идентификации оказывается ложным: вместо $x_{П2\,точн}$ = 135,9 мм или близкого к этому значения, результирующее положение сканирующего «окна» определилось как $\widehat{x}_{П2}$ = 41,8 мм (рис. 3, *c*). Такой ошибочный результат получен в соответствии с минимумом зависимости квадратического отклонения от положения сканирующего «окна» (рис. 3, *d*). Эта зависимость имеет два наименьших локальных минимума — ложный слева и правильный справа. Ложный локальный минимум оказался меньше правильного, что привело к ложной идентификации.

Во втором эксперименте (рис. 4) зона неопределенности прогнозировалась с использованием предложенного в работе математического аппарата. Результаты, вычисленные по первой формуле (16) для H_{max} и H_{min}, получены следующие: $x_{2л}$ = 134,1 мм, $x_{2пр}$ = 136,2 мм. Ширина расчетной зоны неопределенности, таким образом, составила $d_{зн\,расч}$ = 2,1 мм, что более чем в 60 раз меньше, чем $d_{зн}$ в первом эксперименте.

Результирующее положение сканирующего «окна» на изображении составило 134,8 мм (рис. 4, *a*), что близко к точному положению второй точки соответствия; это наглядно видно из сравнения с рис. 3, *b*.

Такая успешная идентификация объясняется следующим. В результате сужения зоны неопределенности более чем в 60 раз зависимость квадратического отклонения от положения сканирующего «окна» имеет только один минимум, соответствующий точному положению второй точки соответствия (рис. 4, *b*). Ложный минимум находится далеко за пределами расчетной зоны неопределенности. Ошибка определения положения второй точки соответствия незначительна при всех размерах «окна» (рис. 4, *c*), в том числе и при размере «окна» w_x = 10,5 мм, при котором в первом эксперименте получилась ложная идентификация.

Предложенный математический аппарат чувствителен к положению контрольной точки соответствия: чем выше она на изображении, тем меньше ширина зоны неопределенности (рис. 5).

И, конечно, расчетная зона неопределенности зависит от максимальной высоты рельефа, что видно на рис. 5: чем рельеф выше, тем больше зона неопределенности. Минимальная высота рельефа остается неизменной и равной нулю для всех представленных графиков.

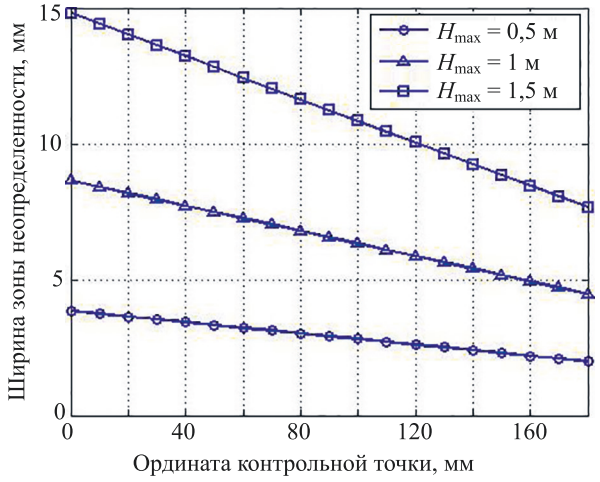


Рис. 5. Зависимости размера зоны неопределенности от положения контрольной точки соответствия при разных значениях H_{max}

Fig. 5. Dependence of the size of the uncertainty zone vs. the control correspondence point position for different values of H_{max}

Заключение

Разработанный математический аппарат позволяет априори, расчетным путем, ограничить зону неопределенности минимальным интервалом, в котором заведомо находится точка соответствия на правом изображении. Такая возможность повышает эффективность идентификации за счет сокращения времени и вычислительных затрат на поисковую процедуру. Кроме этого, расчетное прогнозирование зоны неопределенности существенно уменьшает вероятность ложной идентификации, что наглядно продемонстрировано в компьютерных экспериментах.

Расчетное прогнозирование зоны неопределенности осуществлено по параметрам стереоскопической телевизионной системы, которые известны, и априорной информации о предельных высотах рельефа. Отметим, что зона неопределенности имеет разную ширину в зависимости от положения на изображении контрольной точки, что говорит о гибкости и оптимальности предложенного математического аппарата.

Для плоского рельефа разработанный математический аппарат позволяет прогнозировать положение искомой точки соответствия на втором изображении и стереоскопическое рассогласование по заданным координатам на изображении контрольной точки соответствия.

В работе не рассмотрен случай отсутствия точки соответствия на втором изображении из-за смещения полей зрения камер или из-за окклюзий. Однако предложенный подход позволяет расчетным путем определить положение границ зоны неопределенности на втором изображении. Если эти границы находятся вне области изображения, то очевидно отсутствие точки соответствия на втором изображении вследствие смещения полей зрения камер. Если же имеет место окклюзия, то те же расчетные границы зоны неопределенности позволяют надежнее обнаружить это явление, чем поиск второй точки соответствия в более широкой области, не ограниченной расчетным путем.

Литература

1. Луманн Т., Робсон С., Кайл С., Бом Я. Ближняя фотограмметрия и 3D-зрение / пер. с англ. М.: ЛЕНАНД, 2018. 704 с.
2. Назаров А.С. Фотограмметрия: пособие для студентов вузов / 2-е изд., перераб. и доп. Минск: ТетраСистемс, 2010. 400 с.
3. Pepe M., Costantino D., Alfio V.S., Vozza G., Cartellino E. A novel method based on deep learning, GIS and geomatics software for building a 3D city model from VHR satellite stereo imagery // ISPRS International Journal of Geo-Information. 2021. V. 10. N 10. P. 697. https://doi.org/10.3390/ijgi10100697
4. Ma Y., Li Q., Chu L., Zhou Y., Xu C. Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision // Remote Sensing. 2021. V. 13. N 2. P. 230. https://doi.org/10.3390/rs13020230
5. Ding J., Yan Z., We X. High-accuracy recognition and localization of moving targets in an indoor environment using binocular stereo vision // ISPRS International Journal of Geo-Information. 2021. V. 10. N 4. P. 234. https://doi.org/10.3390/ijgi10040234
6. Albanwan H., Qin R. A comparative study on deep-learning methods for dense image matching of multi-angle and multi-date remote sensing stereo-images // The Photogrammetric Record. 2022. V. 37. N 180. P. 385–409. https://doi.org/10.1111/phor.12430
7. Fan R., Wang H., Cai P., Wu J., Bocus M.J., Qiao L., Liu M. Learning collision-free space detection from stereo images: Homography matrix brings better data augmentation // IEEE/ASME Transactions on Mechatronics. 2022. V. 27. N 1. P. 225–233. https://doi.org/10.1109/tmech.2021.3061077
8. Yang G., Liao Y. An improved binocular stereo matching algorithm based on AANet // Multimedia Tools and Applications. 2023. V. 82. N 26. P. 40987–41003. https://doi.org/10.1007/s11042-023-15183-6
9. Han Y., Pan C., Cheng Z., Xu Y. A PTV-based feature-point matching algorithm for binocular stereo photogrammetry // Measurement Science and Technology. 2023. V. 34. N 12. P. 125602. https://doi.org/10.1088/1361-6501/acf875
10. Wei H., Meng L. An accurate stereo matching method based on color segments and edges // Pattern Recognition. 2023. V. 133. P. 108996. https://doi.org/10.1016/j.patcog.2022.108996
11. Якасова Н.В. Алгоритмы поиска регулярных объектов на изображениях // Образовательные ресурсы и технологии. 2016. № 2. С. 277–281.
12. Фаворская М.Н., Тупицын И.В. Метод повышения устойчивости сопоставлений на стереоизображениях // Механика, управление и информатика. 2012. № 3(9). С. 139–144.
13. Chen M., Duan Z., Lan Z., Yi S. Scene reconstruction algorithm for unstructured weak-texture regions based on stereo vision // Applied

References

1. Luhmann T., Robson S., Kyle S., Boehm J. *Close-Range Photogrammetry and 3D Imaging*. Moscow, LENAND Publ., 2018, 704 p.
2. Nazarov A.S. *Photogrammetry*. 2nd ed., Minsk, TetraSystems Publ., 2010, 400 p. (in Russian)
3. Pepe M., Costantino D., Alfio V.S., Vozza G., Cartellino E. A novel method based on deep learning, GIS and geomatics software for building a 3D city model from VHR satellite stereo imagery. *ISPRS International Journal of Geo-Information*, 2021, vol. 10, no. 10, pp. 697. https://doi.org/10.3390/ijgi10100697
4. Ma Y., Li Q., Chu L., Zhou Y., Xu C. Real-time detection and spatial localization of insulators for UAV inspection based on binocular stereo vision. *Remote Sensing*, 2021, vol. 13, no. 2, pp. 230. https://doi.org/10.3390/rs13020230
5. Ding J., Yan Z., We X. High-accuracy recognition and localization of moving targets in an indoor environment using binocular stereo vision. *ISPRS International Journal of Geo-Information*, 2021, vol. 10, no. 4, pp. 234. https://doi.org/10.3390/ijgi10040234
6. Albanwan H., Qin R. A comparative study on deep-learning methods for dense image matching of multi-angle and multi-date remote sensing stereo-images. *The Photogrammetric Record*, 2022, vol. 37, no. 180, pp. 385–409. https://doi.org/10.1111/phor.12430
7. Fan R., Wang H., Cai P., Wu J., Bocus M.J., Qiao L., Liu M. Learning collision-free space detection from stereo images: Homography matrix brings better data augmentation. *IEEE/ASME Transactions on Mechatronics*, 2022, vol. 27, no. 1, pp. 225–233. https://doi.org/10.1109/tmech.2021.3061077
8. Yang G., Liao Y. An improved binocular stereo matching algorithm based on AANet. *Multimedia Tools and Applications*, 2023, vol. 82, no. 26, pp. 40987–41003. https://doi.org/10.1007/s11042-023-15183-6
9. Han Y., Pan C., Cheng Z., Xu Y. A PTV-based feature-point matching algorithm for binocular stereo photogrammetry. *Measurement Science and Technology*, 2023, vol. 34, no. 12, pp. 125602. https://doi.org/10.1088/1361-6501/acf875
10. Wei H., Meng L. An accurate stereo matching method based on color segments and edges. *Pattern Recognition*, 2023, vol. 133, pp. 108996. https://doi.org/10.1016/j.patcog.2022.108996
11. Yakasova N.V. Search algorithms for the regular objects in the image. *Educational Resources and Technologies*, 2016, no. 2, pp. 277–281. (in Russian)
12. Favorskaya M.N., Toupitsyn I.V. Improving robust method of feature points correspondences on stereo images. *Mechanics, Control and Informatics*, 2012, no. 3(9), pp. 139–144. (in Russian)

Sciences. 2023. V. 13. N 11. P. 6407. https://doi.org/10.3390/app13116407
14. Liu C.W., Wang H., Guo S., Bocus M.J., Chen Q., Fan R. Stereo matching: fundamentals, state-of-the-art, and existing challenges // Autonomous Driving Perception: Fundamentals and Applications. Springer Nature Singapore, 2023. P. 63–100. https://doi.org/10.1007/978-981-99-4287-9_3
15. Фаворская М.Н., Тупицын И.В. Иерархический метод поиска соответствующих точек на стереоизображениях // Вестник Сибирского государственного Аэрокосмического университета имени академика М.Ф. Решетнёва. 2012. № 1(41). С. 62–67.
16. Степанов Д.Н. Методики сопоставления особых точек в задаче визуальной навигации БПЛА // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. 2015. Т. 4. № 4. С. 32–47. https://doi.org/10.14529/cmse150402
17. Гошин Е.В., Фурсов В.А. Метод согласованной идентификации в задаче определения соответственных точек на изображении // Компьютерная оптика. 2012. Т. 36. № 1. С. 131–135.
18. Орлов В.П., Шариков Е. Алгоритм нахождения и классификации особых точек объекта на основе детектора Харриса // Наноиндустрия. 2017. № S(74). С. 171–178.
19. Гук А.П., Алтынцев М.А. Автоматическая идентификация соответственных точек на аэроснимках лесных массивов // Вестник СГУГиТ. 2017. Т. 22. № 4. С. 68–77.
20. Загорский М.Ю., Богданов В.Л., Гарманов В.В., Королева В.П., Рябов Ю.В. Математическая модель снимков стереопары и алгоритм восстановления рельефа местности на ее основе // Современные проблемы дистанционного зондирования Земли из космоса. 2015. Т. 12. № 3. С. 36–51.
21. Самойленко М.В. Векторно-матричный метод восстановления пространственных координат точки в общем случае стереосъемки // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18. № 6. С. 954–960. https://doi.org/10.17586/2226-1494-2018-18-6-954-960
22. Самойленко М.В. Влияние асимметрии интенсивности на ошибку идентификации стереоизображений // Измерительная техника. 2023. № 5. С. 22–28. https://doi.org/10.32446/0368-1025it.2023-5-22-28

13. Chen M., Duan Z., Lan Z., Yi S. Scene reconstruction algorithm for unstructured weak-texture regions based on stereo vision. *Applied Sciences*, 2023, vol. 13, no. 11, pp. 6407. https://doi.org/10.3390/app13116407
14. Liu C.W., Wang H., Guo S., Bocus M.J., Chen Q., Fan R. Stereo matching: fundamentals, state-of-the-art, and existing challenges. *Autonomous Driving Perception: Fundamentals and Applications*. Springer Nature Singapore, 2023, pp. 63–100. https://doi.org/10.1007/978-981-99-4287-9_3
15. Favorskaya M.N., Tupitsyn I.V. Hierarchical method of search of corresponding points at stereoimages. *The Siberian Aerospace Journal*, 2012, no. 1(41), pp. 62–67. (in Russian)
16. Stepanov D.N. Techniques of feature points matching in the problem of uav's visual navigation. *Bulletin of the South Ural State University. Series: Computational Mathematics and Software Engineering*, 2015, vol. 4, no. 4, pp. 32–47. https://doi.org/10.14529/cmse150402
17. Goshin Ye.V., Fursov V.A. Conformed identification in corresponding points detection problem. *Computer Optics*, 2012, vol. 36, no. 1, pp. 131–35. (in Russian)
18. Orlov V.P., Sharikov E.N. Algorithm of finding and classifying special points of object on the basis of Harris's detector. *Journal Nanoindustry*, 2017, no. S(74), pp. 171-178. (in Russian)
19. Guk A.P., Altyntsev M.A. Automatic identification of corresponding points for aerial images of forest areas. *Vestnik of SSUGT*, 2017, vol. 22, no. 4, pp. 68–77. (in Russian)
20. Zagorskii M.Yu., Bogdanov V.L., Garmanov V.V., Koroleva V.P., Ryabov Yu.V. Mathematical model of stereo images and terrain reconstruction algorithm based on it. *Sovremennye problemy distantsionnogo zondirovaniya Zemli iz kosmosa*, 2015, vol. 12, no. 3, pp. 36–51. (in Russian)
21. Samoilenko M.V. Vector-matrix method for restoration of point spatial coordinates in stereo photography general case. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2018, vol. 18, no. 6, pp. 954–960. (in Russian). https://doi.org/10.17586/2226-1494-2018-18-6-954-960
22. Samoilenko M.V. Effect of Intensity asymmetry on stereo image identification error. *Measurement Techniques*, 2023, vol. 66, no. 5, pp. 311–319. https://doi.org/10.1007/s11018-023-02229-2

Авторы

Самойленко Марина Витальевна — кандидат технических наук, доцент, независимый исследователь, Москва, Российская Федерация, sc 57191194098, https://orcid.org/0000-0002-4798-2414, samoi.mar@mail.ru

Хачикян Владимир Александрович — кандидат технических наук, старший научный сотрудник, доцент, Московский авиационный институт (национальный исследовательский университет), Москва, 125993, Российская Федерация, https://orcid.org/0009-0006-1279-7661, zaktol@yandex.ru

Authors

Marina V. Samoilenko — PhD, Associate Professor, Independent Researcher, Moscow, Russian Federation, sc 57191194098, https://orcid.org/0000-0002-4798-2414, samoi.mar@mail.ru

Vladimir A. Hachikian — PhD, Associate Professor, Senior Researcher, Moscow Aviation Institute (National Research University), Moscow, 125993, Russian Federation, https://orcid.org/0009-0006-1279-7661, zaktol@yandex.ru

Статья поступила в редакцию 04.09.2023
Одобрена после рецензирования 09.12.2023
Принята к печати 16.01.2024

Received 04.09.2023
Approved after reviewing 09.12.2023
Accepted 16.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-20-29
УДК 535.417: 539.388.1

Сравнение результатов применения двух спекловых методов изучения многоцикловой усталости конструкционной стали

Александр Петрович Владимиров[1], Иван Сергеевич Каманцев[2], Никита Александрович Друкаренко[3], Константин Евгеньевич Мызнов[4]✉, Константин Вячеславович Наумов[5]

[1,2,3] Институт машиноведения имени Э.С. Горкунова Уральского отделения Российской академии наук, Екатеринбург, 620049, Российская Федерация
[4] Институт физики металлов имени М.Н. Михеева Уральского отделения Российской академии наук, Екатеринбург, 620108, Российская Федерация
[5] Уральский федеральный университет имени первого президента России Б.Н. Ельцина, Екатеринбург, 620002, Российская Федерация

[1] vap52@bk.ru, https://orcid.org/0000-0001-6343-9772
[2] ks@imach.uran.ru, https://orcid.org/0000-0003-0775-4043
[3] targosk@ya.ru, https://orcid.org/0000-0002-5979-7398
[4] myznov@imp.uran.ru✉, https://orcid.org/0009-0002-0411-1712
[5] knaumov98@mail.ru, https://orcid.org/0000-0002-5636-9857

Аннотация

Введение. Новым методом усредненных во времени спекловых изображений и известным ранее методом интерференции двух спекл-полей изучено развитие пластических деформаций, возникающих при многоцикловой усталости конструкционной стали. Выполнена оценка корректности определения деформаций новым методом путем сравнения данных, полученных двумя методами. **Метод.** Исследованы две оптические системы, в которых применены лазерные модули с разными длинами волн. Предложенная оптическая система позволяет определять компоненты Δu_y, Δu_z вектора относительного перемещения двух точек поверхности, расположенных на расстоянии (базе измерения) $\Delta s = 66$ мкм. Известная схема дает возможность описать деформации традиционным методом на базе измерения 470 мкм. Исследуемый образец представлял собой плоскую пластину с двумя боковыми вырезами, выполненную из стали марки 09Г2С. Испытания на усталость проводились на установке резонансного типа при разных амплитудах цикла. **Основные результаты.** Показано, что при всех амплитудах цикла развитие пластических деформаций происходит по механизму циклической ползучести. Существует хорошая корреляция данных, полученных двумя разными спекловыми методами. Вместе с тем деформация, оцененная новым методом, в ряде случаев на порядок превышает значение деформации, рассчитанной известным методом. Очевидно, это связано с существованием локальных малоразмерных (порядка 10^1 мкм) участков деформации, измерение которых традиционными методами невозможно. Рассчитанная новым методом предельная растягивающая деформация $\Delta u_y / \Delta s$ имеет порядок 10^{-1}, что совпадает с аналогичной деформацией, возникающей при испытании стандартных образцов на разрыв. **Обсуждение.** Результаты, полученные новым методом, обосновывают необходимость разработки датчиков и приборов неразрушающего контроля нового поколения, позволяющих по скорости изменения физических величин и по их предельным значениям оценивать время до зарождения усталостной трещины.

Ключевые слова

многоцикловая усталость, сталь, разрушение, спеклы, интерференция, свободное пространство, изображение

Благодарности

Работа выполнена в рамках государственного задания Министерства науки и высшего образования Российской Федерации по темам № АААА-А18-118020790148-1 и «Диагностика» № 122021000030-1. При выполнении работы было использовано оборудование Центра коллективного пользования «Пластометрия» при Институте машиноведения имени Э.С. Горкунова Уральского отделения Российской академии наук.

© Владимиров А.П., Каманцев И.С., Друкаренко Н.А., Мызнов К.Е., Наумов К.В., 2024

Ссылка для цитирования: Владимиров А.П., Каманцев И.С., Друкаренко Н.А., Мызнов К.Е., Наумов К.В. Сравнение результатов применения двух спекловых методов изучения многоцикловой усталости конструкционной стали // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 20–29. doi: 10.17586/2226-1494-2024-24-1-20-29

Comparison of application results of two speckle methods for study multi-cycle fatigue of structural steel

Alexandr P. Vladimirov[1], Ivan S. Kamantsev[2], Nikita A. Drukarenko[3], Konstantin E. Myznov[4]✉, Konstantin V. Naumov[5]

[1,2,3] Institute of Engineering Science, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, 620049, Russian Federation

[4] Institute of Metal Physics, Ural Branch of Russian Academy of Sciences, Ekaterinburg, 620108, Russian Federation

[5] Ural Federal University named after the First President of Russia B.N.Yeltsin, Ekaterinburg, 620002, Russian Federation

[1] vap52@bk.ru, https://orcid.org/0000-0001-6343-9772

[2] ks@imach.uran.ru, https://orcid.org/0000-0003-0775-4043

[3] targosk@ya.ru, https://orcid.org/0000-0002-5979-7398

[4] myznov@imp.uran.ru✉, https://orcid.org/0009-0002-0411-1712

[5] knaumov98@mail.ru, https://orcid.org/0000-0002-5636-9857

Abstract

The new method of time-averaged speckle images and the well-known method of speckle-field interference studied the development of plastic deformations occurring during multicycle fatigue of structural steel. The correctness of strain determination by the new method is evaluated by comparing the data obtained by the two methods. Two optical systems including laser modules with different wavelengths were investigated. The proposed optical system allows to determine the components Δu_y, Δu_z of the relative displacement vector of two surface points located at a distance (measurement base) Δs = 66 μm. The known scheme makes it possible to describe deformations by the traditional method on the base of 470 μm. The object of study was a flat specimen made of 09Г2С steel with two side notches. Fatigue tests were carried out on a resonance-type machine at different cycle amplitudes. It is shown that at all cycle amplitudes the development of plastic deformations occurs by the mechanism of cyclic creep. There is a good correlation between the data obtained by two different speckle methods. At the same time, the strain estimated by the new method is in some cases an order of magnitude higher than the strain calculated by the known method. Obviously, this is due to the existence of local small-sized (of the order of 10^1 μm) strain areas which cannot be measured by conventional methods. The ultimate tensile strain $\Delta u_y/\Delta s$ calculated by the new method is of the order of 10^{-1}, which coincides with the similar strain occurring in tensile testing of standard specimens. The results obtained by the new method justify the need of developing sensors and nondestructive testing devices of a new generation, allowing to estimate the time to fatigue crack initiation by the rate of change of physical quantities and by their limit values.

Keywords

high-cycle fatigue, steel, destruction, speckles, interference, free space, image

Acknowledgements

The research was carried out within the state assignment of the Ministry of Science and Higher Education of the Russian Federation under the topic no. AAAA-18-118020790148-1 and “Diagnostics” no. 122021000030-1. The equipment of the Collective Center “Plastometria” at Institute of Engineering Science of Ural Branch of the Russian Academy of Sciences was used for this work.

For citation: Vladimirov A.P., Kamantsev I.S., Drukarenko N.A., Myznov K.E., Naumov K.V. Comparison of applications results of two speckle methods for study multi-cycle fatigue of structural steel. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 20–29 (in Russian). doi: 10.17586/2226-1494-2024-24-1-20-29

Введение

При освещении когерентным светом шероховатой поверхности, а также прозрачного объекта через матовое стекло, в свободном пространстве и в плоскости изображения объектов образуется пятнистая или спекловая структура рассеянного излучения. Использование спекловых [1], а также других оптических [2] и не оптических методов [3, 4] для изучения многоцикловой усталости материалов является актуальной задачей. В работах [5, 6] описан новый спекловый метод, который по чувствительности и пространственному разрешению на два порядка превосходит имеющиеся оптические методы определения деформации. В [7] путем регистрации на одном кадре трех спекловых изображений, формируемых под тремя ракурсами наблюдения, с использованием нового метода были найдены зависимости трех компонент вектора относительного перемещения точек поверхности от числа циклов деформации. Было обнаружено, что предельная локальная растягивающая деформация, соответствующая зарождению трещины, на базе измерения 66 мкм достигает порядка 10^{-1}. Полученное значение деформации по порядку величин совпадает с предельной деформацией, возникающей при испытаниях стандартных образцов в условиях их квазистатического растяжения. Вследствие важности указанного совпадения, как с научной, так и с практической точек

зрения, необходимо оценить надежность полученного результата. Недостатком метода является определение порядка интерференции с точностью до знака. В настоящей работе данные, полученные новым методом, сравнивались с результатами, полученными с помощью известного спеклового метода [8], позволяющего определять величину и знак порядка интерференции.

Цели настоящей работы: воспроизведение значения предельной растягивающей деформации, соответствующей зарождению трещины; сравнение необратимых относительных перемещений точек поверхности одновременно новым методом и известным методом интерференции двух спекл-полей; изучение особенности локализации необратимых деформаций при разных амплитудах цикла.

Теоретические положения

В работе [8] представлены теория метода интерференции двух спекл-полей в свободном пространстве и возможности его применения для изучения деформаций. Если два малых участка тела осветить когерентным светом от одного источника, то при равенстве направлений освещения этих участков и направлений наблюдения, связь между разностью оптических путей Δu двух волн, отраженных от центров двух освещенных участков поверхности, находящихся на расстоянии Δs, и порядком интерференции Δn в точке наблюдения, расположенной в области наложения двух спекл-полей, определим формулой:

$$\Delta u = \Delta \mathbf{u} \cdot (\mathbf{l}_s + \mathbf{l}) = \lambda \Delta n, \qquad (1)$$

где λ — длина волны излучения; $\mathbf{l}_s$ и $\mathbf{l}$ — единичные векторы, направленные от центров освещенных участков к источнику света и в точку наблюдения; $\Delta \mathbf{u}$ — вектор относительного перемещения двух точек поверхности, расположенных на расстоянии (на базе измерения) Δs.

Теория усредненных во времени спекловых изображений для отражающего и прозрачного объектов приведена в работах [5, 9] и [6, 10, 11]. В принятой модели предположено, что в общем случае центры рассеяния отражающего объекта перемещаются поступательно, совершают периодические движения, а также случайным образом перемещаются в пространстве. Аналогично предполагалось, что тонкий прозрачный объект, расположенный вблизи случайного фазового экрана, перемещается поступательно, совершает периодические движения, а показатель преломления изменяется случайным образом. Регистрируемая величина — среднее по времени цифровое значение интенсивности излучения $\tilde{I}$ в произвольной точке $\mathbf{q}$ плоскости изображения объекта, а также временная автокорреляционная функция $\eta(t_1, t_2)$ этой интенсивности, где t_1 и t_2 — начальный и текущий моменты времени. Отметим, что время усреднения должно быть кратным периоду колебаний объекта. Определяемая величина — разность оптических путей Δu двух волн, распространяющихся на среднем, характерном расстоянии Δs в окрестности точки в плоскости объекта, сопряженной точке $\mathbf{q}$. Размер этой области равен линейному разрешению линзы Δx. При изменении Δu во времени функции $\tilde{I}(t)$ и $\eta(t_1, t_2)$ в общем случае являются квазипериодическими. Если объект является отражающим, то связь между Δu и Δn определяется по формуле (1), где теперь Δn — период указанных функций $\tilde{I}(t)$ и $\eta(t_1, t_2)$. За счет использования нескольких направлений наблюдения становится возможной регистрация изменения компонент вектора $\Delta \mathbf{u}$ во времени. Для прозрачного объекта используется формула $\Delta u = \Delta n_0 L$, где Δn_0 — разность показателя преломления на базе Δs; L — толщина объекта. На практике, как будет показано в разделе «Методика эксперимента», функцию $\eta(t_1, t_2)$ определим по фрагменту спеклового изображения. Этому фрагменту в сопряженной области объекта соответствует область поверхности, содержащая определенное количество участков размером Δx. Если деформация области поверхности однородна, то функция $\eta(t_1, t_2)$ квазипериодически изменяется в интервале от +1 до –1. Если деформация неоднородна, то по недостижению функцией $\eta(t_1, t_2)$ значений +1 и –1 можно определить дисперсию величины Δu.

Преобразуем формулу (1) к следующему виду:

$$\Delta u_d 2\cos(\theta/2) = \lambda \Delta n, \qquad (2)$$

где Δu_d — проекция вектора $\Delta \mathbf{u}$ на биссектрису угла θ между единичными векторами $\mathbf{l}_s$ и $\mathbf{l}$.

Методика эксперимента

Для проведения эксперимента были собраны две оптические системы, в которых применены лазерные модули с разными длинами волн. Первая оптическая система (рис. 1, *а*) использована для формирования двух спекловых изображений (рис. 2, *а*) на одном кадре монохромной видеокамеры, необходимых для определения двух компонент вектора относительных необратимых перемещений. Направление освещения и два направления наблюдения объекта расположены в плоскости *yoz*. Во второй оптической системе (рис. 1, *b*) луч лазера разделен на два пучка, пучки фокусировались на два расположенные рядом участка поверхности в центре образца. Интерференционные полосы (рис. 2, *b*), возникающие в свободном пространстве в области наложения двух спеклов, были зарегистрированы с помощью цветной видеокамеры. Направление освещения и наблюдения также сформированы в плоскости *yoz*. Объект исследования — плоский образец из стали 09Г2С с двумя симметричными выточками диаметром 5 мм и размерами 3,1 × 1,8 мм в минимальном сечении (рис. 3). Образец был подвергнут циклическому деформированию с частотой около 100 Гц и коэффициентом асимметрии цикла 0,1, при этом амплитуда цикла последовательно увеличивалась от 1,0 кН до 1,8 кН с шагом 0,1 кН. Образец выдерживался при постоянной амплитуде сначала в течение 34 и 65 тысяч циклов, затем в течение 100 тысяч циклов.

Рассматриваемые оптические системы были собраны таким образом, чтобы биссектрисы угла θ_1 для одного из направлений наблюдения первой системы и угла θ_2 второй системы совпали. С использованием

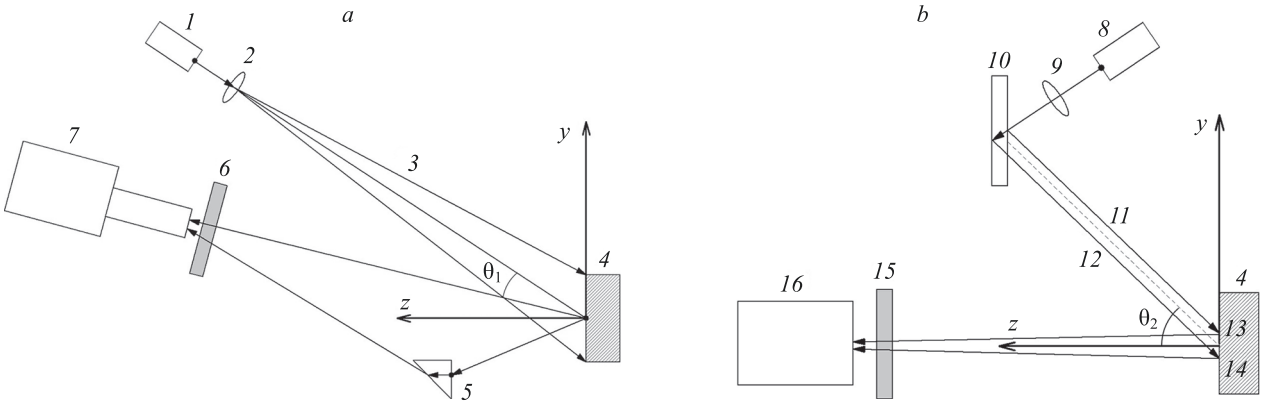


Рис. 1. Оптические системы, использованные для изучения усталости стали спекловыми методами: усредненных во времени спекловых изображений (*a*) и интерференции спекл-полей (*b*).

1 — лазерный модуль (λ = 532 нм); *2* — микрообъектив; *3* — освещающий пучок; *4* — образец; *5* — оптическая призма; *6* — зеленый светофильтр; *7* — монохромная видеокамера; *8* — лазерный модуль (λ = 650 нм); *9* — фокусирующая линза; *10* — плоскопараллельная пластина; *11* и *12* — освещающие пучки; *13* и *14* — освещенные участки; *15* — красный светофильтр; *16* — цветная видеокамера

Fig. 1. Optical systems used to study steel fatigue by speckle methods: time-averaged speckle images (*a*) and speckle-field interference (*b*).

1 — laser module (λ = 532 nm); *2* — microlens; *3* — illuminating beam; *4* — specimen; *5* — optical prism; *6* — green light filter; *7* — monochrome video camera; *8* — laser module (λ = 650 nm); *9* — focusing lens; *10* — parallel-sided plate; *11* and *12* — illuminating beams; *13* and *14* — illuminated areas; *15* — red light filter; *16* — color video camera

двух систем были зарегистрированы три зависимости величины Δn от числа циклов N нагружения. На кадре (рис. 2, *b*) в центре светлой или темной интерференционной полосы выбрана точка (пиксел) наблюдения и по смещению полос относительно выбранной точки визуально с точностью 1/10 полосы определено значение порядка интерференции Δn. На двух спекловых изображениях (рис. 2, *a*) в центре изображений рабочей части образца выбраны пикселы для определения цифрового значения интенсивности $\tilde{I}$, а также два фрагмента изображения размером 3 × 3 пиксела для определения коэффициента корреляции η. По функциям $\tilde{I}(t)$ и $\eta(t_1, t_2)$

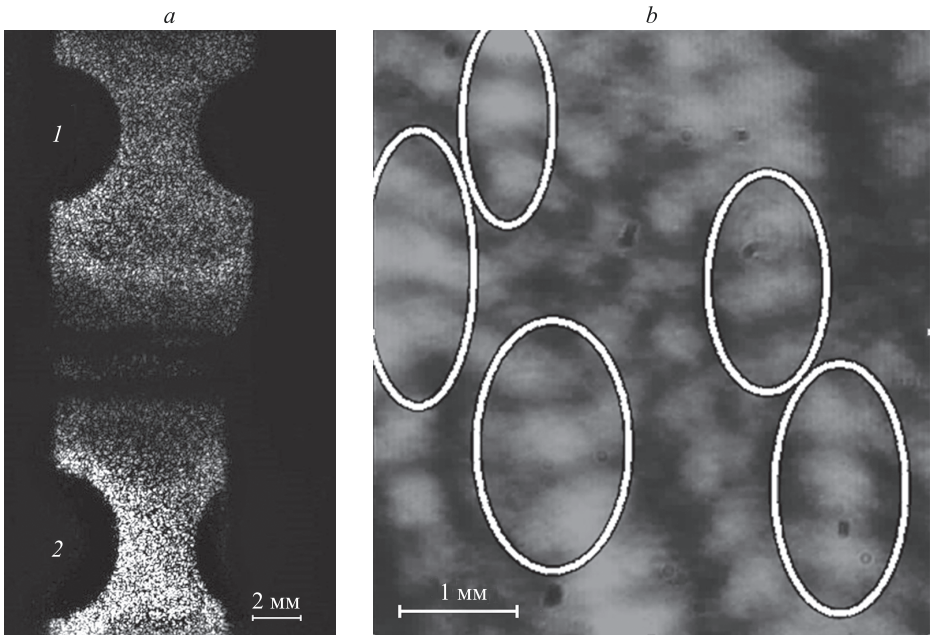


Рис. 2. Кадры, полученные с помощью методов: усредненных во времени спекловых изображений (*a*) и интерференции спекл-полей (*b*).

1 и *2* — спекловые изображения, соответствующие двум направлениям наблюдения под углами 88°12′ и 101°18′ к оси *oy*. Белые овалы — области, по которым можно определять смещение полос

Fig. 2. Frames obtained using the time-averaged speckle image method (*a*) and the speckle-field interference method (*b*).

1 and *2* — speckle images corresponding to two observation directions at angles of 88°12′ and 101°18′ to the *oy* axis. White ellipses — areas by which the fringe shift can be determined

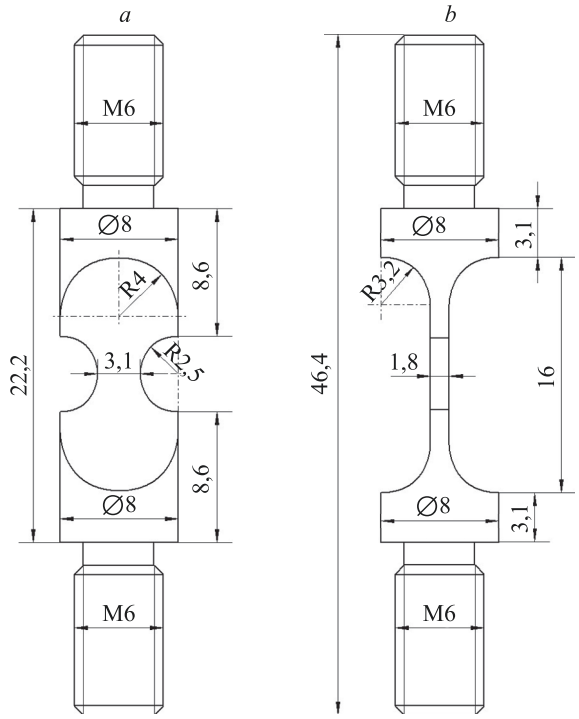


Рис. 3. Плоский образец из стали 09Г2С видов: спереди (*a*) и сбоку (*b*)

Fig. 3. Flat specimen from 09Г2С steel: front (*a*) and side (*b*) views

определены зависимости количества периодов Δn этих функций от числа циклов нагружения N. Аналогично получены зависимости $\Delta n(N)$, соответствующие участку поверхности в зоне зарождения трещины. Величина η, являющаяся коэффициентом корреляции цифровых значений интенсивности в моменты времени t_1 и t_2, определена с точностью порядка 1 % [5, формула (38)].

Особенности локализации пластических деформаций при разных амплитудах прикладываемых усилий изучали по распределениям величины η в изображении образца, расположенного в верхней части кадра (рис. 2, *a*). Для визуализации приращений зон пластических деформаций в начале опыта время t_1 выбрано после 50 циклов нагружения, а время t_2 — в пределах временного интервала, соответствующего постоянному значению амплитуды цикла в 1,1 кН. Размеры фрагментов изображения, равнялись 5 × 5 пикселам. Одному пикселу соответствует участок объекта, равный 66 мкм. Локализация деформаций при увеличении амплитуды цикла еще на 0,1 кН наблюдалась аналогичным образом. Для точного определения начала развития необратимых процессов время t_1 было выбрано до изменения амплитуды цикла, а время t_2 — после. На рис. 4 приведены типичные распределения величины η, полученные по рассмотренной методике, демонстрирующие зоны приращений пластических деформаций. Размер ΔY зоны приращенных пластических деформаций определим по расстоянию вдоль оси *oy* между верхней и нижней границами визуально наблюдаемой неоднородности в распределении величины η. Границу неоднородности определим по уровню η = 0,6, что примерно соответствует приращению пластической деформации на величину порядка 10^{-4}.

Результаты и их обсуждение

На рис. 5 приведены типичные совмещенные зависимости величины Δn от числа циклов нагружения N для центра образца, полученные двумя рассмотренными в разделе «Методика эксперимента» спекловыми методами. Зависимости были получены путем обработки кадров, зарегистрированных после увеличения амплитуды цикла от 1,1 до 1,2 кН. Видно, что имеется хорошая корреляция между значениями массивов Δn двух методов. Для разных амплитуд цикла коэффициенты корреляции двух массивов данных находятся в пределах от 0,86 до 0,996. По максимальным значениям Δn (2,4 и 2,25), приведенным на рис. 5, по формуле (2) с учетом конкретных значений угла θ, для двух методов имеем два значения Δu_d, равные 0,64 и 0,75 мкм. В качестве меры необратимой деформации ε_p рассчитаем отношение Δu_d к базе измерения. Получим, что соответствующее увеличению амплитуды цикла от 1,1

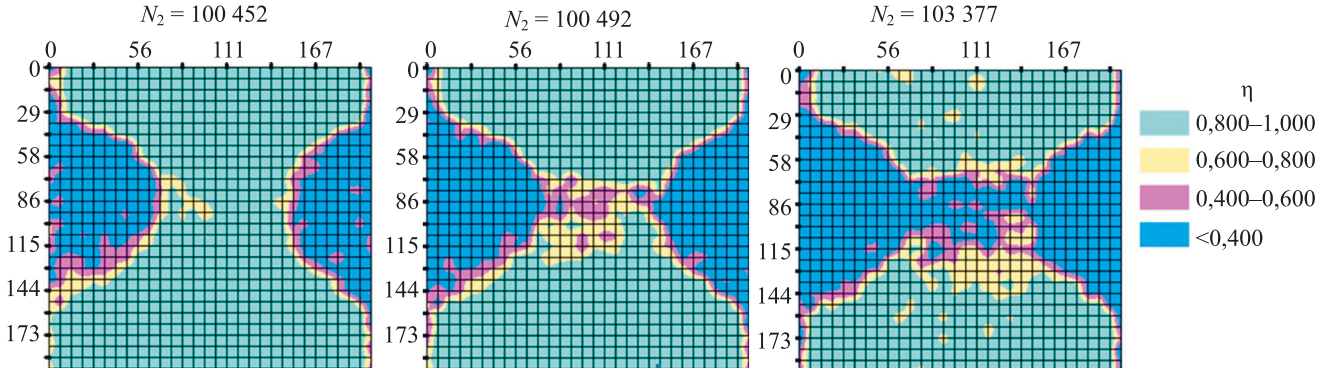


Рис. 4. Распределение величины $\eta(t_1, t_2)$ при увеличении амплитуды цикла от 1,1 до 1,2 кН и значении N_1 равном 99 751 циклов.

N_1 и N_2 — число циклов, соответствующее времени t_1 и t_2

Fig. 4. Distribution of the value of $\eta(t_1, t_2)$ when increasing the double cycle amplitude from 1.1 to 1.2 kN and the value of N_1 equal to 99,751 cycles.

N_1 and N_2 — number of cycles corresponding to time t_1 and t_2

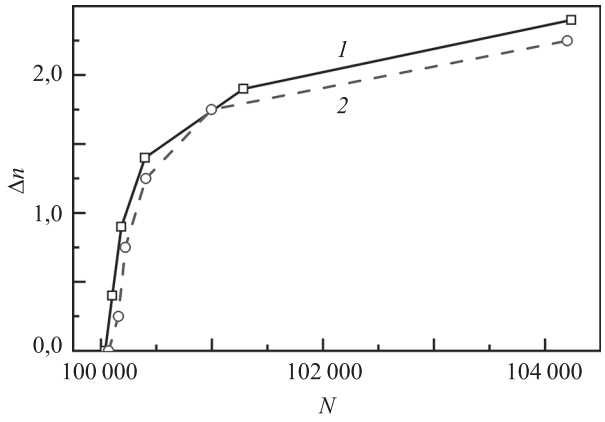


Рис. 5. Типичные зависимости величины Δ*n* от числа циклов нагружения *N*, полученные двумя спекловыми методами: усредненных во времени спекловых изображений (кривая *1*) и интерференции спекл-полей (кривая *2*)

Fig. 5. Typical dependences of Δ*n* value vs. the number of loading cycles *N* obtained by two speckle methods: time-averaged speckle images (curve *1*) and speckle-field interference (curve *2*)

до 1,2 кН значение ε_p на базе измерения 66 мкм, найденное по изменению величины $\tilde{I}$, равно $0{,}97\cdot10^{-2}$, а на базе 470 мкм (расстояние между двумя освещенными участками на поверхности) по смещению полос — $1{,}6\cdot10^{-3}$. Оценка деформаций по всем амплитудам цикла показала, что деформации, найденные методом усредненных во времени спекловых изображений на порядок больше, чем найденные методом интерференции спекл-полей.

На рис. 6 приведены совмещенные зависимости величины, полученные методом усредненных во времени спекловых изображений, от числа циклов нагружения *N* для двух направлений наблюдения, кривые соответствуют центру рабочей части образца. По полученным зависимостям построены графики Δ*n*(*N*). Значения Δ*n* определены по достижению η нуля, а также по минимальным отрицательным и максимальным положительным значениям η. Величина Δ*n* для произвольного значения *N* определена с помощью линейной интерполяции данных по ближайшим двум значениям Δ*n*. По двум зависимостям Δ*n*(*N*) путем решения системы двух уравнений, выведенных из формулы (1), с двумя неизвестными компонентами Δu_y и Δu_z вектора Δ**u** построены зависимости $\Delta u_y(N)$ и $\Delta u_z(N)$ (рис. 7). Видно, что предельные значения растягивающих деформаций ε_{yy}, найденных как отношение максимальных значений Δu_y к базе измерения 66 мкм, в центре образца равны 0,48, а в зоне зарождения трещины — 0,71, что в 1,5 раза больше, чем в центре образца. Полученные значения ε_{yy} имеют порядок 10^{-1}, что воспроизводит значение предельной растягивающей деформации в зоне зарождения трещины, приведенное в работе [7].

Значения Δ*Y* определим по методике, описанной в разделе «Методика эксперимента». Анализ распределений величины η при амплитуде 1,1 кН показал, что уже в самом начале стабилизации резонансной частоты машины необратимые процессы возникают в небольшой зоне на месте возникновения макротрещины. Далее зона пластических деформаций быстро охватывает область длиной 3,3 мм рабочей части образца. Затем наблюдалась полная декорреляция всей картины спеклов, по-видимому, связанная как с притиркой системы «образец-машина» в процессе стабилизации резонансной частоты испытательной машины, так и с деформацией участков размером 66 мкм на всей наблюдаемой области на величину порядка 10^{-3}. В дальнейшем был выбран первый кадр после стабилизации резонансной частоты для сравнения с последующими кадрами. Аналогично выполнен анализ приращения длины зоны

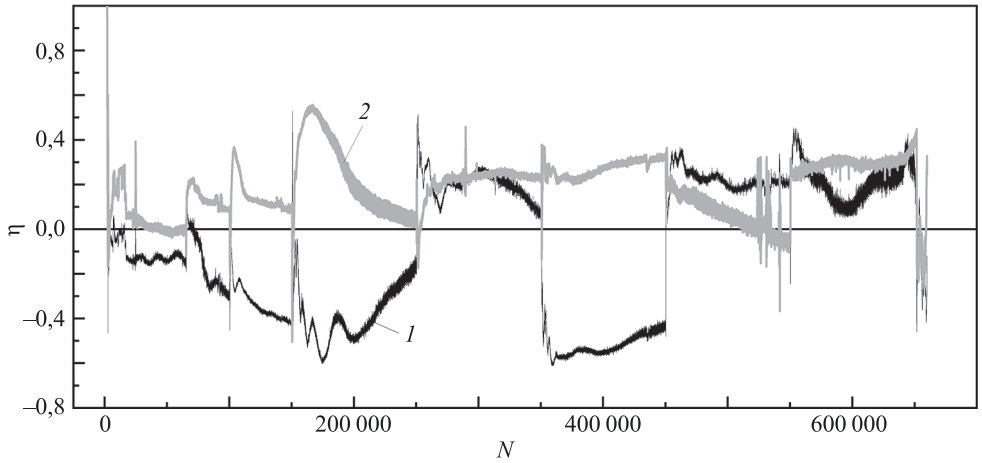


Рис. 6. Зависимости коэффициента корреляции η от числа циклов нагружения *N* для пикселов, находящихся в центре двух изображений рабочей части образца.

Направления наблюдения: под углами 88°12′ (кривая *1*) и 101°18′ (кривая *2*) к оси *oy*

Fig. 6. Dependence of the correlation coefficient η vs. the number of loading cycles *N* for the pixels located in the center of two images of the working part of the specimen.

Observation directions: at angles of 88°12′ (curve *1*) and 101°18′ (curve *2*) to the *oy* axis

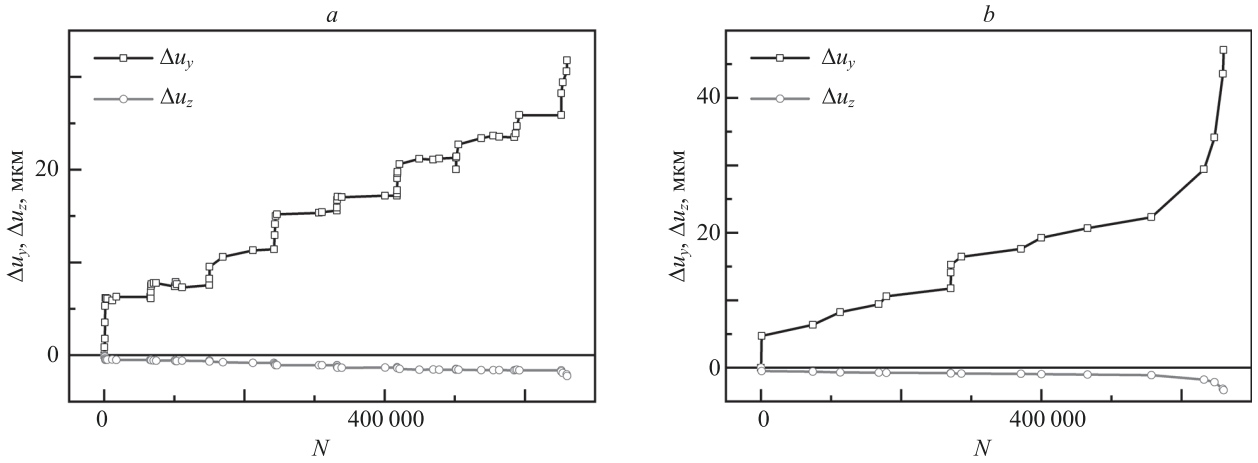


Рис. 7. Зависимости Δu_y, Δu_z от количества циклов N для участка, находящегося в центре рабочей части образца (*a*) и в зоне зарождения трещины (*b*)

Fig. 7. Dependences of Δu_y, Δu_z vs. the number of cycles N for the area located in the center of the working part of the specimen (*a*) and in the crack initiation zone (*b*)

необратимых процессов для остальных амплитуд цикла. Из приведенной на рис. 8 зависимости видно, что по мере увеличения амплитуды цикла первоначальное значение величины ΔY, равное 3,3 мм, достигается при амплитуде, равной 1,5 кН. После увеличения амплитуды цикла до 1,8 кН величина ΔY быстро увеличивается до 5,4 мм.

Данные, полученные в настоящей работе и в предыдущем опыте с острым надрезом типа Шарпи [12] на образце из стали 09Г2С, а также в работах [13–15], позволяют сделать выводы о накоплении повреждений при многоцикловой усталости стали 09Г2С. Как известно [16, 17], многоцикловая усталость возникает при напряжениях ниже, равных и превышающих предел текучести материала. Если максимальное напряжение цикла ниже предела текучести, в локальных областях материала найдутся участки, в которых напряжение превышает предел текучести. В этих областях может начаться так называемая циклическая ползучесть, характерная для циклически разупрочняющегося материала. Особенностью этой ползучести является резкое увеличение пластической деформации на начальной стадии усталости, затем постепенное ее уменьшение и подрастание с малой скоростью (рис. 4). Согласно ранее полученным данным, предельные пластические деформации циклически деформируемых образцов достигают значений 10^{-4} (в отдельных случаях 10^{-3}) [13]. В настоящей работе при использовании нового метода зафиксированы деформации порядка 10^{-2} при первых 780 циклах, и 10^{-1} на стадии предразрушения в центре образца, а также в зоне зарождения трещины (рис. 5). Причина полученных разных значений деформаций заключа-

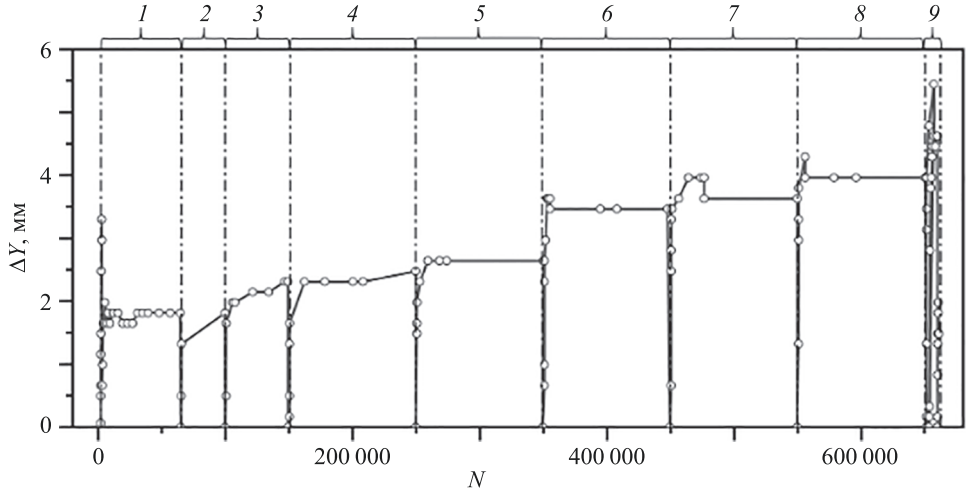


Рис. 8. Зависимость размера зоны приращенных пластических деформаций ΔY от количества циклов N. Числа сверху графика означают номера промежутков с постоянными значениями амплитуд цикла: *1* — 1 кН; *2* — 1,1 кН; *3* — 1,2 кН; *4* — 1,3 кН; *5* — 1,4 кН; *6* — 1,5 кН; *7* — 1,6 кН; *8* — 1,7 кН; *9* — 1,8 кН

Fig. 8. Dependence of the size of the incremented plastic deformation zone ΔY vs. the number of cycles N. The numbers at the top of the graph indicate the numbers of intervals with constant values of cycle amplitudes: *1* — 1 kN; *2* — 1.1 kN; *3* — 1.2 kN; *4* — 1.3 kN; *5* — 1.4 kN; *6* — 1.5 kN; *7* — 1.6 kN; *8* — 1.7 kN; *9* — 1.8 kN

ется в существенной разнице базы измерения: если в настоящей работе базы измерения были порядка 10^1 и 10^2 мкм, то в предыдущих работах до 100 мм и более.

По результатам анализа работ по циклической ползучести [18], разупрочнение материалов может быть вызвано структурно-фазовыми превращениями, накоплением усталостных повреждений и диссипативными процессами. В работе [9] представлены теоретические и экспериментальные данные, которые указывают на наличие нескольких m процессов, одновременно протекающих на поверхности образца с разной скоростью, которые вносят вклад в изменение величины Δu. Определение дисперсии и времени релаксации вкладов Δu_m в суммарное значение Δu по методике [9] является предметом дальнейших исследований.

Отметим, что в настоящее время не найдены конкретные механизмы усталостных повреждений, которые оказывают влияние на сталь 09Г2С. Выявление этих механизмов, анализ их влияния на изменение величин $\tilde{I}$ и η, а также изучение эволюции не только пластических, но и упругих деформаций будут изучены в дальнейшем.

Совпадение по порядку величин предельных деформаций, возникающих при многоцикловой усталости и в опытах на квазистатическое растяжение стандартных образцов, объясняет причину локализации деформаций, и причину не регистрации многоцикловой усталости существующими методами неразрушающего контроля. Как показывает опыт, при испытании и эксплуатации образцов и деталей типичных размеров, в условиях многоцикловой усталости имеют место абсолютные и относительные перемещения точек поверхности около 1 мкм. Тогда деформации металла порядка 10^{-1} будут в областях порядка 10^1 мкм, порядка 10^{-2} — в областях 10^2 мкм, а минимальные пластические деформации 10^{-3} — в областях 1 мм. Существующие методы контроля имеют базу измерения не менее 1 мм, но часто это сантиметры [13] и даже дециметры. Исходя из этого, большие деформации, возникающие в малых областях среды, как правило, регистрируются приборами в виде шума.

На основе сравнения данных, полученных двумя оптическими методами, можно сделать вывод о том, что при всех амплитудах цикла развитие пластических деформаций происходит по механизму циклической ползучести. Существует хорошая корреляция данных, полученных двумя разными спекловыми методами, однако деформация, оцененная новым методом, на порядок превышает значение деформации, рассчитанной известным методом. Различие растягивающих деформаций, оцененных двумя методами, объясняется использованием разных баз измерения, а также указывает на локализацию пластических деформаций на участках размером порядка 10^1 мкм.

Так как оптические методы контроля не всегда применимы на практике, то необходима модернизация существующих методов неразрушающего контроля, которые на данный момент не способны определять деформации на малых базах измерения. К таким методам можно отнести вихретоковый, акустический, магниторезистивный и тепловой. Модернизация возможна в направлении создания приборов с датчиками матричного типа с размером ячеек порядка 10^1 мкм.

Заключение

В работе использован плоский образец из стали 09Г2С с двумя симметричными боковыми вырезами. Образец применен для анализа развития пластических деформаций при многоцикловой усталости.

Для определения пластических деформаций применены новый и известный спекловые методы: усредненных во времени спекловых изображений и интерференции двух спекл-полей. Преимуществами нового метода перед известным является меньшая база измерения, позволяющая выявлять деформации на более малоразмерных участках на поверхности объекта, и возможность оценки сразу нескольких компонент вектора относительного перемещения.

Визуализация зоны пластических деформаций осуществлена по распределению коэффициента корреляции интенсивностей излучения во фрагментах спекловых изображений размером 5 × 5 пикселов, взятых в начальный и текущий моменты времени. Показано, что в самом начале опыта при амплитуде цикла 1,0 кН первые признаки необратимых процессов обнаружены на месте будущей трещины. Затем в течение 780 циклов длина зоны пластических деформаций увеличилась до 3,3 мм. При последующих увеличениях амплитуды до 1,5 кН приращение длины зоны не наблюдалось, необратимые деформации протекали внутри указанной зоны. При увеличении амплитуды с 1,5 до 1,8 кН длина зоны увеличивалась до 5,4 мм. В пределах этой зоны пластические деформации протекали не равномерно, а в локализованных малых участках.

Получена высокая корреляция относительных перемещений, определяемых двумя методами на базах измерений 66 и 470 мкм соответственно. Обоими методами обнаружено, что при увеличении амплитуды цикла накопление пластических деформаций осуществляется по механизму циклической ползучести циклически разупрочняющегося материала. Полученные результаты позволяют сделать вывод о корректности оценки деформаций новым методом. С использованием нового метода подтверждено значение предельной растягивающей деформации в зоне зарождения трещины порядка 10^{-1}.

Показано, что деформация, оцененная новым методом, на порядок превышает значение деформации, рассчитанной известным методом. Отмечено, что это связано с возникновением при многоцикловой усталости локальных участков деформации размером порядка 10^1 мкм, регистрация которых традиционными методами, по причинам большой базы измерения, невозможна.

Вследствие не регистрации неоптическими методами неразрушающего контроля многоцикловой усталости обоснована их модернизация путем разработки и использования датчиков матричного типа с размерами ячеек порядка 10^1 мкм.

Литература

1. Владимиров А.П. Динамическая спекл-интерферометрия микроскопических и макроскопических процессов в деформируемых средах // Diagnostics, Resource and Mechanics of Materials and Structures. 2015. № 6. С. 27–57. https://doi.org/10.17804/2410-9908.2015.6.027-057
2. Новиков И.И., Ермишкин В.А. Физическая механика реальных материалов. М.: Наука, 2004. 328 с.
3. Горкунов Э.С., Саврай Р.А., Макаров А.В., Задворкин С.М. Магнитные методы оценки упругой и пластической деформации при циклическом нагружении сталей // Diagnostics, Resource and Mechanics of Materials and Structures. 2015. № 2. С. 6–15. https://doi.org/10.17804/2410-9908.2015.2.006-015
4. Горкунов Э.С., Поволоцкая А.М., Задворкин С.М., Путилова Е.А., Мушников А.Н., Базулин Е.Г., Вопилкин А.Х. Особенности поведения магнитных и акустических характеристик горячекатаной стали 08Г2Б при циклическом нагружении // Дефектоскопия. 2019. № 11. С. 21–31. https://doi.org/10.1134/S0130308219110034
5. Vladimirov A.P. Speckle metrology of dynamic macro- and microprocesses in deformable media // Optical Engineering. 2016. V. 55. N 12. P. 121727. https://doi.org/10.1117/1.OE.55.12.121727
6. Vladimirov A.P. Dynamic speckle interferometry of high-cycle material fatigue: Theory and some experiments // AIP Conference Proceedings. 2016. V. 1740. P. 040004. https://doi.org/10.1063/1.4952663
7. Владимиров А.П., Друкаренко Н.А, Мызнов К.Е. Использование спекловых изображений для определения локальных пластических деформаций, возникающих при многоцикловой усталости стали 09Г2С // Письма в Журнал технической физики. 2021. Т. 47. № 15. С. 35–38. https://doi.org/10.21883/PJTF.2021.15.51232.18752
8. Владимиров А.П. Динамическая спекл-интерферометрия деформируемых тел. Екатеринбург: УрО РАН, 2004. 240 с.
9. Владимиров А.П., Друкаренко Н.А., Михайленко Ю.А. Спекл-диагностика необратимых процессов, протекающих в некоторых живых и технических объектах // Оптические методы исследования потоков: Труды XVI Международной научно-технической конференции, 28 июня — 02 июля 2021. М.: Издательство «Перо», 2021. С. 51–62 [Электронный ресурс]. URL: https://omfi-conf.ru/omfi2021/OMFI-2021-Proceedings.pdf (дата обращения: 14.12.2022).
10. Vladimirov A.P. Speckle tomography of the living-cell functions // Radiophysics and Quantum Electronics. 2021. V. 63. N 8. P. 592–604. https://doi.org/10.1007/s11141-021-10082-y
11. Vladimirov A.P. Dynamic speckle-interferometry of microscopic processes in thin biological objects // Radiophysics and Quantum Electronics. 2015. V. 57. N 8-9. P. 564–576. https://doi.org/10.1007/s11141-015-9540-2
12. Владимиров А.П., Поносов Ю.С. Применение динамики спеклов и комбинационного рассеяния света для изучения особенностей разрушения трубной стали при многоцикловой усталости // Вестник Пермского национального исследовательского политехнического университета. Механика. 2018. № 3. С. 138–146. https://doi.org/10.15593/perm.mech/2018.3.13
13. Трощенко В.Т. Усталость и неупругость металлов. Киев: Наукова думка, 1971. 268 с.
14. Coffin L.F. Symposium on Internal Stress and Fatigue of Metals. Elsevier, 1959. 451 p.
15. Klesnil M., Lukác P. Fatigue of Metallic Materials. Elsevier, 1992. 270 p.
16. Гаф Г.Дж. Усталость металлов / пер. с англ. М., Л.: Гл. ред. лит. по черн. металлургии, 1935. 304 с.
17. Ботвина Л.Р. Кинетика разрушения конструкционных материалов: учебник. М.: Наука, 1989. 230 с.
18. Голуб В.П. Исследования в области циклической ползучести материалов (обзор) // Прикладная механика. 1987. Т. 23. № 12. С. 3–19.

References

1. Vladimirov A.P. Dynamic speckle interferometry of microscopic and macroscopic processes in deformable media. *Diagnostics, Resource and Mechanics of Materials and Structures*, 2015, no. 6, pp. 27–57. (in Russian). https://doi.org/10.17804/2410-9908.2015.6.027-057
2. Novikov I.I., Ermishkin V.A. *Physical Mechanics of Real Materials*. Moscow, Nauka Publ., 2004, 328 p. (in Russian)
3. Gorkunov E.S., Savrai R.A., Makarov A.V., Zadvorkin S.M. Magnetic techniques for estimating elastic and plastic strains in steels under cyclic loading. *Diagnostics, Resource and Mechanics of Materials and Structures*, 2015, no. 2, pp. 6–15. (in Russian). https://doi.org/10.17804/2410-9908.2015.2.006-015
4. Gorkunov E.S., Povolotskaya A.M., Zadvorkin S.M., Putilova E.A., Mushnikov A.N., Bazulin E.G., Vopilkin A.K. Some features in the behavior of magnetic and acoustic characteristics of hot-rolled 08G2B steel under cyclic loading. *Russian Journal of Nondestructive Testing*, 2019, vol. 55, no. 11, pp. 827–836. https://doi.org/10.1134/S1061830919110044
5. Vladimirov A.P. Speckle metrology of dynamic macro- and microprocesses in deformable media. *Optical Engineering*, 2016, vol. 55, no. 12, pp. 121727. https://doi.org/10.1117/1.OE.55.12.121727
6. Vladimirov A.P. Dynamic speckle interferometry of high-cycle material fatigue: Theory and some experiments. *AIP Conference Proceedings*, 2016, vol. 1740, pp. 040004. https://doi.org/10.1063/1.4952663
7. Vladimirov A.P., Drukarenko N.A., Myznov K.E. Using speckle images for determining the local plastic strains arising at high-cycle fatigue of 09G2S steel. *Technical Physics Letters*, 2021, vol. 47, no. 11, pp. 777–780. https://doi.org/10.1134/S1063785021080137
8. Vladimirov A.P. *Dynamic Speckle Interferometry of Deformable Bodies*. Yekaterinburg, Ural Branch of the Russian Academy of Sciences, 2004, 240 p. (in Russian)
9. Vladimirov A.P. Drukarenko N.A., Mikhailenko Yu.A. Speckle-diagnostics of irreversible processes occurring in some living and technical objects. *Proc. of the International Conference Optical Methods of Flow Investigation (OMFI 2021) 28 June – 2 July 2021*. Available at: https://omfi-conf.ru/omfi2021/OMFI-2021-Proceedings.pdf (accessed: 14.12.2022). (in Russian)
10. Vladimirov A.P. Speckle tomography of the living-cell functions. *Radiophysics and Quantum Electronics*, 2021, vol. 63, no. 8, pp. 592–604. https://doi.org/10.1007/s11141-021-10082-y
11. Vladimirov A.P. Dynamic speckle-interferometry of microscopic processes in thin biological objects. *Radiophysics and Quantum Electronics*, 2015, vol. 57, no. 8-9, pp. 564–576. https://doi.org/10.1007/s11141-015-9540-2
12. Vladimirov A.P., Ponosov Yu.S. Application of speckle dynamics and Raman light scattering to study the fracture features of pipe steel at high-cycle fatigue. *PNRPU Mechanics Bulletin*, 2018, no. 3, pp. 138–146. (in Russian). https://doi.org/10.15593/perm.mech/2018.3.13
13. Troshchenko V.T. *Fatigue and Inelasticity of Metals*. Kiev, Naukova dumka Publ., 1971, 268 p. (in Russian)
14. Coffin L.F. *Symposium on Internal Stress and Fatigue of Metals*. Elsevier, 1959, 451 p.
15. Klesnil M., Lukác P. *Fatigue of Metallic Materials*. Elsevier, 1992, 270 p.
16. Gough H.J. *The Fatigue of Metals*. E. Benn, 1926, 304 p.
17. Botvina L.R. *Fracture Kinetics of Structural Materials*. Moscow, Nauka Publ., 1989, 230 p. (in Russian)
18. Golub V.P. Investigations into cyclic creep of materials (review). *Soviet Applied Mechanics*, 1987, vol. 23, no. 12, pp. 1107–1121. https://doi.org/10.1007/BF00884887

Авторы

Владимиров Александр Петрович — доктор технических наук, старший научный сотрудник, старший научный сотрудник, Институт машиноведения имени Э.С. Горкунова Уральского отделения Российской академии наук, Екатеринбург, 620049, Российская Федерация, sc 56204831400, https://orcid.org/0000-0001-6343-9772, vap52@bk.ru

Каманцев Иван Сергеевич — кандидат технических наук, заведующий сектором, старший научный сотрудник, Институт машиноведения имени Э.С. Горкунова Уральского отделения Российской академии наук, Екатеринбург, 620049, Российская Федерация, sc 36696097700, https://orcid.org/0000-0003-0775-4043, ks@imach.uran.ru

Друкаренко Никита Александрович — младший научный сотрудник, Институт машиноведения имени Э.С. Горкунова Уральского отделения Российской академии наук Екатеринбург, 620049, Российская Федерация, sc 57200041012, https://orcid.org/0000-0002-5979-7398, targosk@ya.ru

Мызнов Константин Евгеньевич — младший научный сотрудник, Институт физики металлов имени М.Н. Михеева Уральского отделения Российской академии наук, Екатеринбург, 620108, Российская Федерация, https://orcid.org/0009-0002-0411-1712, myznov@imp.uran.ru

Наумов Константин Вячеславович — студент, Уральский федеральный университет имени первого президента России Б.Н. Ельцина, Екатеринбург, 620002, Российская Федерация, https://orcid.org/0000-0002-5636-9857, knaumov98@mail.ru

Authors

Alexandr P. Vladimirov — D.Sc., Senior Researcher, Institute of Engineering Science, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, 620049, Russian Federation, sc 56204831400, https://orcid.org/0000-0001-6343-9772, vap52@bk.ru

Ivan S. Kamantsev — PhD, Head of Sector, Senior Researcher, Institute of Engineering Science, Ural Branch of the Russian Academy of Sciences, sc 36696097700, https://orcid.org/0000-0003-0775-4043, ks@imach.uran.ru

Nikita A. Drukarenko — Junior Researcher, Institute of Engineering Science, Ural Branch of the Russian Academy of Sciences, Ekaterinburg, 620049, Russian Federation, sc 57200041012, https://orcid.org/0000-0002-5979-7398, targosk@ya.ru

Konstantin E. Myznov — Junior Researcher, Institute of Metal Physics, Ural Branch of Russian Academy of Sciences, Ekaterinburg, 620108, Russian Federation, https://orcid.org/0009-0002-0411-1712, myznov@imp.uran.ru

Konstantin V. Naumov — Student, Ural Federal University named after the First President of Russia B.N.Yeltsin, Ekaterinburg, 620002, Russian Federation, https://orcid.org/0000-0002-5636-9857, knaumov98@mail.ru

Статья поступила в редакцию 14.09.2023
Одобрена после рецензирования 12.12.2023
Принята к печати 20.01.2024

Received 14.09.2023
Approved after reviewing 12.12.2023
Accepted 20.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-30-40
УДК 538.975

Лазерно-индуцированное тепловое воздействие на электрические характеристики фоточувствительных пленок селенида свинца

Анастасия Александровна Ольхова[1], Алина Александровна Патрикеева[2], Мария Алексеевна Бутяева[3], Александра Евгеньевна Пушкарева[4], Екатерина Александровна Авилова[5], Михаил Константинович Москвин[6], Максим Михайлович Сергеев[7]✉, Вадим Павлович Вейко[8]

[1,2,3,4,5,6,7,8] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] olkhova.a.a@mail.ru, https://orcid.org/0000-0001-9048-3031
[2] patrikeeva17@gmail.com, https://orcid.org/0000-0002-5274-9692
[3] maria.dubkova@mail.ru, https://orcid.org/0000-0003-3251-6602
[4] alexandra.pushkareva@gmail.com, https://orcid.org/0000-0003-0082-984X
[5] eaavilova@itmo.ru, https://orcid.org/0000-0002-4520-6761
[6] mkmoskvin@itmo.ru, https://orcid.org/0000-0001-7399-7022
[7] maxim.m.sergeev@gmail.com✉, https://orcid.org/0000-0003-2854-9954
[8] vpveiko@itmo.ru, https://orcid.org/0000-0001-6071-3449

Аннотация

Введение. В работе представлено исследование влияния лазерного облучения кристаллических халькогенидных пленок селенида свинца на их электрические характеристики, вызванные необратимой модификацией структуры за счет валентной переконфигурации свинца в результате его окисления. **Метод.** Исследование особенностей модификации электрических свойств пленок выполнено в результате лазерного воздействия наносекундных импульсов с длиной волны 1064 нм. Измерения электрических характеристик пленок селенида свинца проводились с помощью четырехзондового метода. **Основные результаты.** Показано, что при направлении тока параллельно лазерным трекам, записанным в режиме потемнения, сопротивление модифицированной пленки снизилось на 44 % по сравнению с исходным образцом, а при перпендикулярном направлении тока сопротивление возросло на 153 %. Сопротивление пленки увеличилось более чем в 27 раз после лазерного облучения в режиме просветления вне зависимости от направления тока относительно лазерных треков. Экспериментально измеренные температура и ее градиент по лазерному пятну на пленке в режимах потемнения и просветления оказались в хорошем соответствии с предложенной математической моделью теплового воздействия лазерных импульсов. Показано, что процессы лазерной модификации пленок происходят при более низких температурах, чем при стандартной тепловой обработке в печи. **Обсуждение.** Полученные результаты могут быть применены при разработке фотодетекторов в среднем инфракрасном диапазоне спектра на основе пленки селенида свинца.

Ключевые слова

лазерная модификация, пленки PbSe, оптические характеристики, режим потемнения, режим просветления, термообработка, лазерные импульсы

Благодарности

Работа выполнена за счет средств гранта Российского научного фонда и гранта Санкт-Петербургского научного фонда (проект № 23-29-10081).

Исследование структуры методом СЭМ-электронной микроскопии проведено в Центре нанотехнологий Научного парка Санкт-Петербургского государственного университета в рамках проекта № АААА-А19-119091190094.

Ссылка для цитирования: Ольхова А.А., Патрикеева А.А., Бутяева М.А., Пушкарева А.Е., Авилова Е.А., Москвин М.К., Сергеев М.М., Вейко В.П. Лазерно-индуцированное тепловое воздействие на электрические характеристики фоточувствительных пленок селенида свинца // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 30–40. doi: 10.17586/2226-1494-2024-24-1-30-40

© Ольхова А.А., Патрикеева А.А., Бутяева М.А., Пушкарева А.Е., Авилова Е.А., Москвин М.К., Сергеев М.М., Вейко В.П., 2024

Laser-induced thermal effect on the electrical characteristics of photosensitive PbSe films

Anastasiia A. Olkhova[1], Alina A. Patrikeeva[2], Maria A. Butyaeva[3], Alexandra E. Pushkareva[4], Ekaterina A. Avilova[5], Mikhail K. Moskvin[6], Maksim M. Sergeev[7]✉, Vadim P. Veiko[8]

[1,2,3,4,5,6,7,8] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] olkhova.a.a@mail.ru, https://orcid.org/0000-0001-9048-3031
[2] patrikeeva17@gmail.com, https://orcid.org/0000-0002-5274-9692
[3] maria.dubkova@mail.ru, https://orcid.org/0000-0003-3251-6602
[4] alexandra.pushkareva@gmail.com, https://orcid.org/0000-0003-0082-984X
[5] eaavilova@itmo.ru, https://orcid.org/0000-0002-4520-6761
[6] mkmoskvin@itmo.ru, https://orcid.org/0000-0001-7399-7022
[7] maxim.m.sergeev@gmail.com✉, https://orcid.org/0000-0003-2854-9954
[8] vpveiko@itmo.ru, https://orcid.org/0000-0001-6071-3449

Abstract

The paper presents a study of the effect of laser irradiation of crystalline chalcogenide films of lead selenide (PbSe) on their electrical characteristics caused by irreversible modification of the structure due to valence reconfiguration of lead as a result of its oxidation. The study of the modification features of the electrical properties of the films was carried out because of laser exposure to nanosecond pulses with a wavelength of 1064 nm. Measurements of the electrical characteristics of PbSe films were carried out using the four-probe method. It was shown that when the current was directed parallel to the laser tracks recorded in the darkening mode, the resistance of the modified film decreased by 44 % compared to the original sample, and with the perpendicular direction of the current, the resistance increased by 153 %. The resistance of the film increased more than 27 times after laser irradiation in the bleaching mode, regardless of the direction of the current relative to the laser tracks. The experimentally measured temperature and its gradient along the laser spot on the film in the darkening and bleaching modes turned out to be in good agreement with the proposed mathematical model of the thermal effect of laser pulses. It has been shown that the processes of laser modification of films occur at lower temperatures than during standard heat treatment in a furnace. The obtained results can be applied in the development of photodetectors in the middle IR range of the spectrum based on PbSe film.

Keywords

laser modification, PbSe films, optical characteristics, darkening mode, bleaching mode, heat treatment, laser pulses

Acknowledgements

This research was funded by the Russian Science Foundation grant and a grant from the Saint Petersburg Science Foundation (project no. 23-29-10081).

The investigation of the structure by means SEM electron microscopy was carried out at the IRC for Nanotechnology of the Science Park of Saint Petersburg State University within the framework of project No. AAAA-A19-119091190094.

For citation: Olkhova A.A., Patrikeeva A.A., Butyaeva M.A., Pushkareva A.E., Avilova E.A., Moskvin M.K., Sergeev M.M., Veiko V.P. Laser-induced thermal effect on the electrical characteristics of photosensitive PbSe films. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 30–40 (in Russian). doi: 10.17586/2226-1494-2024-24-1-30-40

Введение

В работе приводятся исследования элементов датчиков, использующих чувствительные элементы на основе халькогенидов (в настоящей работе исследован селенид свинца (PbSe)). Эти датчики служат основой аналитических приборов в области газового анализа, микроаналитики и фотовольтаики. Одним из самых ответственных процессов при производстве таких пленок является их отжиг для увеличения чувствительности, который обычно осуществляют в муфельных печах. Данный способ не позволяет обеспечить необходимую повторяемость параметров, поэтому был предложен метод лазерной модификации пленок. На основе данного метода выполнена модернизация существующего производственного цикла с целью повышения качества сенсоров и сокращения трудозатрат.

Для создания фотодетекторов в среднем инфракрасном (ИК) диапазоне спектра и солнечных элементов широко применяются PbSe пленки, которые обладают высокой поглощающей способностью в среднем ИК (1–4 мкм) [1, 2]. Эти особенности также представляют интерес для микроаналитики органических молекул в виде газовой смеси или жидкости [3, 4]. В данном спектральном диапазоне находятся пики поглощения оксида углерода, углекислого газа и оксида азота, а также различных углеводородов, включая пары топлива и анестезирующих газов, которые являются элементами на основе пленок PbSe.

Пленки PbSe применяются также в качестве фоторезисторов [5], их спектральная чувствительность лежит в диапазоне от 0,5 до 5,0 мкм. Важным свойством фоторезисторов из соединений свинца является смещение спектральной чувствительности в длинноволновую область спектра при охлаждении материала, что расширяет потенциальную область их применения. PbSe по сравнению с теллуридами свинца и германия является более высокотемпературным термоэлектриком [6].

Одна из наиболее важных характеристик, используемых в устройствах газового анализа для определения наличия и концентрации газа — соотношение сигнал/шум, которое связано с чувствительностью детектора и его электрическим сопротивлением. Работы различных научных групп, основанные на измерениях отража-

тельной способности при комнатной температуре в ИК диапазоне от 1 до 5 мкм [7] и на измерениях поглощения в диапазоне от 4,5 до 9,0 мкм [8], показали, что граница полосы поглощения PbSe находится около 5,0 мкм [9]. В таком случае PbSe является фотопроводящим материалом в спектральном диапазоне до 5,0 мкм. В работе [10] отмечена закономерность, в которой уровень чувствительности в толстых пленках (от 2 мкм) был намного ниже, чем в более тонких пленках (до 1 мкм).

В настоящей работе исследованы особенности изменения электрических и фотоэлектрических характеристик пленок PbSe в результате лазерного облучения наносекундными импульсами ближнего ИК излучения. В эксперименте использована методика измерения поверхностного сопротивления пленок, регламентируемая ГОСТ IEC/TS 62607-2-1-2017[1] [11]. Испытания проводились с помощью четырехзондового метода, основанного на измерении поверхностного сопротивления без и при засветке образца ИК излучением.

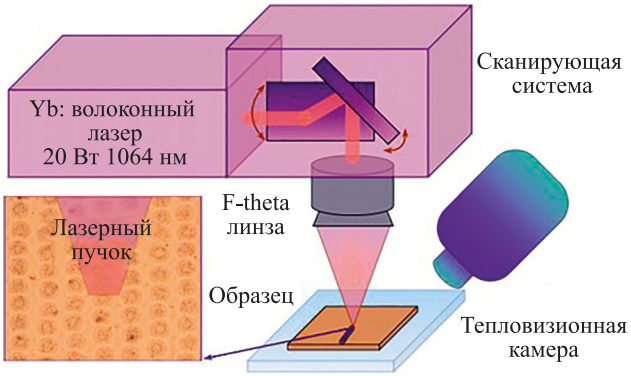


Рис. 1. Схема экспериментальной установки для лазерной модификации структуры пленок селенида свинца с помощью импульсного инфракрасного излучения

Fig. 1. Scheme of the experimental setup for PbSe film structure laser modification using pulsed infrared radiation

Материалы и методы

В работе применен метод вакуумно-термического напыления для изготовления исследуемых пленок на плоскопараллельной подложке из покровного стекла толщиной 0,2 мм. Образцы пленки изготовлены компанией ООО «Оптосенс». Традиционная тепловая обработка исходных образцов выполнена в печи открытого типа при атмосферных условиях. Термическая обработка продолжалась в течение 90 с и представляла собой нагрев пленки до температуры 540 °C с последующей выдержкой и нагревом до 630 °C. Данная обработка состояла из процессов активации и сенсибилизации образцов [12].

Для лазерной модификации структуры пленки использовался наносекундный Yb:волоконный лазер со средней мощностью 20 Вт, длиной волны 1064 нм. Движение лазерного пятна по заданной траектории в плоскости обработки осуществлялось двухзеркальным гальванометрическим сканатором. После зеркал сканатора лазерный пучок фокусировался F-theta линзой с полем обработки 100 × 100 мм (рис. 1).

Обработка пленок происходила в сходящемся лазерном пучке, формируемым F-theta линзой. Максимальная импульсная плотность мощности падающего на пленку лазерного излучения в лазерном пятне с диаметром 70–90 мкм достигла 18,2 МВт/см^2 при длительности импульсов от 4 до 50 нс и частоте их следования от 5 до 120 кГц. Скорость лазерного пятна изменялась от 50 до 150 мм/с.

В табл. 1 представлены режимы просветления и потемнения при модификации образца с помощью импульсного лазерного источника. Просветление и потемнение — нелинейно-оптические эффекты, просветление связано с уменьшением показателя поглощения среды на определенной длине волны излучения, а потемнение — с увеличением. Режим потемнения сопровождался непосредственным (визуальным) потемнением пленки PbSe в области лазерного воздействия и наблюдался при уменьшении плотности мощности излучения относительно режима просветления. Изменение визуального восприятия пленки после лазерной обработки связано с ее нагревом, при этом потемнение образца появлялось при меньшей температуре по сравнению с просветлением. Под оптимальным режимом подразумевается режим, при котором лазерная модификация не вызывает разрушение образца, но при этом изменяет его оптические свойства, приводя к потемнению/просветлению пленки PbSe. Регистрация температуры и ее градиента теплового источника в зоне облучения осуществлялось с помощью тепловизионной камеры FLIR Titanium 520M (FLIR Systems, Inc., Wilsonville, OR), обладающей калибровочным диапазоном температуры от 20 до 400 °C без высокотемпературного фильтра. Для изучения распределения температуры от небольшого теплового источника применялась германиевая линза с фокусным расстоянием 50 мм, фокальная плоскость которой была выровнена с плоскостью пленки в области, где размещалось лазерное пятно, при этом камера находилась под углом около 45° относительно плоскости образца. Во время измерения температуры тепловизионная камера находилась на расстоянии 5 см от образца, измерения проводились в фокальной плоскости объектива камеры.

Измерение электрических и фотоэлектрических характеристик. Измерения электрических характеристик пленок PbSe выполнено на облученных участках пленки и на исходной пленке. Измерение поверхностного сопротивления осуществлено с помощью четырехзондового метода [11] (рис. 2). Метод измерения заключался в том, что через зонды *1* и *4* проходит ток, в то время как между зондами *2* и *3* измеряется разность потенциалов. В схеме использован четырехзондовый щуп (ST2253-F01 Suzhou Jingge Electronic Corporation) с фиксированным расстоянием между щупами *b* равным 1 мм. Расстояние от ИК лампы до образца составило 10 мм. Для визуализации процесса измерения и кон-

[1] ГОСТ IEC/TS 62607-2-1-2017. Производство нанотехнологическое. Контроль основных характеристик. Часть 2-1. Материалы из углеродных нанотрубок. Методы определения поверхностного сопротивления. Введен 01.09.2018. М. Стандартинформ, 2017. 9 с.

Таблица 1. Режимы модификации пленки PbSe с помощью импульсного лазерного источника с длиной волны 1064 нм
Table 1. Modification modes of chalcogenide film using a pulsed laser source with a wavelength of 1064 nm

Режим		Параметры					
		P_{av}, мВт	f, кГц	v, мм/с	t, нс	q, МВт/см^2	d, мкм
Потемнение	Порог	10	5	150	4	6,4	100
	Оптимальное	14	5	100	4	18,2	70
	Разрушение	14	5	65	4	18,2	70
Просветление	Порог	970	120	50	50	0,9	150
	Оптимальное	930	120	50	50	2,4	90
	Разрушение	700	50	50	50	4,4	90

Примечание. P_{av} — средняя мощность падающего излучения; f — частота следования импульсов; v — скорость сканирования; t — длительность импульсов; q — импульсная плотность мощности; d — диаметр лазерного пятна на пленке PbSe

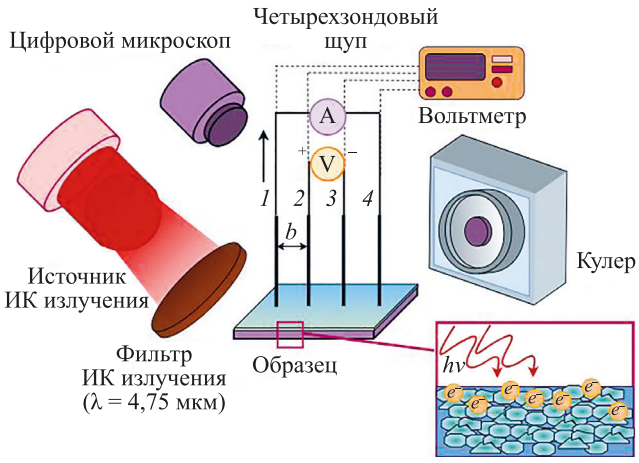


Рис. 2. Схема экспериментальной установки для измерения электрических характеристик.
hv — энергия кванта; e — заряд электрона

Fig. 2. Diagram of the experimental setup for measuring the electrical characteristics of chalcogenide films (hv — quantum energy, e — electron charge)

троля контакта между образцом и щупами установка оснащена цифровым микроскопом. Измерение электрических характеристик проведено при экспонировании анализирующим источником ИК излучением лампы (G4 12V 10W MILBLOON) в спектральном диапазоне 0,7–11,0 мкм и без источника, с целью первичной оценки фоточувствительности образцов. Для селектирования широкополосного излучения ИК лампы применен ИК фильтр с полосой пропускания на длине волны $\lambda = 4{,}75$ мкм. Термостабилизация образца получена путем конвекции при контакте нагреваемой поверхности с воздухом с помощью вентилятора (кулера).

Микроскопия. Для оценки поверхности пленок PbSe с помощью сканирующей электронной микроскопии использован сканирующий электронный микроскоп Merlin Zeiss (Германия) с разрешением 0,6 нм.

Результаты и обсуждение

Электрическая характеризация. Электрическая характеризация пленок PbSe проведена до и после лазерной модификации структуры образца, а также после тепловой обработки в печи.

Измерение электрических характеристик после лазерной обработки выполнено при параллельных и перпендикулярных расположениях зондов относительно лазерных треков. Полученные результаты представлены на рис. 3. Температура нагрева образцов при экспонировании анализирующим источником ИК излучением стабилизировалась с помощью вентилятора во всех экспериментах, а значит, измерение сопротивления образцов происходило при близких условиях. Мощность ИК излучения изменялась в пределах от 0 до 25 мВт. Измерение поверхностного сопротивления определено в Ом/м^2, однако данные результаты не были нормированы на какой-либо эталон, поэтому в настоящей работе приведены результаты измерений поверхностного сопротивления в процентах.

В процессе экспонирования анализирующим источником ИК излучения пленок PbSe получен небольшой рост температуры, который обычно сопровождается ростом дрейфовой скорости носителей заряда [13], что в результате приводит к росту сопротивления пленки [14]. Как было отмечено в разделе «Введение», наиболее важной характеристикой является соотношение сигнал/шум, которое напрямую связано с чувствительностью детектора и его сопротивлением.

В работе [10] отмечено, что наибольшая чувствительность детектора находится около максимума сопротивления [10]. При этом чувствительность на коротких длинах волн (от 1 до 3 мкм) заметно больше в тонкой пленке (толщиной до 1 мкм), чем в толстой пленке (толщиной более 2 мкм). Отметим [10], что чувствительность в длинноволновой области спектра существенно не меняется при увеличении толщины пленки. Поскольку уровень шума уменьшается с увеличением толщины пленки, то для получения наилучшей чувствительности необходимо использовать толстые пленки (толщиной от 1 мкм).

Появление оксидных фаз на поверхности образца [15], связанное с нагревом пленок PbSe в температурном диапазоне 473–673 К, привело к уменьшению концентрации носителей заряда и их подвижности в окисленной зоне, что способствует росту их темнового сопротивления (рис. 3, *a*, *b*) [15]. Увеличение сопротив-

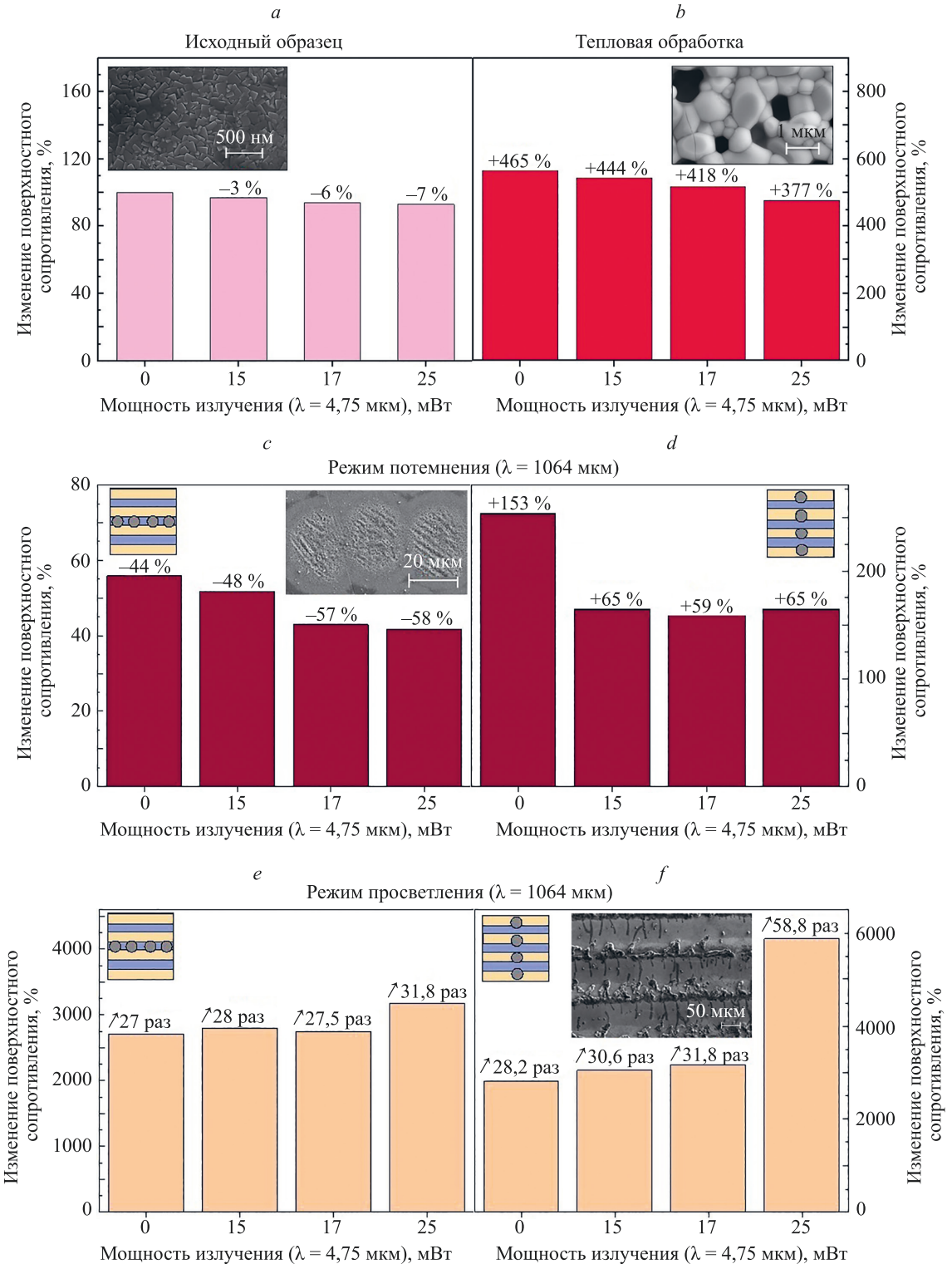


Рис. 3. Поверхностные сопротивления пленки без и во время экспонирования (*a*, *b*) для «потемненной» (*c*, *d*) и «просветленной» (*e*, *f*) пленки при λ = 1064 нм во время экспонирования анализирующим источником инфракрасного излучения лампы G4 12V 10W MILBLOON различной мощности: до модификации (*a*) и после термообработки (*b*) образца; зонды расположены параллельно (*c*, *e*) и перпендикулярно (*d*, *f*) лазерным трекам.

Значение 100 % соответствует исходной пленке без экспонирования анализирующим источником, значения над столбцами отсчитываются относительно исходной пленки без экспонирования

Fig. 3. Surface resistances of the film without and during exposure (*a*, *b*), “darkened” (*c*, *d*) and “cleared” (*e*, *f*) film during exposure to the analyzing source of infrared radiation of the G4 12V 10W MILBLOON lamp of various powers: before modification (*a*) and after heat treatment (*b*); the probes are located parallel (*c*, *e*) and perpendicular (*d*, *f*) to the laser tracks.

The value of 100 % corresponds to the original film without exposure by the analyzing source; the values above the columns are relative to the original film without exposure

ления пленки в связи с образованием оксидной фазы наблюдалось также при исследовании у образцов, подвергшихся лазерной обработке в режиме просветления (рис. 3, *e*, *f*). Данная тенденция может быть объяснена тем, что в центре лазерного пятна температура образца достигала 353 °С (рис. 5).

Как видно из рис. 3, *c*, *d*, в результате лазерного воздействия на пленку в режиме потемнения снижается ее сопротивление, что может быть связано с модификацией ионно-ковалентных и межмолекулярных связей в структуре пленки PbSe при нагревании. В результате происходит повышение фоточувствительности пленок PbSe. При параллельном расположении зондов сопротивление модифицированной пленки меньше на 44 %, чем у исходного образца, а при перпендикулярном — больше в 16 раз. Отметим, что большее изменение сопротивления при экспонировании анализирующим источником ИК излучения наблюдалось при перпендикулярном расположении зондов. Полученные результаты подтвердили связь между сопротивлением и направлением движения носителей электрического заряда перпендикулярно или параллельно лазерным трекам. Увеличение мощности ИК излучения при экспонировании анализирующим источником пленки привело к увеличению фототока и уменьшению сопротивления образца.

При модификации в режиме просветления (рис. 3, *e*, *f*) произошел значительный рост сопротивления за счет образования оксидной фазы, обладающей большим сопротивлением [16]. Оксид свинца имеет меньшую концентрацию электронов проводимости по сравнению с основным материалом пленки PbSe. В режиме просветления с ростом мощности ИК излучения растет сопротивление, это связано с образованием оксида свинца и его значимой толщиной (массовой долей) в пленке. Оксид свинца является диэлектриком и, таким образом, он не участвует в фотопроводимости при используемой интенсивности облучения.

Как и в случае модификации в режиме потемнения видно, что движение зарядов перпендикулярно лазерным трекам сильно затруднено, по сравнению с их движением вдоль треков, поэтому значение сопротивления при параллельном расположении зондов меньше, чем при перпендикулярном. Этот факт подтверждает, что на движение свободных носителей заряда оказывает влияние не только перестройка структуры в пределах лазерного трека, но также рельеф пленки и ее толщина. Кроме того, возможно образование локальных зон с различной концентрацией носителей заряда, что также может способствовать увеличению сопротивления.

В работе [14] показано, что лазерное воздействие вызывает значительное снижение отражения пленок PbSe в видимом диапазоне спектра в результате их просветления [14]. Проводимость пленки при просветлении снижается из-за образования диэлектрического слоя оксида. Таким образом, значительно возрастает (более, чем в 2,1 раза) чувствительность регистрации зондирующего излучения.

Лазерное нагревание поликристаллического PbSe приводит к образованию стекловидного аморфного слоя оксида свинца на поверхности пленки. Удельный объем аморфной фазы больше, чем у исходной кристаллической пленки, в результате чего на границе двух фаз: PbSe и оксида свинца [17] возникают механические напряжения, которые могут иметь как термомеханическую, так и термоструктурную природу. Когда величина таких напряжений превышает разрушающие значения, в пленке PbSe образуются трещины и другие дефекты, которые могут приводить к деградации и дестабилизации ее характеристик во времени (рис. 4).

Однако в процессе термообработки пленки PbSe в печи происходят кардинально другие процессы, связанные с ростом оксидной фазы, что соответствует данным из патента № WO 2013/154462 A2 [12]. Медленное, по сравнению с лазерным облучением, тепловое воздействие в печи открытого типа чаще приводит к растрескиванию образовавшейся аморфной структуры пленки за счет увеличения продолжительности действия напряжений [14]. Возможность направленного проведения таких индуцированных локальных изменений структуры и состава пленок PbSe при коротких (1–100 нс) временах протекания лазерной модификации открывают новые пути использования технологии их структурно-фазовой перестройки в приложениях газового анализа.

Характеризация теплового источника

Лазерная модификация полупроводниковых пленок при низких температурах имеет свои особенности, отличающиеся от процессов, происходящих при высоких температурах. При температуре пленки порядка 300 °С, нагреваемой в пределах лазерного пятна, происходит локализованное воздействие на поверхность материала, что позволяет получить микроскопические модификации и сохранить целостность пленки. Данная температура нетипична для тепловой обработки пленок PbSe и значительно ниже температуры активации структурно-фазовых переходов, реализуемых стандартным нагреванием в печи. Лазерная обработка в режиме просветления также обладает высокими градиентом температуры 1,8 К/мкм и скоростью нагрева 4,5 К/мс.

Один из основных физических процессов, происходящих в полупроводниковых пленках при таком лазерном воздействии — нагревание. При низкой температуре относительно термообработки, а также за счет малых размеров лазерного пятна, нагрев ограничен небольшим объемом материала, что позволяет проводить точную локализацию и контроль над модификацией [18].

Кроме того, низкая температура нагреваемой области в пределах лазерного пятна позволяет управлять фазовыми переходами, а именно, аморфизацией пленки и перестройкой конфигурации структуры в полупроводниковых пленках. Это позволяет получать новые физические свойства и улучшать характеристики материала. Природа возникновения модификаций при низкой температуре нагрева пленки связана с точечным нагревом поверхности материала.

В ходе модификации импульсным излучением (рис. 5) это значение не превышало 353 °С для режима просветления и 101 °С для режима потемнения.

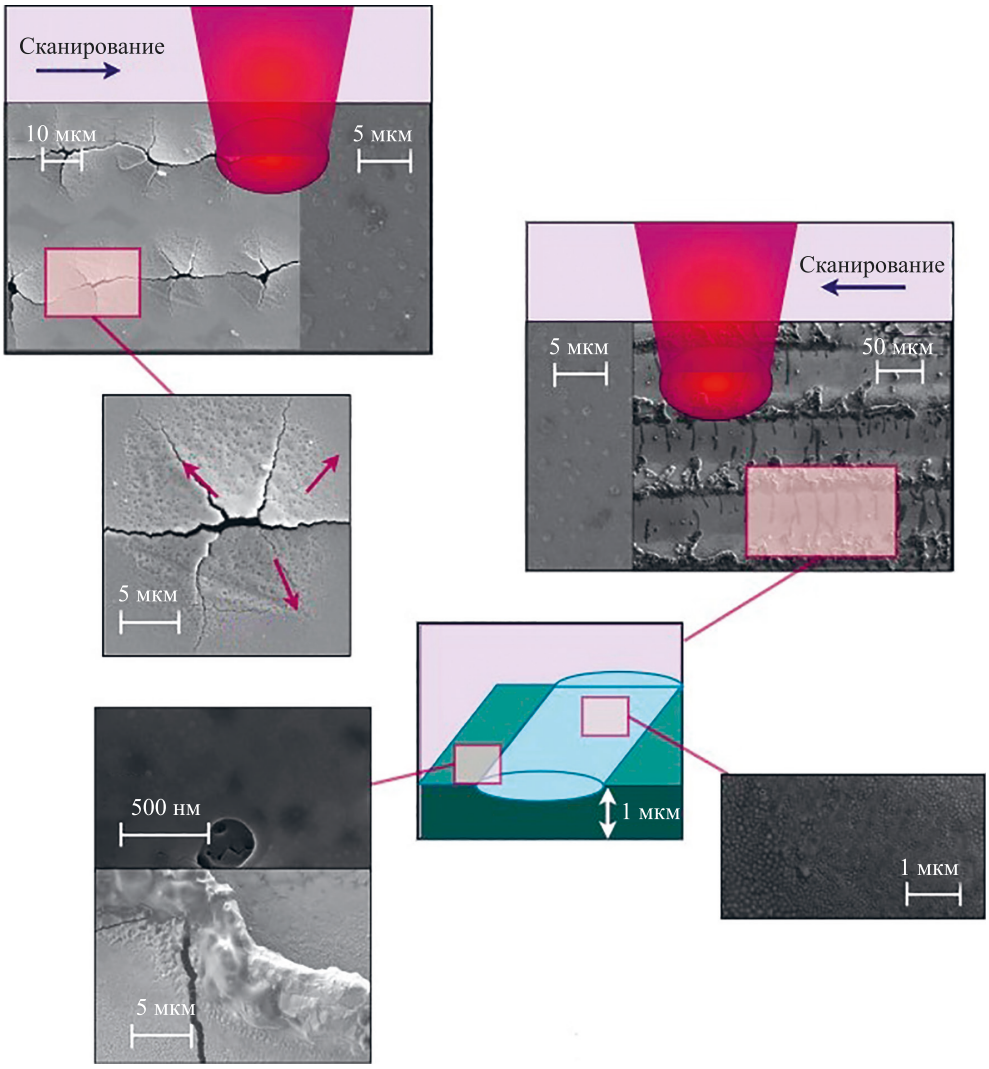


Рис. 4. Механизм возникновения дефектных центров в области модификации пленки селенида свинца

Fig. 4. Defect centers formation mechanism in the area of lead selenide film modification

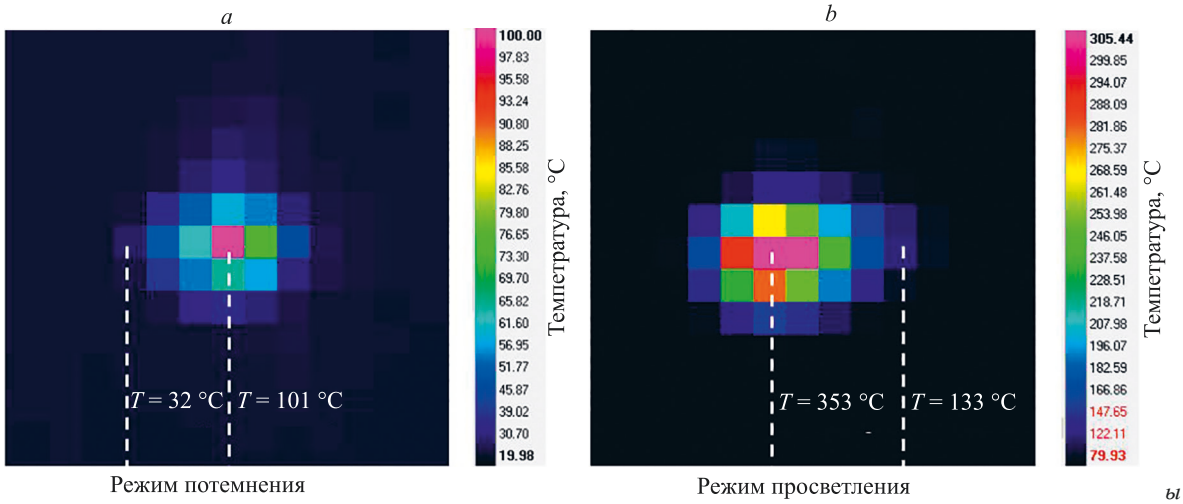


Рис. 5. Тепловизионные изображения лазерного пучка при лазерной модификации импульсным излучением в режимах: потемнения (*a*) и просветления (*b*)

Fig. 5. Thermal images of a laser beam during laser modification with pulsed radiation in the following modes: darkening (*a*) and bleaching (*b*)

Процессы, связанные с просветлением пленки, как и предполагалось, протекали при более высоких температурах, по сравнению с потемнением пленки. Так как на начальном этапе воздействия пленка темнела, что приводило к увеличению ее поглощения, доля поглощенной энергии лазерного излучения со временем увеличивалась. В результате пленка нагревалась до более высоких температур к образованию оксидной фазы в зоне воздействия. При этом лазерная модификация в режиме потемнения также характеризовалась высокой скоростью нагрева 2,3 К/мс и градиентом температуры 0,8 К/мкм, в 2,4 раз меньшим, чем при обработке в режиме просветления.

Таким образом, при лазерной модификации полупроводниковых пленок следует отдавать предпочтение низкотемпературным процессам. Например, использовать структурно-фазовые переходы в пленке PbSe, включая образование оксидов свинца в аморфном состоянии, возникающие вследствие теплового (фототермического) воздействия лазерного излучения в пятне ограниченных размеров. В результате можно получить экстремально высокие градиенты температур, скорости нагревания и охлаждения.

Возможность создания микроскопических модификаций с размером модифицированных областей до 100 мкм — одно из ключевых преимуществ лазерной модификации при низкой температуре относительно термообработки.

Моделирование теплового источника

Для сопоставления измеренных значений температуры и моделирования ее распределения в области воздействия используем физико-математическую модель теплового источника, возникающего от серии коротких [19]. Модель описывает мгновенный приток тепла во время импульса и отток тепла между импульсами по каждой координате (x, y, z):

$$T(x, y, z, t) = (A_f E_p)/(64 c_f \rho_f \omega^2 l) \sum_{i=0}^{N} \times$$
$$\times[(\mathrm{erfc}\{(x + iv/f - \omega)/2\sqrt{a(i/f + t)}\} -$$
$$- \mathrm{erfc}\{(x + iv/f + \omega)/2\sqrt{a(i/f + t)}\})\times$$
$$\times(\mathrm{erfc}\{(y - \omega)/2\sqrt{a(i/f + t)}\} - \quad (1)$$
$$- (\mathrm{erfc}\{(y + \omega)/2\sqrt{a(i/f + t)}\}\times$$
$$\times(\mathrm{erfc}\{(z - l)/2\sqrt{a(i/f + t)}\} -$$
$$- (\mathrm{erfc}\{(z + l)/2\sqrt{a(i/f + t)}\}],$$

где T — температура; t — время; A_f — поглощательная способность пленки; E_p — энергия лазерного импульса; ρ_f и c_f — плотность и удельная теплоемкость пленки; ω — радиус пятна; v — скорость движения лазерного источника; l — толщина прогреваемого слоя; f — частота следования импульсов; a — температуропроводность материала, которая определяется как

$$a = \kappa_{air}/(\rho_{air} c_{air}) \text{ при } z < -l_f,$$
$$a = \kappa_f/(\rho_f c_f) \text{ при } -l_f \le z \le l_f,$$
$$a = \kappa_{sub}/(\rho_{sub} c_{sub}) \text{ при } z > l_f,$$

где κ_f — значение теплопроводности пленки; ρ_{air}, c_{air} и κ_{air} — плотность, удельная теплоемкость и теплопроводность окружающей среды (воздух); ρ_{sub}, c_{sub} и κ_{sub} — плотность, удельная теплоемкость и теплопроводность подложки.

Выражение (1) для расчета динамики температуры позволяет описывать динамику быстрого нагрева материала с последующим охлаждением, что соответствует процессу нагрева короткими (1–50 нс) лазерными импульсами с частотами их следования до 120 кГц.

Модель разработана с учетом следующих приближений:

— излучение лазера поглощалось материалом пленки и не поглощалось материалом подложки;
— отведение тепла от нагретой области происходило в окружающую среду (воздух) и подложку (плавленый кварц) посредством теплопроводности и теплоотдачи на границах (коэффициент теплоотдачи на границе пленка-воздух принят равным 25 Вт/(м^2·К), на границе пленка-подложка — 100 Вт/(м^2·К));
— распределение энергии в пучке равномерное;
— поглощательная способность пленки и теплофизические характеристики материалов не зависят от температуры.

На рис. 6 приведены зависимости температуры от количества импульсов в центре пятна при лазерной обработке в режиме просветления и потемнения. С увеличением количества импульсов происходит рост температуры. При 50 импульсах смоделированные значения температуры теплового источника для режимов потемнения и просветления близки к измеренным с помощью

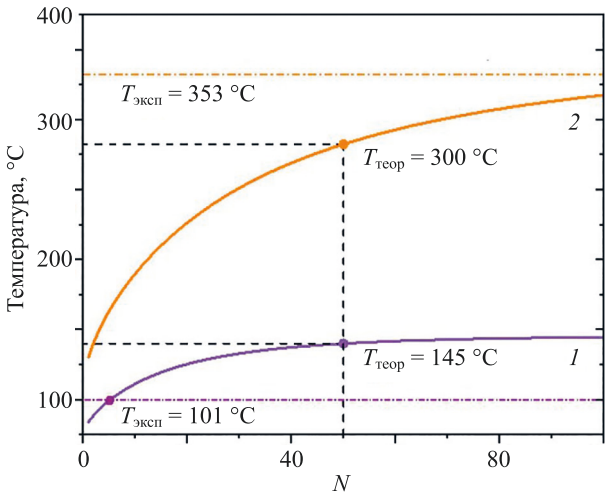


Рис. 6. Температура в центре лазерного пятна на поверхности пленки в зависимости от количества импульсов N при модификации в режимах потемнения (кривая *1*) и просветления (кривая *2*)

Fig. 6. Temperature in the center of the laser spot on the film surface depending on the number of pulses (N) during modification: in darkening mode (purple curve), in bleaching mode (orange curve)

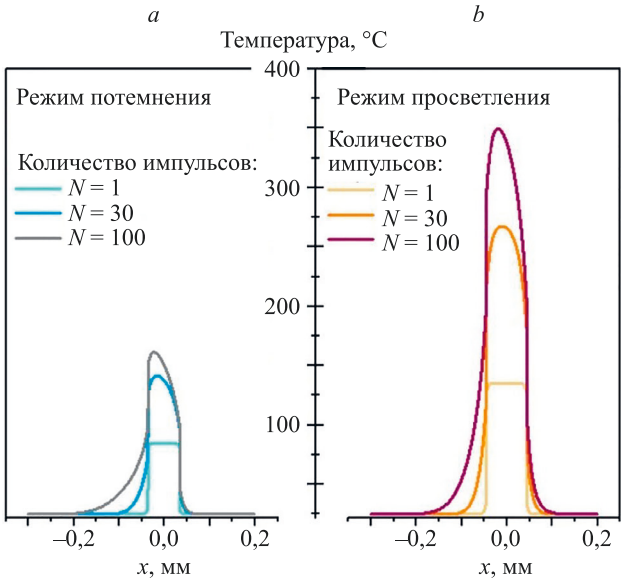


Рис. 7. Температура на поверхности пленки в зависимости от координаты *x*, для импульсов *N* = 1, 30, 100 при модификации в режимах потемнения (*a*) и просветления (*b*)

Fig. 7. Temperature on the film surface vs. the *x* coordinate, for pulses *N* = 1, 30, 100 when modified in the darkening (*a*) and clearing (*b*) modes

тепловизора (рис. 5). В режиме потемнения температура на поверхности пленки в области лазерного пятна составляет 145 °C, в режиме просветления 300 °C. Незначительное расхождение данных, полученных при моделировании и реальных измерениях, может быть связано с тем, что при моделировании не учитывалась нелинейность нагревания материала образца.

На рис. 7 показано распределение температуры на поверхности пленки по координате *x* (по направлению движения лазерного пучка) при различном количестве импульсов 1, 30, 100.

Полученные кривые позволяют выбрать оптимальные режимы лазерной обработки и подтверждают прямую корреляцию между количеством импульсов от лазерного источника и температурой в области модификации образца в выбранном диапазоне параметров.

Таким образом, при воздействии импульсным лазерным излучением относительно малой мощности до 1 Вт, при длительности импульсов воздействия 1–50 нс и частоте их следования 5–120 кГц, можно достаточно точно контролировать температуру в зоне воздействия, скорости нагрева и охлаждения, градиенты температурного поля, что позволяет управлять термоструктурными и термохимическими процессами в пленках PbSe. Данное преимущество особенно важно при обработке пленок толщиной до 500 нм, где повышенная температура может вызывать дефекты и деградацию структуры.

Заключение

В результате проведенных исследований выявлено, что лазерное воздействие в режиме потемнения приводит к снижению сопротивления, что проявляется в увеличении чувствительности модифицированных пленок. Более значительные изменения сопротивления при экспонировании анализирующим источником (инфракрасной лампой) наблюдались при перпендикулярном расположении зондов. Обнаружено, что увеличение мощности анализирующего источника приводит к уменьшению сопротивления образца, модифицированного в режиме потемнения, что обеспечивает увеличение соотношения сигнал/шум.

При модификации в режиме просветления отмечено значительное увеличение сопротивления в результате образования аморфной фазы оксида свинца, что проявляется в повышении фоточувствительности образца.

Важно отметить, что движение зарядов перпендикулярно лазерным трекам затруднено по сравнению с их движением параллельно трекам, что объясняет наблюдаемые различия в значениях сопротивления при параллельном и перпендикулярном расположениях зондов. При параллельном расположении зондов сопротивление модифицированной пленки оказалось меньше на 44 % по сравнению с исходным образцом, а при перпендикулярном — больше на 153 % для образца, модифицированного в режиме потемнения. В случае обработки в режиме просветления, вне зависимости от расположения зондов, сопротивление пленки увеличивалось более чем в 27 раз.

Использование импульсного лазерного источника небольшой мощности (до 25 мВт) является перспективным подходом для точной и неразрушающей модификации полупроводниковых пленок. Контролируемая температура обработки в диапазоне до 353 °C минимизирует тепловое воздействие и позволяет создавать микро- и наноструктуры с высокой пространственной разрешающей способностью до 100 лин/мм. Такой подход имеет большой потенциал для применения в области разработки новых устройств и технологий, требующих высокой точности и надежности при модификации полупроводниковых пленок. Дальнейшие исследования и оптимизация параметров облучения могут дополнительно расширить возможности и эффективность лазерной обработки с низкой температурой.

Литература

1. Tan C.L., Mohseni H. Emerging technologies for high performance infrared detectors // Nanophotonics. 2018. V. 7. N 1. P. 169–197. https://doi.org/10.1515/nanoph-2017-0061
2. Karim A., Andersson J.Y. Infrared detectors: Advances, challenges and new technologies // IOP Conference Series: Materials Science and Engineering. 2013. V. 51. N 1. P. 012001. https://doi.org/10.1088/1757-899x/51/1/012001
3. Kasiyan V., Dashevsky Z., Schwarz C.M., Shatkhin M., Flitsiyan E., Chernyak L., Khokhlov D. Infrared detectors based on semiconductor p-n junction of PbSe // Journal of Applied Physics. 2012. V. 112. N 8. P. 086101. https://doi.org/10.1063/1.4759011
4. Weng B., Qiu J., Yuan Z., Larson P.R., Strout G.W., Shi Z. Responsivity enhancement of mid-infrared PbSe detectors using CaF_2 nano-structured antireflective coatings // Applied Physics Letters. 2014. V. 104. N 2. P. 021109. https://doi.org/10.1063/1.4861186
5. Томаев В.В., Егоров С.В., Стоянова Т.В. Исследование фоточувствительности композита из селенида и селенита свинца в ультрафиолетовой области спектра // Физика и химия стекла. 2014. Т. 40. № 2. С. 268–276.
6. Alekseeva G.T., Gurieva E.A., Konstantinov P.P., Prokofeva L.V., Fedorov M.I. Thermoelectric figure of merit of hetero- and isovalently doped PbSe // Semiconductors. 1996. V. 30. N 12. P. 1125–1127.
7. Avery D.G., Goodwin D.W., Lawson W.D., Moss T.S. Optical and photo-electrical properties of indium antimonide // Proceedings of the Physical Society. Section B. 1954. V. 67. N 10. P. 761. https://doi.org/10.1088/0370-1301/67/10/304
8. Paul W., Jones D.A., Jones R.V. Infra-Red Transmission of Galena // Proceedings of the Physical Society. Section B. 1951. V. 64. N 6. P. 528. https://doi.org/10.1088/0370-1301/64/6/109
9. Gibson A.F. The absorption spectra of single crystals of lead sulphide, selenide and telluride // Proceedings of the Physical Society. Section B. 1952. V. 65. N 7. P. 555. https://doi.org/10.1088/0370-1301/65/7/516
10. Humphrey J.N., Scanlon W.W. Photoconductivity in lead selenide. Experimental // Physical Review. 1957. V. 105. N 2. P. 469–476. https://doi.org/10.1103/physrev.105.469
11. Danilov E.A., Veretennikov M., Dronova M., Kalyakin T., Stepashkin A.A., Tcherdyntsev V.V., Samoilov V. Simple route to increase electrical conductivity and optical transmittance in graphene/silver nanoparticles hybrid suspensions // Applied Sciences. 2023. V. 13. N 3. P. 1922. https://doi.org/10.3390/app13031922
12. Непомнящий С.В., Погодина С.Б. Способ изготовления полупроводниковой структуры на основе селенида свинца. Патент RU 2493632 C1. Бюл. 2013. № 26.
13. Kolobov A.V., Tominaga J. Chalcogenide glasses in optical recording: recent progress // Journal of Optoelectronics and Advanced Materials. 2002. V. 4. N 3. P. 679–686.
14. Olkhova A.A., Patrikeeva A.A., Sergeev M.M. Electrical and optical properties of laser-induced structural modifications in PbSe films // Applied Sciences. 2022. V. 12. N 19. P. 10162. https://doi.org/10.3390/app121910162
15. Silverman S.J., Levinstein H. Electrical properties of single crystals and thin films of PbSe and PbTe // Physical Review. 1954. V. 94. N 4. P. 871–876. https://doi.org/10.1103/physrev.94.871
16. Ahmed R., Gupta M.C. Mid-infrared photoresponse of electrodeposited PbSe thin films by laser processing and sensitization // Optics and Lasers in Engineering. 2020. V. 134. P. 106299. https://doi.org/10.1016/j.optlaseng.2020.106299
17. Вейко В.П., Киеу К.К. Лазерная аморфизация стеклокерамик: основные закономерности и новые возможности изготовления микрооптических элементов // Квантовая электроника. 2007. Т. 37. № 1. С. 92–98.
18. Voznyi A., Kosyak V., Onufrijevs P., Grase L., Vecstaudža J., Opanasyuk A., Medvid' A. Laser-induced SnS_2-SnS phase transition and surface modification in SnS_2 thin films // Journal of Alloys and Compounds. 2016. V. 688. Part B. P. 130–139. https://doi.org/10.1016/j.jallcom.2016.07.103
19. Miyamoto I., Horn A., Gottmann J., Wortmann D., Yoshino F. Fusion welding of glass using femtosecond laser pulses with high-repetition rates // Journal of Laser Micro/Nanoengineering. 2007. V. 2. N 1. P. 57–63. https://doi.org/10.2961/jlmn.2007.01.0011

References

1. Tan C.L., Mohseni H. Emerging technologies for high performance infrared detectors. *Nanophotonics*, 2018, vol. 7, no. 1, pp. 169–197. https://doi.org/10.1515/nanoph-2017-0061
2. Karim A., Andersson J.Y. Infrared detectors: Advances, challenges and new technologies. *IOP Conference Series: Materials Science and Engineering*, 2013, vol. 51, no. 1, pp. 012001. https://doi.org/10.1088/1757-899x/51/1/012001
3. Kasiyan V., Dashevsky Z., Schwarz C.M., Shatkhin M., Flitsiyan E., Chernyak L., Khokhlov D. Infrared detectors based on semiconductor p-n junction of PbSe. *Journal of Applied Physics*, 2012, vol. 112, no. 8, pp. 086101. https://doi.org/10.1063/1.4759011
4. Weng B., Qiu J., Yuan Z., Larson P.R., Strout G.W., Shi Z. Responsivity enhancement of mid-infrared PbSe detectors using CaF_2 nano-structured antireflective coatings. *Applied Physics Letters*, 2014, vol. 104, no. 2, pp. 021109. https://doi.org/10.1063/1.4861186
5. Tomaev V.V., Egorov S.V., Stoyanova T.V. Investigation into the photosensitivity of a composite from lead selenide and selenite in UV region of spectrum. *Glass Physics and Chemistry*, 2014, vol. 40, no. 2, pp. 208–214. https://doi.org/10.1134/s1087659614020229
6. Alekseeva G.T., Gurieva E.A., Konstantinov P.P., Prokofeva L.V., Fedorov M.I. Thermoelectric figure of merit of hetero- and isovalently doped PbSe. *Semiconductors*, 1996, vol. 30, no. 12, pp. 1125–1127.
7. Avery D.G., Goodwin D.W., Lawson W.D., Moss T.S. Optical and photo-electrical properties of indium antimonide. *Proceedings of the Physical Society. Section B*, 1954, vol. 67, no. 10, pp. 761. https://doi.org/10.1088/0370-1301/67/10/304
8. Paul W., Jones D.A., Jones R.V. Infra-Red Transmission of Galena. *Proceedings of the Physical Society. Section B*, 1951, vol. 64, no. 6, pp. 528. https://doi.org/10.1088/0370-1301/64/6/109
9. Gibson A.F. The absorption spectra of single crystals of lead sulphide, selenide and telluride. *Proceedings of the Physical Society. Section B*, 1952, vol. 65, no. 7, pp. 555. https://doi.org/10.1088/0370-1301/65/7/516
10. Humphrey J.N., Scanlon W.W. Photoconductivity in lead selenide. Experimental. *Physical Review*, 1957, vol. 105, no. 2, pp. 469–476. https://doi.org/10.1103/physrev.105.469
11. Danilov E.A., Veretennikov M., Dronova M., Kalyakin T., Stepashkin A.A., Tcherdyntsev V.V., Samoilov V. Simple route to increase electrical conductivity and optical transmittance in graphene/silver nanoparticles hybrid suspensions. *Applied Sciences*, 2023, vol. 13, no. 3, pp. 1922. https://doi.org/10.3390/app13031922
12. Nepomnjashchij S.V., Pogodina S.B. Method of making lead selenide-based semiconductor structure. *Patent RU 2493632 C1*. 2013. (in Russian)
13. Kolobov A.V., Tominaga J. Chalcogenide glasses in optical recording: recent progress. *Journal of Optoelectronics and Advanced Materials*, 2002, vol. 4, no. 3, pp. 679–686.
14. Olkhova A.A., Patrikeeva A.A., Sergeev M.M. Electrical and optical properties of laser-induced structural modifications in PbSe films. *Applied Sciences*, 2022, vol. 12, no. 19, pp. 10162. https://doi.org/10.3390/app121910162
15. Silverman S.J., Levinstein H. Electrical properties of single crystals and thin films of PbSe and PbTe. *Physical Review*, 1954, vol. 94, no. 4, pp. 871–876. https://doi.org/10.1103/physrev.94.871
16. Ahmed R., Gupta M.C. Mid-infrared photoresponse of electrodeposited PbSe thin films by laser processing and sensitization. *Optics and Lasers in Engineering*, 2020, vol. 134, pp. 106299. https://doi.org/10.1016/j.optlaseng.2020.106299
17. Veiko V.P., Kieu K.K. Laser amorphisation of glass ceramics: basic properties and new possibilities for manufacturing microoptical elements. *Quantum Electronics*, 2007, vol. 37, no. 1, pp. 92–98. https://doi.org/10.1070/qe2007v037n01abeh008992
18. Voznyi A., Kosyak V., Onufrijevs P., Grase L., Vecstaudža J., Opanasyuk A., Medvid' A. Laser-induced SnS_2-SnS phase transition and surface modification in SnS_2 thin films. *Journal of Alloys and Compounds*, 2016, vol. 688, part B, pp. 130–139. https://doi.org/10.1016/j.jallcom.2016.07.103
19. Miyamoto I., Horn A., Gottmann J., Wortmann D., Yoshino F. Fusion welding of glass using femtosecond laser pulses with high-repetition rates. *Journal of Laser Micro/Nanoengineering*, 2007, vol. 2, no. 1, pp. 57–63. https://doi.org/10.2961/jlmn.2007.01.0011

Авторы

Ольхова Анастасия Александровна — аспирант, младший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57814489800, https://orcid.org/0000-0001-9048-3031, olkhova.a.a@mail.ru
Патрикеева Алина Александровна — студент, инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57815254200, https://orcid.org/0000-0002-5274-9692, patrikeeva17@gmail.com
Бутяева Мария Алексеевна — студент, инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 58125613000, https://orcid.org/0000-0003-3251-6602, maria.dubkova@mail.ru
Пушкарева Александра Евгеньевна — кандидат технических наук, ведущий инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 12791000500, https://orcid.org/0000-0003-0082-984X, alexandra.pushkareva@gmail.com
Авилова Екатерина Александровна — аспирант, инженер-исследователь, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0002-4520-6761, eaavilova@itmo.ru
Москвин Михаил Константинович — младший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 57194008858, https://orcid.org/0000-0001-7399-7022, mkmoskvin@itmo.ru
Сергеев Максим Михайлович — кандидат технических наук, доцент, старший научный сотрудник, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 55624732300, https://orcid.org/0000-0003-2854-9954, maxim.m.sergeev@gmail.com
Вейко Вадим Павлович — профессор, доктор технических наук, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 7005095644, https://orcid.org/0000-0001-6071-3449, vpveiko@itmo.ru

Authors

Anastasiia A. Olkhova — PhD Student, Junior Researcher, ITMO University. Saint Petersburg, 197101, Russian Federation, sc 57814489800, https://orcid.org/0000-0001-9048-3031, olkhova.a.a@mail.ru
Alina A. Patrikeeva — Student, Engineer, ITMO University. Saint Petersburg, 197101, Russian Federation, sc 57815254200, https://orcid.org/0000-0002-5274-9692, patrikeeva17@gmail.com

Maria A. Butyaeva — Student, Engineer, ITMO University. Saint Petersburg, 197101, Russian Federation, sc 58125613000, https://orcid.org/0000-0003-3251-6602, maria.dubkova@mail.ru

Alexandra E. Pushkareva — PhD, Leading Engineer, ITMO University. Saint Petersburg, 197101, Russian Federation, sc 12791000500, https://orcid.org/0000-0003-0082-984X, alexandra.pushkareva@gmail.com

Ekaterina A. Avilova — PhD Student, Research Engineer, ITMO University. Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0002-4520-6761, eaavilova@itmo.ru
Mikhail K. Moskvin — Junior Researcher, ITMO University. Saint Petersburg, 197101, Russian Federation, sc 57194008858, https://orcid.org/0000-0001-7399-7022, mkmoskvin@itmo.ru

Maksim M. Sergeev — PhD, Associate Professor, Senior Researcher, ITMO University. Saint Petersburg, 197101, Russian Federation, sc 55624732300, https://orcid.org/0000-0003-2854-9954, maxim.m.sergeev@gmail.com
Vadim P. Veiko — D.Sc., Full Professor, ITMO University. Saint Petersburg, 197101, Russian Federation, sc 7005095644, https://orcid.org/0000-0001-6071-3449, vpveiko@itmo.ru

Статья поступила в редакцию 22.11.2023
Одобрена после рецензирования 13.12.2023
Принята к печати 26.01.2024

Received 22.11.2023
Approved after reviewing 13.12.2023
Accepted 26.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

КОМПЬЮТЕРНЫЕ СИСТЕМЫ И ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

COMPUTER SCIENCE

doi: 10.17586/2226-1494-2024-24-1-41-50
УДК 004.021, 004.827

Алгоритм распознавания омографов на основе евклидовой метрики

Элиса Салаудиновна Израилова[1]✉, Арсланбек Виситович Астемиров[2], Айшат Салауддиновна Бадаева[3], Зелимхан Аюбович Султанов[4], Салаудин Мусаевич Умархаджиев[5], Мохмад-Салех Лейчевич Хехаев[6], Мадина Лечаевна Ясаева[7]

[1,2,3,4,5,6,7] Академия наук Чеченской Республики, Грозный, 364043, Российская Федерация
[1,2,3,4,5,6] Комплексный научно-исследовательский институт им. Х.И. Ибрагимова Российской академии наук, Грозный, 364051, Российская Федерация

[1] uelisa@yandex.ru✉, https://orcid.org/0000-0003-1655-4414
[2] astemirow@mail.ru, https://orcid.org/0009-0003-8416-7587
[3] ayshatbs@gmail.com, https://orcid.org/0009-0009-5351-229X
[4] zelimkhan.sultanov@yandex.ru, https://orcid.org/0009-0001-1391-5475
[5] usmu@mail.ru, https://orcid.org/0000-0002-8283-1515
[6] saleh_1991@mail.ru, https://orcid.org/0009-0009-7672-4148
[7] madina.yasaeva@internet.ru, https://orcid.org/0009-0001-3565-4974

Аннотация

Введение. Проблема разрешения неопределенностей, связанных с омонимией, для чеченского языка стала особенно актуальной после создания систем синтеза речи. Основным недостатком синтезаторов речи на чеченском языке являются ошибки чтения слов-омографов, различающихся долготой/краткостью гласных — долгота таких звуков никак не отображается при письме. Вызывает проблемы воспроизведение дифтонгов, которые обозначаются на письме так же, как близкие им по звучанию монофтонги. Для повышения качества синтезируемой речи на чеченском языке необходима программа автоматического распознавания омографов. Для решения этой проблемы рассмотрена задача устранения неоднозначности смысла слов Word Sense Disambiguation (WSD). **Метод.** Для чеченского языка выбраны алгоритмические (управляемые) методы, работающие на основе заранее размеченной базы данных. Эти методы являются наиболее распространенными при устранении неоднозначности смысла слов. Реализация таких методов возможна при наличии больших размеченных корпусов, которые недоступны для большинства языков мира, в том числе и для чеченского. Чеченский язык относится к малоресурсным языкам, для которых оптимальным подходом, с точки зрения экономии трудовых и временных ресурсов, является полууправляемый гибридный метод распознавания омографов, основанный на использовании алгоритмических и статистических методов. **Основные результаты.** Разработан алгоритм распознавания омографов по шести соседним словам в предложении. Алгоритм реализован в виде программы. Предварительная подготовка исходных данных для работы алгоритма включает разметку предложений по значениям омографов, выполняемую «вручную». Результаты работы программы оценены с использованием общепризнанных метрик точности и составили F1 — 39 %, Accuracy — 45 %. **Обсуждение.** Сравнительный анализ полученных данных с результатами других методов и моделей показал, что точность представленного алгоритма наиболее близка к результатам точности алгоритмов на основе метода Леска. По методу Леска для английского языка получены результаты точности F1 — 41,1% (простой Леск) и 51,1 % (Леск расширенный). Методы, использующие нейросетевые алгоритмы, дают более высокие показатели точности по WSD для большинства языков, однако для их реализации требуется наличие больших корпусов данных, что не всегда доступно для малоресурсных языков, в том числе и для чеченского.

Ключевые слова

графическая омонимия, омографы, WSD, синтез речи, чеченский язык, малоресурсные языки, текстовый корпус

Ссылка для цитирования: Израилова Э.С., Астемиров А.В., Бадаева А.С., Султанов З.А., Умархаджиев С.М., Хехаев М.-С.Л., Ясаева М.Л. Алгоритм распознавания омографов на основе евклидовой метрики // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 41–50. doi: 10.17586/2226-1494-2024-24-1-41-50

© Израилова Э.С., Астемиров А.В., Бадаева А.С., Султанов З.А., Умархаджиев С.М., Хехаев М.-С.Л., Ясаева М.Л., 2024

Homograph recognition algorithm based on Euclidean metric

Elisa S. Izrailova[1]✉, Arslanbek V. Astemirov[2], Ayshat S. Badaeva[3], Zelimkhan A. Sultanov[4], Salaudin M. Umarkhadzhiev[5], Mokhmad-Salekh L. Khekhaev[6], Madina L. Yasaeva[7]

[1,2,3,4,5,6,7] Academy of Sciences of the Chechen Republic, Grozny, 364043, Russian Federation
[1,2,3,4,5,6] Kh. Ibragimov Complex Institute of the Russian Academy of Sciences, Grozny, 364051, Russian Federation

[1] uelisa@yandex.ru✉, https://orcid.org/0000-0003-1655-4414
[2] astemirow@mail.ru, https://orcid.org/0009-0003-8416-7587
[3] ayshatbs@gmail.com, https://orcid.org/0009-0009-5351-229X
[4] zelimkhan.sultanov@yandex.ru, https://orcid.org/0009-0001-1391-5475
[5] usmu@mail.ru, https://orcid.org/0000-0002-8283-1515
[6] saleh_1991@mail.ru, https://orcid.org/0009-0009-7672-4148
[7] madina.yasaeva@internet.ru, https://orcid.org/0009-0001-3565-4974

Abstract

The problem of resolving the uncertainties associated with homonymy for the Chechen language has become especially relevant after the creation of speech synthesis systems. The main disadvantage of speech synthesizers in the Chechen language are errors in reading homograph words that differ in the length / brevity of vowels — the longitude of such sounds is not displayed in any way when writing. The reproduction of diphthongs, which are indicated on the letter in the same way as monophthongs close to them in sound, causes problems. To improve the quality of synthesized speech in the Chechen language, an automatic homograph recognition program is needed. To solve this problem, the article considers the task of eliminating the ambiguity of the meaning of the words WSD (Word Sense Disambiguation). Algorithmic (supervised) methods based on a pre-marked database have been selected for the Chechen language. These methods are the most common solutions for eliminating the ambiguity of the meaning of words. The implementation of such methods is possible in the presence of large marked-up corpora that are inaccessible to most languages of the world including Chechen. The Chechen language belongs to low-resource languages for which the optimal approach from the point of view of saving labor and time resources is a semi-controlled hybrid method of homograph recognition based on the use of algorithmic and statistical methods. The algorithm created by the authors for recognizing homographs by six adjacent words in a sentence is presented. The method is implemented as a program. Preliminary preparation of the initial data for the operation of the algorithm includes marking of proposals by the values of homographs performed "manually". The results of the program were evaluated using generally recognized accuracy metrics and amounted to F1 — 39 %, Accuracy — 45 %. A comparative analysis of the data obtained with the results of other methods and models showed that the accuracy of the algorithm presented in this article is closest to the results of the accuracy of algorithms based on the Lesk method. Using Lesk method for English, the results of F1 accuracy were obtained — 41.1 % (simple Lesk) and 51.1 % (extended Lesk). Methods using neural network algorithms provide higher WSD accuracy rates for most languages; however, their implementation requires large data bodies, which is not always available for low-resource languages, including Chechen.

Keywords

graphic homonymy, homographs, WSD, speech synthesis, Chechen language, low resource languages, text corpus

For citation: Izrailova E.S., Astemirov A.V., Badaeva A.S., Sultanov Z.A., Umarkhadzhiev S.M., Khekhaev M.-S.L., Yasaeva M.L. Homograph recognition algorithm based on Euclidean metric. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 41–50 (in Russian). doi: 10.17586/2226-1494-2024-24-1-41-50

Введение

Системы искусственного интеллекта, связанные с обработкой естественных языков, такие как системы синтеза и распознавания речи, автоматические переводчики, морфологические анализаторы, поисковые системы, несомненно, должны включать функцию лексического анализа для устранения многозначности и выявления единственно правильного решения в рамках заданной проблемы.

В настоящей работе рассмотрена проблема устранения неоднозначности омографов в текстах на чеченском языке.

Последние несколько лет в отделе прикладной семиотики Академии наук Чеченской Республики ведутся исследования в области речевых технологий, в результате которых получены практические результаты. В частности, создано онлайн-приложение синтеза чеченской речи на основе базы данных Speeche1 [1] и базы чеченских текстов, являющихся обучающим корпусом для нейросетевых речевых систем. На основе данного корпуса была создана система синтеза чеченской речи, состоящая из различных функциональных модулей: модуль для автоматического преобразования орфографической записи текста в транскрипцию[1], нормализатор для расшифровки числительных и аббревиатур[2], модуль обучения, состоящий из двух нейронных сетей; модуль синтезирования речи на основе вокодера.

После создания синтезатора речи были проведены тестирование, MOS-оценка (Mean Opinion Score) и

[1] Свидетельство № 2019664423, 06.11.2019. Российская Федерация. Elp-Az: свидетельство об официальной регистрации программы для ЭВМ / С.М. Умархаджиев, Э.С. Израилова, А.С. Бадаева, М.Х. Бекаев, З.А. Султанов, Х.И. Асхабов, Я.В. Эльсаев, А.А. Абдулкадыров.

[2] Свидетельство № 2020660527, 04.09.2020. Российская Федерация. TerahYoza: свидетельство об официальной регистрации программы для ЭВМ / С.М. Умархаджиев, Э.С. Израилова, А.С. Бадаева, А.В. Астемиров, З.А. Султанов.

анализ качества синтезируемого программой речевого сигнала [2]. В результате анализа был выявлен недостаток синтезатора — программа не всегда правильно читает слова-омографы, различающиеся долготой/краткостью гласных, а также дифтонгами/монофтонгами. Как показывает опыт создания систем синтеза речи при некорректном определении признаков слова-омографа и последующем синтезе такого слова затрудняется восприятие смысла всего предложения. Для чеченского языка эта проблема особенно актуальна ввиду наличия в нем большого количества омографов. Обусловлено это, как правило, наличием некоторых несоответствий графики, используемой для чеченского языка, особенностям его фонетики и орфоэпии. Главным образом, речь идет об отсутствии специальных средств отображения на письме дифтонгов и долготы гласных. Исходя из этого, было решено провести оптимизацию созданной системы синтеза речи, чтобы решить проблему распознавания омографов в текстах на чеченском языке для их правильного прочтения.

Основная трудность при решении проблемы распознавания омографов — малоресурсность чеченского языка, т. е. отсутствие необходимых текстовых обучающих корпусов и баз данных для реализации различных алгоритмов распознавания омографов.

Для чеченского языка проблема устранения многозначности слов никогда до этого не решалась. В основе настоящей работы лежит исследование существующих методов Word Sense Disambiguation (WSD), при этом учтена специфика чеченского языка, ограниченность ресурсов и особенности чеченской орфографии.

Цель работы — описание особенностей устранения многозначности омографов для чеченского языка, рассмотрение существующих методов WSD и процесса подготовки данных, представление разработанного алгоритма распознавания омографов на основе весового евклидова расстояния между предложениями.

Методы устранения лексической многозначности (WSD)

Все методы WSD делятся на три основные группы. Одну из этих групп составляют методы, основанные на знаниях (knowledge-based methods) и использующие электронные словари. Основным методом первой группы является метод Леска — классический алгоритм разрешения лексической многозначности, основанный на знаниях. Метод предложен Майклом Леском в 1986 году. Идея метода заключается в поиске значения слова в толковых словарях и списках словарных определений с учетом контекста, где это слово использовано [3].

Существует множество модификаций метода Леска, одна из них предлагает, кроме использования самих определений слов, непосредственно входящих в фразу, применить концепцию иерархии в базе WordNet [4]. Еще один метод, основанный на концепции Леска, метод семантической схожести (semantic similarity), включающий разные меры схожести. Общей для всех этих методов является идея того, что схожие слова могут определяться контекстом [5]. В работе [6] рассмотрены современные нейронные расширения метода Леска, основанные на определениях для вычисления смысловых вложений, использующихся в качестве обучающих. Хорошие результаты показала модель SyntagRank, в которой применены пары совпадающих слов с устраненной неоднозначностью, включенные в SyntagNet — ресурс лексико-семантических комбинаций для выполнения современного WSD [7]. Модель интересна тем, что может быть масштабирована для множества языков, которые представлены в многоязычном словаре BabelNet, включающем как лексикографические, так и энциклопедические термины.

Вторая группа методов включает методы машинного обучения с учителем, так называемые контролируемые методы (supervised methods), работающие на основе заранее размеченной базы данных. В эту группу входят методы деревьев решений (decision trees) [8–10], метод опорных векторов (Support Vector Machine) [11], нейронные сети [12, 13].

Методы, использующие нейронные сети с учителем, получили в последнее время наибольшее распространение и наилучшие результаты в устранении неоднозначности. Среди них можно выделить несколько наиболее интересных моделей: модели WSD, полностью управляемые данными, обучающиеся с помощью градиентного анализа [14, 15]; контролируемые модели WSD, использующие глоссы — текстовые определения, закодированные в виде векторов [16, 17]; модели, основанные на знаниях и реляционных графах в WordNet [18, 19].

Методы, применяющие нейросетевые алгоритмы, обеспечивают высокие показатели точности по WSD для большинства языков. В работе [20] получен высокий результат для русского языка с помощью нейросетевой архитектуры ELMo (Embeddings from Language Model), основа которой — вложения из языковых моделей. Используя ELMo, вычислены векторы контекста неконтролируемым образом с помощью двух уровней двунаправленного LSTM (Long Short-Term Memory), принимающих встраивания символов из сверточного слоя в качестве входных данных. Результат точности ELMo для русского языка F1 составил 77 %, а с использованием модели BERT (Bidirectional Encoder Representations from Transformers) — 67 % [21]. В [22] для арабского языка при различных модификациях модели BERT (ArBERT, AraBERTv2) на разных корпусах данных получены результаты точности F1 от 74 до 89 %. В архитектуре модели BERT использован многоуровневый двунаправленный преобразователь-кодер и токены WordPiece, т. е. единицы подслова, что помогает избежать проблемы слов, которых нет в словарном запасе. Для татарского языка модель разрешения морфологической многозначности [23], основанная на рекуррентной нейронной сети с долгой краткосрочной памятью (LSTM), показала результат точности F1 — 79 %, а гибридный метод разрешения многозначности на основе скрытых марковских моделей PurePos — 86 %.

Третью группу методов составляют методы, не требующие ни заранее размеченной базы, ни использования словарей. Они направлены на автоматическое извлечение значений слов и предполагают в качестве основы кластеризацию слов и кластеризацию контек-

ста. Главная идея этих подходов состоит в том, что слова с одинаковыми смыслами будут иметь схожие соседние слова [24, 25].

Ученые Санкт-Петербургского Центра речевых технологий для снятия омонимии в системе синтеза русской речи использовали метод лингвистического анализа, включающий поиск в предложении ключевых слов или выражений [26]. Этот подход включает анализ слов непосредственно рядом с текущим омографом, т. е. анализируется грамматическое окружение и выполняется поиск согласованных слов в предложении. Для формализации этого принципа были введены грамматические правила, увеличивающие условный «вес» словоформы в зависимости от ее окружения. Правила хранятся в формализованном виде, позволяющем быстро оценивать и корректировать работу системы.

Проблема устранения неоднозначности омографов для чеченского языка

В компьютерной лингвистике WSD рассмотрена в виде задачи определения правильного смысла слова в контексте. Проблема устранения многозначности слов на сегодняшний день является одной из актуальных проблем обработки большинства естественных языков. При этом одна из задач, которую требуется решить — неоднозначность омографов. Данную проблему возможно устранить либо посредством полноценного семантического анализа, либо с помощью автономного подхода, который выполняется независимо от лексических и морфологических норм языка.

Для решения проблемы устранения многозначности слов необходимо наличие больших текстовых корпусов, различных словарей и баз данных для реализации методов WSD, как алгоритмических моделей, так и методов машинного обучения. Метки, идентифицирующие слова-омографы и контекстно-зависимые сочетания слов, производятся специалистами вручную, что требует больших временных и кадровых ресурсов. В настоящее время для лексического анализа наиболее часто используется ресурс WordNet [27]. WordNet — большая база данных лексических отношений для английского и других языков. В стадии разработки находятся базы для русского языка.

Основные научные исследования в области WSD и созданные на их основе методы также получены для английского и других языков, достаточно обеспеченных электронными ресурсами. В настоящей работе для проведения сравнительного анализа были изучены материалы по проблеме WSD по некоторым малоресурсным языкам (турецкий, татарский, восточноафриканский, хинди и др.). В чеченском языке много заимствований из русского и арабского языков, поэтому были изучены методы устранения многозначности для этих языков.

Основная проблема, с которой столкнулись авторы данной работы — для чеченского языка не созданы базы данных и текстовые корпусы. Для решения проблемы были собраны максимально разнообразные по тематике и жанрам тексты для формирования всевозможных явлений омонимии и контекстов омонимов. В настоящее время собранный банк текстов насчитывает около 15 млн слов и используется как инструмент для языковых исследований, разработки и настройки различных автоматизированных систем, в том числе и для решения проблемы WSD, в частности для подготовки данных, необходимых при реализации методов распознавания омографов и омонимов [28]. Банк текстов постоянно пополняется новыми текстами.

Для чеченского языка вышеприведенные методы, основанные на знаниях, не являются приемлемыми, поскольку чеченский язык по международной классификации относится к малоресурсным языкам. Основные проблемы, возникающие при разработке прикладных программ для чеченского языка — нехватка электронных ресурсов для обработки языка и речи, в том числе одноязычных корпусов, двуязычных электронных словарей, орфоэпических и грамматических словарей, тезаурусов, толковых словарей, словарей омонимов и других различных лингвистических ресурсов [29].

В результате авторами настоящей работы создана новая база данных для реализации WSD: собран текстовый корпус; извлечены тексты с предложениями, содержащими омографы в различных контекстах; выполнена «вручную» разметка текстов по значениям омографов. Для чеченского языка существуют два основных словаря, которые были оцифрованы в отделе прикладной семиотики Академии наук Чеченской Республики: «Русско-чеченский словарь» А.Г. Карасаева, А.Т. Мациева и «Чеченско-русский словарь» А.Т. Мациева, а также несколько отраслевых русско-чеченских словарей. На основе сведений, содержащихся в этих словарях, разработаны алгоритмы и программы, созданы обучающие корпуса и текстовые базы для проверки орфографии и статистического анализа.

Решение проблемы разрешения графической омонимии в текстах на чеченском языке состоит из следующих функциональных задач:

— подготовка текстового корпуса и словаря омографов;
— сбор лексических признаков омографов и их компоновка в словаре;
— разработка программы для автоматизации поиска и сбора предложений с омографами;
— создание базы предложений, содержащих омографы в различных контекстах;
— разработка программы для автоматизации выборки тегов — слов, однозначно идентифицирующих омографы в контексте;
— создание словаря тегов;
— статистический частотный анализ омографов и тегов;
— создание алгоритма и программной реализации программы распознавания омографов;
— распознавание омографов в тексте при создании автоматических транскрипций;
— применение полученных результатов для автоматического синтеза чеченской речи.

Подготовка данных

В процессе изучения проблемы WSD выполнено исследование алгоритмических (управляемых) методов, работающих на основе заранее размеченной базы

данных (вторая группа методов, рассмотренная в разделе «Методы устранения лексической многозначности (WSD)»), так как они наиболее распространены при устранении неоднозначности слов. Однако для реализации таких методов необходимо наличие больших размеченных корпусов, которые недоступны для большинства языков мира. Так как чеченский язык относится к малоресурсным языкам, оптимальным методом разрешения графической омонимии должен быть тот, который использует для реализации минимальные ресурсы. Реализация такого «экономичного» метода возможна благодаря полууправляемому гибридному методу распознавания омографов.

Полууправляемые методы могут решить проблему распознавания омографов, используя большой размеченный корпус, одна часть которого применена для анализа частотности и индексации тегов, вторая часть — для проверки точности распознавания омографов, т. е. для тестирования созданной программы. В процентном соотношении эти части обычно составляют 80/20, причем большая часть текстового корпуса используется для индексации тегов.

Предлагаемый гибридный метод устранения графической омонимии основан на использовании алгоритмических методов и статистики. Данный метод наиболее подходит для разрабатываемого алгоритма, так как при чтении текста читателю приходится анализировать контекст, чтобы правильно произнести слова-омографы. Статистический метод направлен на анализ контекста омографов по следующим признакам: позиция в предложении; наличие служебных слов и аффиксов; контекст слова справа и слева.

В предлагаемом алгоритме рассмотрен только один признак омографа — контекст слова справа и слева. Алгоритм программы распознавания омонимов рассматривает слова в их контексте, что позволяет сократить список возможных слов до очень ограниченного числа высоковероятных слов-тегов, которые позволяют однозначно идентифицировать омограф в контексте. Далее статистика тегов использована при реализации полученного алгоритма.

Для реализации метода статистического анализа были определены и собраны 100 омографов, наиболее часто встречающихся в литературе и устной речи. Для каждого омографа осуществлен поиск в базе текстов на чеченском языке, в ходе которого выбрана база предложений, содержащих данный омограф. Далее файлы были обработаны созданной по алгоритму программой, проводящей анализ статистики слов-тегов, идентифицирующих в контексте омографы. Программа выполнила индексацию тегов в соответствии с частотой их использования в контексте заданного омографа, показывающая количество повторов соответствующих тегов в базе предложений. Полученная статистика позволила выявить наиболее часто встречающиеся в контексте с омографом теги, из которых был сформирован словарь тегов [30].

База предложений является основой для разработки и реализации предложенного алгоритма. Предложения с омографами собраны автоматически из сформированного банка текстов, результатом программы стала электронная таблица Excel, в которой все предложения с омографами находятся в первом столбце. Дальнейшая работа проведена вручную. База предложений с конкретным омографом была разбита на два столбца в зависимости от значения омографа (в основном по долготе гласной буквы). При письме долгота гласных букв никак не обозначается, поэтому и возникла проблема синтезирования слов-омографов в чеченской речи [30].

В табл. 1 приведен фрагмент таблицы из базы предложений для омографа «бала» со значениями «бала» (с кратким звуком «а») и «бАла» (с долгим звуком «а»).

Для исследования было выбрано 13 наиболее частотных омографов, собраны предложения из созданного авторами текстового корпуса, содержащие эти омографы в различных контекстах. Так, для омографа «дан» база содержит 3150 предложений, «де» — 973, «бала» — 482, «белира» — 800, «яха» — 196, «йолу» — 2130, «бен» — 1511, «дала» — 1301, «дакъа» — 490, «шун» — 543, «цIе» — 2000, «ча» — 288, «лар» — 1171. Общее количество предложений составило 15 035.

Алгоритм AWEN распознавания омографов на основе евклидовой метрики

Рассмотрим разработанный алгоритм AWEN (algorithm, weight, Euclid, number of tags) распознавания омографов на основе евклидовой метрики.

Исходные данные и обозначения. Пусть *om* — омограф (слово, имеющее два значения om_i, где $i = 1, 2$).

Таблица 1. Фрагмент таблицы из базы предложений для омографа «бала»
Table 1. Fragment of a table from the database of sentences for the homograph «бала»

Вариант произношения	
бала (краткий звук)	бАла (долгий звук)
Адаман ойланех цуьнан амалех кхетам бала тарлун цхьацца методикаш хуьлу малхбузен литературехь юьйцуш амма чlагlдойла яц цара дерриг а нийса дуьйцу аьлла	Иштта цхьа дагна бАла lаьткъича бен кху тlаьхьарчу шина баттахь цигаьрка ца узуш волчу lалхас шена цигаьрка а хьарчийна боккха баькkхина кlур пехашна кlоргге чууьйзира
Уггар хьалха хьомсара государашх хьомсара государынаш хастам бала беза сийлахь-доккхачу оьрсийн муьтlахьаллина	Болх луш верг а болх беш верг а цуьнан бАла кхочуш ца хилча хала хир ду оцу гlуллакха тlехь кхиаме кхача
Дена шен доьзал мичахь бу хиъчахьана тоуьйту аса-м аьлла шашена ела а къежна хьешана кхача бала дагахь тохаелира хlусамнана	Шен дагара бАла гучу ца болуьйтуш хала садеттара lаббаса

Обозначим буквой p предложение, содержащее омограф om; B_i, где $i = 1, 2$ — две базы предложений $p_{i,j}, j = 1, 2, \ldots, kp_i$, содержащих kp_i предложений. В результате получим два множества предложений:

$$B_1 = \{p_{1,1}, p_{1,2}, \ldots, p_{1,kp1}\} \text{ и } B_2 = \{p_{2,1}, p_{2,2}, \ldots, p_{2,kp2}\}.$$

Буквой N обозначим выбранное количество ближайших к омографу om слов — тегов, по которым алгоритм будет идентифицировать значение омографа. Для удобства выберем число N четным. Обозначим $sp_{i,j,m}$, где $m = 1, 2, \ldots, N$ — слова в предложении $p_{i,j}$. Из них $\frac{N}{2}$ слов $sp_{i,j,m}$, $m = 1, 2, \ldots, \frac{N}{2}$, слева от омографа om и $\frac{N}{2}$ слов $sp_{i,j,m}$, $m = \frac{N}{2} + 1, \frac{N}{2} + 2, \ldots, N$, справа от омографа om.

Такое симметричное разбиение количества тегов необязательно. Отметим, что оптимальность разбиения зависит от порядка слов в предложении в естественном языке. Этот выбор можно сделать экспериментально.

Базы слов. Сформируем две базы S_i, $i = 1, 2$ различных слов $s_{i,k}$, $k = 1, 2, \ldots, ks_i$, содержащихся хотя бы в одном предложении из баз B_i, $i = 1, 2$, где ks_i — количество слов в базе S_i. Будем считать, что каждому слову $s_{i,k}$ из базы S_i соответствуют неотрицательные целые числа $ns_{i,k,m}$, $m = 1, 2, \ldots, N$, равные количеству предложений из базы B_i, в которых слово $s_{i,k}$ совпадает со словом $sp_{i,j,m}$ на m-ом месте в предложении $p_{i,k}$:

$$ns_{i,k,m} = \sum_{\substack{k=1, \\ s_{i,k}=sp_{i,k,m}}}^{kp_i} 1, \; m = 1, 2, \ldots, N, \; i = 1, 2.$$

Таким образом, каждому слову $s_{i,k}$ из баз S_i, $i = 1, 2$, будет соответствовать N-мерный числовой вектор

$$\mathbf{vs_{i,k}} = (ns_{i,k,1}, ns_{i,k,2}, \ldots, ns_{i,k,N}).$$

Заметим, что в некоторых случаях, когда омограф в предложении расположен достаточно близко к началу или к концу предложения, или предложение состоит из малого числа слов, количество тегов может оказаться меньше числа N. В таких случаях соответствующая недостающему слову координата будет равна нулю.

***N*-мерное векторное пространство.** Если каждому предложению $p_{i,j}, j = 1, 2, \ldots kp_i$, из баз B_i, $i = 1, 2$, соответствует числовой вектор

$$\mathbf{vp_{i,j}} = (np_{i,j,1}, np_{i,j,2}, \ldots, np_{i,j,N}),$$

то координаты вектора равны сумме соответствующих координат вектора $\mathbf{vs_{i,j}}$:

$$np_{i,j,m} = \sum_{s_{i,j} \in S_i} \sum_{m=1}^{N} ns_{i,j,m}, \; i = 1, 2, j = 1, 2, \ldots, kp_i.$$

Таким образом, каждое предложение $p_{i,j}$ есть точка N-мерного векторного пространства. В этом пространстве имеем два множества B_1 и B_2 точек $p_{1,j}$ и $p_{2,j}$, $j = 1, 2, \ldots, N$. Идея рассматриваемого алгоритма заключается в том, что необходимо выяснить к какому из двух этих множеств «ближе» содержащее данный омограф тестовое предложение. Исходя из этого, в полученном векторном пространстве необходимо ввести понятие расстояния между предложениями.

Весовое евклидово пространство. Введем в этом пространстве метрику Евклида: расстояние между числовыми векторами $\mathbf{a} = (a_1, a_2, \ldots, a_N)$ и $\mathbf{b} = (b_1, b_2, \ldots, b_N)$ определим формулой

$$\rho(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_{k=1}^{N} (a_k - b_k)^2}. \quad (1)$$

Естественно предположить, что чем дальше тег в предложении от омографа, тем меньше его влияние на значение омографа. Чтобы отразить в алгоритме данное условие, введем в формулу (1) так называемый весовой вектор

$$\mathbf{w} = (w_1, w_2, \ldots, w_N),$$

координаты которого удовлетворяют условиям

$$0 < w_1 \le w_2 \le \ldots \le w_{\frac{N}{2}} = 1 = w_{\frac{N}{2}+1} \ge w_{\frac{N}{2}+2} \ge \ldots \ge w_N > 0,$$

где N — четное число. Тогда весовая метрика (1) примет вид

$$\rho(\mathbf{a}, \mathbf{b}) = \sqrt{\sum_{k=1}^{N} w_k (a_k - b_k)^2}.$$

Тестовое предложение. Пусть tp — тестовое предложение, содержащее омограф om и не входящее в базы B_i, $i = 1, 2$. Необходимо распознать значение омографа om.

Обозначим теги вокруг омографа om в предложении tp через ts_i, $i = 1, 2, \ldots, N$. Координаты $ts_{i,k}$, $k = 1, 2, \ldots, N$, слов ts_i вычислим по обоим базам B_1 и B_2, значения которых равны количеству предложений, в которых слово $ts_{i,k}$ содержится на k-ом месте от омографа:

$$ts_{i,k} = \begin{cases} sp_{1,i,k}, \text{ если } ts_i = sp_{1,i} \\ 0, \text{ если } ts_i \ne sp_{1,i} \end{cases} + \\ + \begin{cases} sp_{2,i,k}, \text{ если } ts_i = sp_{2,i} \\ 0, \text{ если } ts_i \ne sp_{2,i} \end{cases},$$

$$i = 1, 2, \ldots, N, \; k = 1, 2, \ldots, N.$$

Таким образом, каждый тег ts_i в предложении tp имеет N координат:

$$\mathbf{ts_i} = (ts_{i,1}, ts_{i,2}, \ldots, ts_{i,N}), \; i = 1, 2, \ldots, N.$$

Просуммировав соответствующие координаты этих слов, получим координаты тестового предложения:

$$tv_i = \sum_{k=1}^{N} ts_{i,k}, \; i = 1, 2, \ldots, N.$$

Следовательно, тестовому предложению tp соответствует числовой вектор

$$\mathbf{tv} = (tv_1, tv_2, \ldots, tv_N).$$

Финальный этап. Теперь вычислим арифметические средние расстояний от предложения *tp* до предложений из баз B_1 и B_2, полагая, что

$$\rho(tp, p_{i,j}) = \rho(tv, vp_{i,j}), j = 1, 2, \ldots, kp_i, i = 1, 2,$$

$$r_i = \frac{1}{kp_i} \sum_{j=1}^{kp_i} \sqrt{\sum_{k=1}^{N} w_k (tv_k - np_{i,j,k})^2}, i = 1, 2.$$

Сравнивая полученные числа, выбираем значение омографа: если $r_1 < r_2$, то $om = om_1$, если $r_1 > r_2$, то $om = om_2$ и если $r_1 = r_2$, то $om = om_1$ при $kp_1 > kp_2$, иначе $om = om_2$.

Результаты работы программы на основе алгоритма AWEN

Для исходных параметров алгоритма экспериментальным путем установим оптимальные значения: количество тегов $N = 6$, т. е. значение омографа определяется по шести словам в предложении: по трем — слева и по трем — справа от омографа; весовой вектор $\mathbf{w} = (0{,}9; 0{,}95; 1; 1; 0{,}95; 0{,}9)$.

Для тестирования результатов работы алгоритма создан модуль самотестирования программы распознавания омографов. Для модуля выделены случайным образом выбранные части баз B_1 и B_2. На оставшихся частях выполнено обучение. Экспериментально определено оптимальное процентное соотношение между обучающей и тестовой частями баз — 80/20. Цикл обучения на подготовленных базах составил 100 итераций, в итоге программа вывела средние значения для разных метрик, которые отображены в табл. 2.

Таким образом, результаты работы программы, созданной по представленному в работе алгоритму распознавания омографов на основе евклидовой метрики, составили F1 — 39 %, Accuracy — 45 % (табл. 2).

Проведенный сравнительный анализ показал, что результат точности разработанного алгоритма наиболее близок к результатам алгоритмов на основе метода Леска, которые для английского языка варьируются в пределах F1 41–51 %.

Проанализированы результаты точности различных нейросетевых моделей для нескольких языков, наиболее близких к чеченскому. Например, показатель точности модели BERT для русского языка составил F1 — 67 %. Для арабского языка различные модификации этой модели (ArBERT, AraBERTv2) на разных корпусах данных показали результаты метрики F1 от 74 до 89 %. Для татарского языка модель разрешения морфологической многозначности, основанная на рекуррентной нейронной сети (LSTM) дала результат точности F1 — 79 %.

Таблица 2. Точность распознавания омографов, %

Table 2. Accuracy of homograph recognition, %

Омограф	F1	Accuracy
яха	53	49
ча	41	42
дакъа	40	41
белира	66	70
бала	34	36
дала	69	87
де	31	31
цIе	28	32
шун	28	32
йолу	18	20
бен	29	35
лар	31	41
дан	31	41

Методы, использующие нейросетевые алгоритмы, дают более высокие показатели точности по WSD для большинства языков, однако для их реализации требуется наличие больших корпусов данных, что не всегда доступно для малоресурсных языков, в том числе и для чеченского.

Заключение

Заметим, что алгоритм AWEN нетрудно дополнить для идентификации значений не только омографов, но омонимов, имеющих более двух значений. Однако подготовленные текстовые корпуса с предложениями содержат данные только для омографов. Такое ограничение было принято с целью экономии временных ресурсов. Поскольку необходимости в распознавании омонимов в целом для улучшения качества синтеза речи не было, имеющиеся ресурсы были направлены на автоматическое распознавание значений омографов, так как правильность именно их прочтения значительно влияет на качество синтеза речи.

Результаты работы алгоритма AWEN проанализированы и сопоставлены с результатами различных методов для разных языков. Для сравнения полученных оценок приведены значения мер F1 и Accuracy.

Для малоресурсного чеченского языка результаты работы программы, созданной по алгоритму распознавания омографов на основе весовой евклидовой метрики, составили F1 — 39 %, Accuracy — 45 %.

Литература

1. Израилова Э.С. Процесс создания системы синтеза чеченской речи // Известия Российского государственного педагогического университета им. А.И. Герцена. 2020. № 198. С. 171–177. https://doi.org/10.33910/1992-6464-2020-198-171-177
2. Izrailova E.S., Badaeva A.S. Analysis of the speech signal quality of the chechen speech synthesis system // Automatic Documentation and Mathematical Linguistics. 2021. V. 55. N 2. P. 74–78. https://doi.org/10.3103/S0005105521020059
3. Lesk M. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone // Proc. of the 5th Annual International Conference on Systems Documentation. 1986. P. 24–26. https://doi.org/10.1145/318723.318728
4. Banerjee S., Pedersen T. An adapted lesk algorithm for word sense disambiguation using WordNet // Lecture Notes in Computer Science. 2002. V. 2276. P. 136–145. https://doi.org/10.1007/3-540-45715-1_11
5. Lastra-Diaz J.J., Goikoetxea J., Taieb M.A.H., Garcia-Serrano A., Aouicha M.B., Agirre E. A reproducible survey on word embeddings and ontology-based methods for word similarity: linear combinations outperform the state of the art // Engineering Applications of Artificial Intelligence. 2019. V. 85. P. 645–665. https://doi.org/10.1016/j.engappai.2019.07.010
6. Kumar S., Jat S., Saxena K., Talukdar P. Zero-shot word sense disambiguation using sense definition embeddings // Proc. of the 57th Annual Meeting of the Association for Computational Linguistics. 2019. P. 5670–5681. https://doi.org/10.18653/v1/p19-1568
7. Scozzafava F., Maru M., Brignone F., Torrisi G., Navigli R. Personalized PageRank with syntagmatic information for multilingual Word Sense Disambiguation // Proc. of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations. 2020. P. 37–46. https://doi.org/10.18653/v1/2020.acl-demos.6
8. Escudero G., Marquez L., Rigau G., Salgado J.G. On the portability and tuning of supervised word sense disambiguation systems: Research report. 2000.
9. Manning C.D., Clark K., Hewitt J., Khandelwal U., Levy O. Emergent linguistic structure in artificial neural networks trained by self-supervision // Proceedings of the National Academy of Sciences. 2020. V. 117. N 48. P. 30046–30054. https://doi.org/10.1073/pnas.1907367117
10. Lin D. Automatic retrieval and clustering of similar words // Proc. of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. V. 2. 1998. P. 768–774. https://doi.org/10.3115/980691.980696
11. Hadiwinoto C., Ng H.T., Gan W.C. Improved Word Sense Disambiguation using pre-trained contextualized word representations // Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 2019. P. 5297–5306. https://doi.org/10.18653/v1/D19-1533
12. Vial L., Lecouteux B., Schwab D. Sense vocabulary compression through the semantic knowledge of WordNet for neural Word Sense Disambiguation // Proc. of the 10th Global Wordnet Conference. 2019. P. 108–117.
13. Scarlini B., Pasini T., Navigli R. SensEmBERT: Context-enhanced sense embeddings for multilingual Word Sense Disambiguation // Proceedings of the AAAI Conference on Artificial Intelligence. 2020. V. 34. N 5. P. 8758–8765. https://doi.org/10.1609/aaai.v34i05.6402
14. Scarlini B., Pasini T., Navigli R. With more contexts comes better performance: Contextualized sense embeddings for all-round Word Sense Disambiguation // Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 2020. P. 3528–3539. https://doi.org/10.18653/v1/2020.emnlp-main.285
15. Zhang C.X., Liu R., Gao X.Y., Yu B. Graph convolutional network for word sense disambiguation // Discrete Dynamics in Nature and Society. 2021. V. 2021. P. 2822126. https://doi.org/10.1155/2021/2822126
16. Conia S., Navigli R. Framing Word Sense Disambiguation as a multi-label problem for model-agnostic knowledge integration // Proc. of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. 2021. P. 3269–3275. https://doi.org/10.18653/v1/2021.eacl-main.286

References

1. Izrailova E. Creating a system for synthesizing the Chechen speech. *Izvestia: Herzen University Journal of Humanities & Sciences*, 2020, no. 198, pp. 171–177. (in Russian). https://doi.org/10.33910/1992-6464-2020-198-171-177
2. Izrailova E.S., Badaeva A.S. Analysis of the speech signal quality of the chechen speech synthesis system. *Automatic Documentation and Mathematical Linguistics*, 2021, vol. 55, no. 2, pp. 74–78. https://doi.org/10.3103/S0005105521020059
3. Lesk M. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. *Proc. of the 5th Annual International Conference on Systems Documentation*, 1986, pp. 24–26. https://doi.org/10.1145/318723.318728
4. Banerjee S., Pedersen T. An adapted lesk algorithm for word sense disambiguation using WordNet. *Lecture Notes in Computer Science*, 2002, vol. 2276, pp. 136–145. https://doi.org/10.1007/3-540-45715-1_11
5. Lastra-Diaz J.J., Goikoetxea J., Taieb M.A.H., Garcia-Serrano A., Aouicha M.B., Agirre E. A reproducible survey on word embeddings and ontology-based methods for word similarity: linear combinations outperform the state of the art. *Engineering Applications of Artificial Intelligence*, 2019, vol. 85, pp. 645–665. https://doi.org/10.1016/j.engappai.2019.07.010
6. Kumar S., Jat S., Saxena K., Talukdar P. Zero-shot word sense disambiguation using sense definition embeddings. *Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, 2019, pp. 5670–5681. https://doi.org/10.18653/v1/p19-1568
7. Scozzafava F., Maru M., Brignone F., Torrisi G., Navigli R. Personalized PageRank with syntagmatic information for multilingual Word Sense Disambiguation. *Proc. of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, 2020, pp. 37–46. https://doi.org/10.18653/v1/2020.acl-demos.6
8. Escudero G., Marquez L., Rigau G., Salgado J.G. *On the portability and tuning of supervised word sense disambiguation systems*. Research report, 2000.
9. Manning C.D., Clark K., Hewitt J., Khandelwal U., Levy O. Emergent linguistic structure in artificial neural networks trained by self-supervision. *Proceedings of the National Academy of Sciences*, 2020, vol. 117, no. 48, pp. 30046–30054. https://doi.org/10.1073/pnas.1907367117
10. Lin D. Automatic retrieval and clustering of similar words. *Proc. of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics. V. 2*, 1998, pp. 768–774. https://doi.org/10.3115/980691.980696
11. Hadiwinoto C., Ng H.T., Gan W.C. Improved Word Sense Disambiguation using pre-trained contextualized word representations. *Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, 2019, pp. 5297–5306. https://doi.org/10.18653/v1/D19-1533
12. Vial L., Lecouteux B., Schwab D. Sense vocabulary compression through the semantic knowledge of WordNet for neural Word Sense Disambiguation. *Proc. of the 10th Global Wordnet Conference*, 2019, pp. 108–117.
13. Scarlini B., Pasini T., Navigli R. SensEmBERT: Context-enhanced sense embeddings for multilingual Word Sense Disambiguation. *Proceedings of the AAAI Conference on Artificial Intelligence*, 2020, vol. 34, no. 5, pp. 8758–8765. https://doi.org/10.1609/aaai.v34i05.6402
14. Scarlini B., Pasini T., Navigli R. With more contexts comes better performance: Contextualized sense embeddings for all-round Word Sense Disambiguation. *Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020, pp. 3528–3539. https://doi.org/10.18653/v1/2020.emnlp-main.285
15. Zhang C.X., Liu R., Gao X.Y., Yu B. Graph convolutional network for word sense disambiguation. *Discrete Dynamics in Nature and Society*, 2021, vol. 2021, pp. 2822126. https://doi.org/10.1155/2021/2822126
16. Conia S., Navigli R. Framing Word Sense Disambiguation as a multi-label problem for model-agnostic knowledge integration. *Proc. of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, 2021, pp. 3269–3275. https://doi.org/10.18653/v1/2021.eacl-main.286

17. Amrami A., Goldberg Y. Towards better substitution-based word sense induction // arXiv. 2019. arXiv:1905.12598. https://doi.org/10.48550/arXiv.1905.12598
18. Arefyev N., Sheludko B., Panchenko A. Combining lexical substitutes in neural word sense induction // Proc. of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019). 2019. P. 62–70. https://doi.org/10.26615/978-954-452-056-4_008
19. Vasilescu F., Langlais P., Lapalme G. Evaluating variants of the lesk approach for disambiguating words // Proc. of the Fourth International Conference on Language Resources and Evaluation (LREC'04). 2004.
20. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding // Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 2019. P. 4171–4186.
21. El-Razzaz M., Fakhr M.W., Maghraby F.A. Arabic Gloss WSD Using BERT // Applied Sciences. 2021. V. 11. N 6. P. 2567. https://doi.org/10.3390/app11062567
22. Kilgarriff A., Rosenzweig J. Framework and results for English SENSEVAL // Computers the Humanities. 2000. V. 34. N 1. P. 15–48. https://doi.org/10.1023/A:1002693207386
23. Гатауллин Р.Р., Гильмуллин Р.А., Хакимов Б.Э. Разрешение морфологической многозначности в корпусе татарского языка на основе статистико-вероятностной модели Purepos и нейросетевой модели LSTM // VI Международная конференция по компьютерной обработке тюркских языков «TurkLang 2018» (труды конференции). Ташкент: Издательско-полиграфический дом «Navoiy Universiteti», 2018. С. 133–138.
24. Haveliwala T.H. Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search // IEEE Transactions on Knowledge and Data Engineering. 2003. V. 15. N 4. P. 784–796. https://doi.org/10.1109/tkde.2003.1208999
25. Peters M.E., Neumann M., Iyyer M., Gardner M., Clark C., Lee K., Zettlemoyer L. Deep contextualized word representations // Proc. of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers). 2018. P. 2227–2237. https://doi.org/10.18653/v1/N18-1202
26. Хомицевич О.Г., Рыбин С.В., Аничкин И.М. Использование лингвистического анализа для нормализации текста и снятия омонимии в системе синтеза русской речи // Известия высших учебных заведений. Приборостроение. 2013. Т. 56. № 2. С. 42–46.
27. WordNet: An Electronic Lexical Database // ed. by Ch. Fellbaum. Cambridge, MA: MIT Press, 1998. 423 p.
28. Ясаева М.Л. Создание баз данных чеченских текстов для обработки алгоритмов распознавания омографов компьютерными системами // Всероссийская научно-практическая конференция «Актуальные проблемы исследования родного языка и литературы». Грозный, 2022. С. 65–69.
29. Карпов А.А., Верходанова В.О. Речевые технологии для малоресурсных языков мира // Вопросы языкознания. 2015. № 2. С. 117–135.
30. Израилова Э.С., Астемиров А.В. Статистический контекстный анализ для снятия графической омонимии в текстах на чеченском языке // Материалы Международной научной конференции «Актуальные проблемы развития современной науки» посвященная 30-летию Академии наук Чеченской Республики. Махачкала: Академия наук Чеченской Республики, 2023. С. 478–485.

17. Amrami A., Goldberg Y. Towards better substitution-based word sense induction. *arXiv*, 2019, arXiv:1905.12598. https://doi.org/10.48550/arXiv.1905.12598
18. Arefyev N., Sheludko B., Panchenko A. Combining lexical substitutes in neural word sense induction. *Proc. of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, 2019, pp. 62–70. https://doi.org/10.26615/978-954-452-056-4_008
19. Vasilescu F., Langlais P., Lapalme G. Evaluating variants of the lesk approach for disambiguating words. *Proc. of the Fourth International Conference on Language Resources and Evaluation (LREC'04)*, 2004.
20. Devlin J., Chang M.-W., Lee K., Toutanova K. BERT: Pre-training of deep bidirectional transformers for language understanding. *Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, 2019, pp. 4171–4186.
21. El-Razzaz M., Fakhr M.W., Maghraby F.A. Arabic Gloss WSD Using BERT. *Applied Sciences*, 2021, vol. 11, no. 6, pp. 2567. https://doi.org/10.3390/app11062567
22. Kilgarriff A., Rosenzweig J. Framework and results for English SENSEVAL. *Computers the Humanities*, 2000, vol. 34, no. 1, pp. 15–48. https://doi.org/10.1023/A:1002693207386
23. Gataullin R.R., Gilmullin R.A., Khakimov B.E. Morphological disambiguation in the national corpus of tatar language using Purepos and LSTM models. *VI International Conference on Computer Processing of Turkic Languages "TurkLang 2018" (Proceedings of the Conference)*, Tashkent, Navoiy Universiteti Publ., 2018, pp. 133–138. (in Russian)
24. Haveliwala T.H. Topic-sensitive pagerank: A context-sensitive ranking algorithm for web search. *IEEE Transactions on Knowledge and Data Engineering*, 2003, vol. 15, no. 4, pp. 784–796. https://doi.org/10.1109/tkde.2003.1208999
25. Peters M.E., Neumann M., Iyyer M., Gardner M., Clark C., Lee K., Zettlemoyer L. Deep contextualized word representations. *Proc. of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, 2018, pp. 2227–2237. https://doi.org/10.18653/v1/N18-1202
26. Khomitsevich O.G., Rybin S.V., Anichkin I.M. Application of linguistic analysis for text normalization and homonymy resolution in Russian text-to-speech system. *Journal of Instrument Engineering*, 2013, vol. 56, no. 2, pp. 42–46. (in Russian)
27. *WordNet: An Electronic Lexical Database*. Ed. by Ch. Fellbaum. Cambridge, MA, MIT Press, 1998, 423 p.
28. Yasaeva M.L. Creation of databases of Chechen texts for processing homograph recognition algorithms by computer systems. *All-Russian Scientific and Practical Conference "Current Problems of Native Language and Literature Research"*, Grozny, 2022, pp. 65–69. (in Russian)
29. Karpov A.A., Verkhodanova V.O. Speech technologies for under-resourced languages of the world. *Voprosy Jazykoznanija*, 2015, no. 2, pp. 117–135. (in Russian)
30. Israilov E.S., Astemirov A.V. Statistical context analysis program for removing graphic homonymy in texts in the chechen language. *Proceedings of the International Scientific Conference "Current Issues in the Development of Modern Science" theme-based to the 30th anniversary of the Academy of Sciences of the Chechen Republic*, Makhachkala, Chechen Academy of Sciences, 2023, pp. 478–485. (in Russian)

Авторы

Израилова Элиса Салаудиновна — старший научный сотрудник, Академия наук Чеченской Республики, Грозный, 364043, Российская Федерация; младший научный сотрудник, Комплексный научно-исследовательский институт им. Х.И. Ибрагимова Российской академии наук, Грозный, 364051, Российская Федерация, https://orcid.org/0000-0003-1655-4414, uelisa@yandex.ru

Астемиров Арсланбек Виситович — научный сотрудник, Академия наук Чеченской Республики, Грозный, 364043, Российская Федерация; младший научный сотрудник, Комплексный научно-исследовательский институт им. Х.И. Ибрагимова Российской академии наук, Грозный, 364051, Российская Федерация, https://orcid.org/0009-0003-8416-7587, astemirow@mail.ru

Authors

Elisa S. Izrailova — Senior Researcher, Academy of Sciences of the Chechen Republic, Grozny, 364043, Russian Federation; Junior Researcher, Kh. Ibragimov Complex Institute of the Russian Academy of Sciences, Grozny, 364051, Russian Federation, https://orcid.org/0000-0003-1655-4414, uelisa@yandex.ru

Arslanbek V. Astemirov — Scientific Researcher, Academy of Sciences of the Chechen Republic, Grozny, 364043, Russian Federation; Junior Researcher, Kh. Ibragimov Complex Institute of the Russian Academy of Sciences, Grozny, 364051, Russian Federation, https://orcid.org/0009-0003-8416-7587, astemirow@mail.ru

Бадаева Айшат Салауддиновна — научный сотрудник, Академия наук Чеченской Республики, Грозный, 364043, Российская Федерация; младший научный сотрудник, Комплексный научно-исследовательский институт им. Х.И. Ибрагимова Российской академии наук, Грозный, 364051, Российская Федерация, https://orcid.org/0009-0009-5351-229X, ayshatbs@gmail.com

Султанов Зелимхан Аюбович — научный сотрудник, Академия наук Чеченской Республики, Грозный, 364043, Российская Федерация; младший научный сотрудник, Комплексный научно-исследовательский институт им. Х.И. Ибрагимова Российской академии наук, Грозный, 364051, Российская Федерация, https://orcid.org/0009-0001-1391-5475, zelimkhan.sultanov@yandex.ru

Умархаджиев Салаудин Мусаевич — доктор физико-математических наук, доцент, заведующий отделом, Академия наук Чеченской Республики, Грозный, 364043, Российская Федерация; заведующий лабораторией, Комплексный научно-исследовательский институт им. Х.И. Ибрагимова Российской академии наук, Грозный, 364051, Российская Федерация, SC 37089765500, https://orcid.org/0000-0002-8283-1515, usmu@mail.ru

Хехаев Мохмад-Салех Лейчевич — научный сотрудник, Академия наук Чеченской Республики, Грозный, 364043, Российская Федерация; младший научный сотрудник, Комплексный научно-исследовательский институт им. Х.И. Ибрагимова Российской академии наук, Грозный, 364051, Российская Федерация, https://orcid.org/0009-0009-7672-4148, saleh_1991@mail.ru

Ясаева Мадина Лечаевна — научный сотрудник, Академия наук Чеченской Республики, Грозный, 364043, Российская Федерация, https://orcid.org/0009-0001-3565-4974, madina.yasaeva@internet.ru

Ayshat S. Badaeva — Scientific Researcher, Academy of Sciences of the Chechen Republic, Grozny, 364043, Russian Federation; Junior Researcher, Kh. Ibragimov Complex Institute of the Russian Academy of Sciences, Grozny, 364051, Russian Federation, https://orcid.org/0009-0009-5351-229X, ayshatbs@gmail.com

Zelimhan A. Sultanov — Scientific Researcher, Academy of Sciences of the Chechen Republic, Grozny, 364043, Russian Federation; Junior Researcher, Kh. Ibragimov Complex Institute of the Russian Academy of Sciences, Grozny, 364051, Russian Federation, https://orcid.org/0009-0001-1391-5475, zelimkhan.sultanov@yandex.ru

Salaudin M. Umarkhadzhiev — D.Sc (Physics & Mathematics), Associate Professor, Head of Department, Academy of Sciences of the Chechen Republic, Grozny, 364043, Russian Federation; Head of Laboratory, Kh. Ibragimov Complex Institute of the Russian Academy of Sciences, Grozny, 364051, Russian Federation, SC 37089765500, https://orcid.org/0000-0002-8283-1515, usmu@mail.ru

Mokhmad-Salekh L. Khekhaev — Scientific Researcher, Academy of Sciences of the Chechen Republic, Grozny, 364043, Russian Federation; Junior Researcher, Kh. Ibragimov Complex Institute of the Russian Academy of Sciences, Grozny, 364051, Russian Federation, https://orcid.org/0009-0009-7672-4148, saleh_1991@mail.ru

Madina L. Yasaeva — Scientific Researcher, Academy of Sciences of the Chechen Republic, Grozny, 364043, Russian Federation, https://orcid.org/0009-0001-3565-4974, madina.yasaeva@internet.ru

Статья поступила в редакцию 19.04.2023
Одобрена после рецензирования 29.11.2023
Принята к печати 13.01.2024

Received 19.04.2023
Approved after reviewing 29.11.2023
Accepted 13.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-51-61

An improved performance of RetinaNet model for hand-gun detection in custom dataset and real time surveillance video

Pyone Pyone Khin[1]✉, Nay Min Htaik[2]

[1,2] Mandalay Technological University, Mandalay, 05072, Myanmar

[1] pyonekhin.ppk@gmail.com✉, https://orcid.org/0009-0002-0512-6414
[2] nayminhtaik@gmail.com, https://orcid.org/0009-0009-8295-6914

Abstract
The prevalence of armed robberies has become a significant concern in today's world, necessitating the development of effective detection systems. While various detection devices exist in the market, they do not possess the capability to automatically detect and alarm the presence of guns during robbery activities. In order to address this issue, a deep learning-based approach using gun detection using RetinaNet model is proposed. The objective is to accurately detect guns and subsequently alert either the police station or the bank owner. RetinaNet, the core of the system, comprises three main components: the Residual Neural Network (ResNet), the Feature Pyramid Network (FPN), and the Fully Convolutional Networks (FCN). These components work together to enable real-time detection of guns without the need for human intervention. Proposed implementation uses a custom robbery detection dataset that consists of gun, no-gun and robbery activity classes. By evaluating the performance of the proposed model on our custom dataset, it is evident that the ResNet50 backbone architecture yields outperforms for the accuracy in robbery detection that reached in 0.92 of Mean Average Precision (mAP). The model effectiveness lies in its ability to accurately identify the presence of guns during robbery activities.

Keywords
robbery activities, RetinaNet, ResNet50, FPN, FCN

For citation: Khin P.P., Htaik N.M. An improved performance of RetinaNet model for hand-gun detection in custom dataset and real time surveillance video. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 51–61. doi: 10.17586/2226-1494-2024-24-1-51-61

УДК 004.032/26

Улучшенная производительность модели RetinaNet для обнаружения огнестрельного оружия в пользовательском наборе данных и видеонаблюдения в реальном времени

Пьоне Пьоне Кхин[1]✉, Най Мин Хтайк[2]

[1,2] Мандалайский технологический университет, Мандалай, 05072, Мья́нма

[1] pyonekhin.ppk@gmail.com✉, https://orcid.org/0009-0002-0512-6414
[2] nayminhtaik@gmail.com, https://orcid.org/0009-0009-8295-6914

Аннотация
Распространенность вооруженных ограблений стала серьезной проблемой в современном мире, что требует разработки эффективных систем обнаружения. Существующие разнообразные устройства обнаружения не обладают способностью автоматически выявлять и предупреждать о наличии оружия во время осуществления вооруженных ограблений. Для решения этой проблемы предлагается подход, основанный на глубоком обучении, с использованием модели RetinaNet. В результате его применения возможно точное обнаружение оружия и дальнейшее предупреждение об ограблении полицейского участка или владельца банка. Ядро модели RetinaNet состоит из трех основных компонентов: остаточной сети (Residual Neural Network, ResNet), функциональной пирамидальной сети (Feature Pyramid Net, FPN) и полностью сверточной сети (Fully Convolutional Networks, FCN). Эти компоненты работают вместе, обеспечивая обнаружение оружия в режиме реального времени без

© Khin P.P., Htaik N.M., 2024

вмешательства человека. Предлагаемая реализация использует специальный набор данных для обнаружения грабежей, который состоит из классов активности с применением огнестрельного оружия, без оружия и грабежей. Оценка производительности предлагаемой модели на разработанном специальном наборе данных показал, что магистральная архитектура ResNet50 превосходит точность обнаружения ограблений, достигая меры оценки качества ранжирования (Mean Average Precision, mAP) 0,92. Эффективность модели заключается в ее способности точно определять наличие оружия во время ограбления.

Ключевые слова
вооруженное ограбление, RetinaNet, ResNet50, FPN, FCN

Ссылка для цитирования: Кхин П.П., Хтайк Н.М. Улучшенная производительность модели RetinaNet для обнаружения огнестрельного оружия в пользовательском наборе данных и видеонаблюдения в реальном времени // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 51–61 (на англ. яз.). doi: 10.17586/2226-1494-2024-24-1-51-61

Introduction

Robberies have indeed been a persistent and common criminal activity, causing loss and anxiety for the public. A robbery prevention system that is simple to use and mostly free of false alerts is required to prevent the rising rate of robberies throughout the world [1]. To address these requirements, a system is being developed that can detect robberies in stores or surveillance areas and promptly alert the police or the owner, enabling them to take appropriate action. The proposed system utilizes deep learning to detect robbery activities by focusing on gun detection. By employing deep learning techniques, the system can accurately identify the presence of a firearm and generate an alert message to notify the police station or the owner. The system will have a simple user interface and operate in real-time ensuring the protection of individuals and their valuable possessions.

In today's dynamic and evolving environment, ensuring security against robberies is a major concern. People's precious possessions require enhanced security measures. While many banks, commercial shops, and surveillance areas already have security cameras in place but need an anti-robbery system that can look after the safety of their property even when they are not there [2]. The proposed system specifically targets the security of many organizations in Myanmar. Installing a comprehensive security system typically incurs additional expenses. Although various security devices are available for crime detection and prevention, many organizations often have limited resources to invest in intelligent devices. Hence, the primary objective of the proposed system is to detect robbery activities in banks using deep learning techniques, particularly gun detection, and promptly alert the police station or bank owner through an alert message. The contributions of this paper are as follows:

— The deep learning-based robbery activities detection system that can provide the promising Mean Average Precision (mAP) results is proposed.
— Real time robbery activities detection system is proposed which can detect robbery activities quickly and accurately.
— Standard dataset (self-annotating from scratch) including small arms, robbery activity (kneeling and hand up), and no gun (cell phone, thermal gun and metal detector) that consist of seven challenges and can actually use in real world is created available for robbery activities.

Literature Reviews

Human supervision remains an essential component in surveillance systems, ensuring effective monitoring and response. However, recent advancements in computer vision have emerged as a pivotal trend in video surveillance offering significant efficiency gains. In paper [3], the authors developed theft detection and tracking system using Closed Circuit Television (CCTV) images that employs image processing techniques to detect theft incidents and track the movement of thieves, without the need for additional sensors. The primary focus of this paper lies in object detection, specifically identifying and monitoring individuals involved in theft activities. By leveraging real-time analysis of human movement in CCTV footage, this system enables security personnel to be promptly notified about suspicious individuals engaged in burglary.

The authors introduced a fully automated computer-based system designed for identifying handguns and rifles which are fundamental armaments of concern [4]. Recent advancements at deep learning and transfer learning had described significant improvement in the detection and recognition of object. Our approach involved implementing the You Only Look Once (YOLO) v3 object detection model, trained on a customized dataset specifically tailored to our task. The validation of training results indicated that YOLOv3 outperforms both YOLOv2 and traditional Convolutional Neural Network (CNN) models. The approach applied this paper did not require intensive GPU or high computational resources, as transfer learning techniques for training the model was applied. This paper protected human life and reduced instances of manslaughter or mass killings. Furthermore, this system has the potential to be deployed in high-end surveillance and security robots, enabling the detection of weapons or unsafe assets and mitigating the risks of assault or harm to human life.

The paper [5] focused on the application of three CNN approaches for automatic handguns detection at video surveillance images. This paper explored the potential reduction of false positive detections with the incorporation of pose information regarding with how the handguns are held in the training dataset. The system evaluation showed that RetinaNet fine-tuned by the unfrozen Residual Neural Network, ResNet50, backbone provided the highest Average Precision (AP) by 0.9636 and recall by 0.9723. YOLOv3 exhibited consistent improvement of around 2 % as the explicit consideration of pose information during

training, distinguishing it from the other architectures evaluated in this paper.

The real-time object detection system for automatic weapon detection was proposed in video surveillance systems [6]. This framework introduced an early weapon detection technique using state-of-the-art, real-time object detection systems like YOLO and Single Shot Multi-Box Detector (SSD). Furthermore, the importance of minimizing false alarms was focused for ensuring the applicability of the model in real-life conditions. The developed model was well-suited to indoor surveillance cameras deployed in various circumstances, containing banks, supermarkets, malls, gas stations, and other similar environments. The system implementation showed that it can serve as a preemptive system for deterring potential robberies with the integration of outdoor surveillance cameras.

In paper [7], the authors developed a Hybrid Weapon detection system. In phase I, image processing techniques were employed, resulting in an average accuracy of 0.9464 after several dataset partitions. In phase II, the line integration method was applied, yielding a normalized electric field of $2.25 \cdot 10^{-9}$ V/m, indicating the electromagnetic waves expected to have passed through the metal object (weapon). This low value, relative to that of air (448 °C), demonstrates the minimal electromagnetic wave transmission through the metal, thereby facilitating the identification of the object. Furthermore, the fuzzy logic system exhibited an 83 % decision rate suggesting that it can reliably issue commands in weapon detection scenarios.

The authors described the real-time visual detection of handguns in videos [8]. This paper utilized the YOLOv3 model and performed the comparison of the false positive and false negative rates with the Faster Region-based CNN (RCNN) model. For the improvement in the evaluation, a dataset was compromised by handguns from various angles; it was created and combined with the ImageNet dataset. This combined dataset was then trained utilizing the YOLOv3 model. Four different videos were utilized for the validation in the performance of YOLOv3 in comparison with Faster RCNN. The results demonstrated the improved performance in handguns detection across various scenes, encompassing different rotations, scales, and shapes. The system implementation described that YOLOv3 can perform as a viable element to Faster RCNN supporting faster speed, with accuracy and suitability for real-time applications.

For the reduction of false positives and false negatives, a binary classification approach was introduced, with the pistol class designated as the reference class [9]. The relevant confusion objects were introduced for refining the detection operation. As there is lack of standard dataset for real-time environments, the creation of custom dataset was done with the collection of weapon photos through in-house camera, the manual collection of images by the internet, the deduction of data through YouTube CCTV videos, applying GitHub repositories, data from the University of Granada, and the Internet Movies Firearms Database (IMFDB) imfdb.org. Two methods were utilized: sliding window/classification and region proposal/object detection. Various approaches, including VGG16, Inception-V3, Inception-ResNetV2, SSDMobileNetV1, Faster RCNN Inception-ResNetV2 (FRIRv2), YOLOv3, and YOLOv4, were employed. The evaluation parameters as precision and recall are utilized in the system evaluation. Among all the algorithms tested, YOLOv4 outperforms other approaches with F1-score by 0.91 and a mAP by 0.9173.

In paper [10], the comparative analysis of two state-of-the-art models, YOLOv3 and YOLOv4, for weapons detection was performed. A weapons dataset was created for training including images through Google Images and supplemented with various assets. The images are manually annotated in different formats considering that YOLO requires annotation files in text format, while other models require XML format. Both models were trained on a large dataset of weapons, and their results are tested for comparative analysis. The paper demonstrated that YOLOv4 outperforms YOLOv3 in terms of processing time and sensitivity, while the precision metric is used to compare the two models.

The paper [11] described the automatic gun detection system using the Faster RCNN model. The CNN architecture was employed as a feature extractor in Faster RCNN by the experiments with Inception-ResNetV2, ResNet50, VGG16, and MobileNetV2. The system evaluation of those proposed architectures was performed in the comparison with YOLOv2. The results showed that the promising performance was achieved by Faster RCNN with Inception-ResNetV2 as the feature extractor. However, YOLOv2 provided the shortest training and testing time followed by VGG16, MobileNetV2, ResNet50, and Inception-ResNetV2 which achieved the longest training and testing time.

Theoretical Background

This section gives some theoretical background of the paper such as RetinaNet, different backbones (ResNet50, VGG19 and VGG16).

RetinaNet

RetinaNet is indeed a one-stage object detection model that addresses the challenges of imbalanced data and objects of different sizes. It achieves this through the use of a Feature Pyramid Network (FPN) and a specialized loss function called Focal Loss. By incorporating these components, RetinaNet can effectively detect objects in an image. The Focal Loss function plays a crucial role in RetinaNet architecture [12]. It tackles the issue of imbalanced data by assigning higher weights to difficult examples which are objects that the model struggles to detect. This weighting mechanism allows RetinaNet to focus more on challenging instances, improving its ability to handle imbalanced datasets. RetinaNet comprises three main subnetworks: a backbone network, an FPN, and a Fully Convolutional Network (FCN) for classification and regression. The backbone network, such as ResNet50, VGG19, or VGG16, serves as the feature extractor. It processes the input image and extracts relevant features.

Residual Neural Network

The ResNet architecture introduced a revolutionary bottom-up pathway that consists of residual blocks to improve training deep neural networks. The ResNet

consists of many convolution modules, each has many convolution layers. The output of each convolution module is used in the top-down pathway. ResNet is used as backbone architecture to extract multi-scale features in RetinaNet [13]. The output feature maps with various resolutions from backbone are sent to FPN.

Feature Pyramid Network

The FPN plays a central role in the architecture of RetinaNet. It is responsible for addressing the challenge of detecting objects at different scales by combining multi-scale features through top-down and lateral connections. The process begins with an input image that is passed through the backbone network. The FPN takes these feature maps produced by the backbone network as input. It consists of two main components: the top-down pathway and the lateral connections. In the top-down pathway, the feature maps from the topmost level of the backbone network are progressively upsampled to a higher resolution while reducing their spatial dimensions. This upsampling process can be accomplished using techniques like nearest-neighbor upsampling or transposed convolutions. Simultaneously, the lateral connections enable the FPN to combine high-level semantic information from the upsampled maps with low-level fine-grained details from the backbone feature maps. By repeating the process of upsampling and merging through the top-down and lateral connections, the FPN creates a feature pyramid with different scales. The feature pyramid captures a range of spatial resolutions, allowing RetinaNet to effectively detect objects of varying sizes. These final feature output combinations from the FPN are then used for subsequent tasks, such as object classification and bounding box regression.

Fully Convolutional Network

In RetinaNet architecture, the FCN component is responsible for performing both the classification and box regression tasks. The classification subnetwork within the FCN is designed to handle all the classification tasks. For instance, it determines whether a gun is present or not in the given image. The classification subnetwork takes the feature maps generated by the preceding stages of the network, such as the FPN, and performs classification on these features. It assigns a class label to each anchor or region of interest indicating the presence or absence of a gun. On the other hand, the box regression task is performed by the box subnetwork within the FCN. This subnetwork is responsible for refining the localization of the gun in the image and recording the bounding box coordinates. It takes the same set of feature maps as the classification subnetwork and predicts the offset or adjustment required for the anchor boxes or regions of interest to accurately localize the gun. By regressing the box coordinates, the box subnetwork refines the initial estimates and produces more precise bounding box predictions for the detected objects. Together, the classification and box regression subnetworks in the FCN collaborate to identify the presence of guns in the image and accurately localize them by predicting the bounding box coordinates.

ResNet50

ResNet50 is a type of CNN which has a 50-layered architecture. ResNet is a principal neural network that is utilized as the backbone in computer vision fields. The pretrained network can be loaded more than a million images through the database of ImageNet. This network possesses the input image size of 224 × 224. The ResNet50 architecture contains 5 stages. In the first stage, it consists of convolution and pooling. In the rest stages, each step includes a convolution and identity block [14].

VGG19

VGG19 is a CNN architecture that was used to win the 2014 ILSVR (ImageNet) competition; it is a variation of the VGG model with 19 layers. The VGG19 sets the input image size to 224 × 224. The rectified linear unit (ReLU) activation function is applied.

VGG16

VGG16 is a CNN architecture that was applied for achieving the winner at ILSVR (ImageNet) competition at 2014. In VGG16 there are thirteen convolutional layers, five Max Pooling layers, and three Dense layers which sum up to 21 layers but it has only sixteen weight layers, i.e., learnable parameters layer [15].

Proposed System Architecture

We propose a new method for addressing the problems with the current research because it will help to reduce the unnecessary utilization of hardware and reduce the cost of the project. Our proposed system will reduce the classification errors with utilizing a variety of methods, such as dataset creation, model training, classification, and detection.

Dataset Creation

The process of data collection is an essential initial step in constructing a gun detection model. To address the challenges associated with the object detection, such as viewpoint variation, deformation, lighting conditions, cluttered or textured backgrounds, and more, appropriate data needs to be collected. In this system, our own standard dataset that consists of seven challenges and can actually use in real world to detect the robbery activities is created. This dataset consists of robbery photos from our own camera, manually collected images from internet, data extracted from YouTube CCTV videos. The robbery detection dataset is divided into the following three classes: Gun, No-Gun, Robbery activity.

Robbery Detection Dataset Categories for Gun Class

Gun class consists of pistol and rifle categories as shown in Fig. 1.

Robbery Detection Dataset Categories for No-Gun Class

To reduce the number of false positives and negatives, No-Gun class that can most be confused with Gun class is added. No-Gun class consists of thermal gun, cell phone and metal detector as depicted in Fig. 2.

Robbery Detection Dataset Categories for Robbery Activity Class

In order to increase overall accuracy and precision, robbery activity class is added in our annotated dataset too. Robbery activity class consists of kneeling and hand-up categories as shown in Fig. 3.



Fig. 1. Sample dataset for Gun class (including pistols and rifles) with various sources, resolutions and conditions



Fig. 2. Sample dataset for No-Gun class (including thermal gun (*a*), cell phone (*b*) and metal detector (*c*))

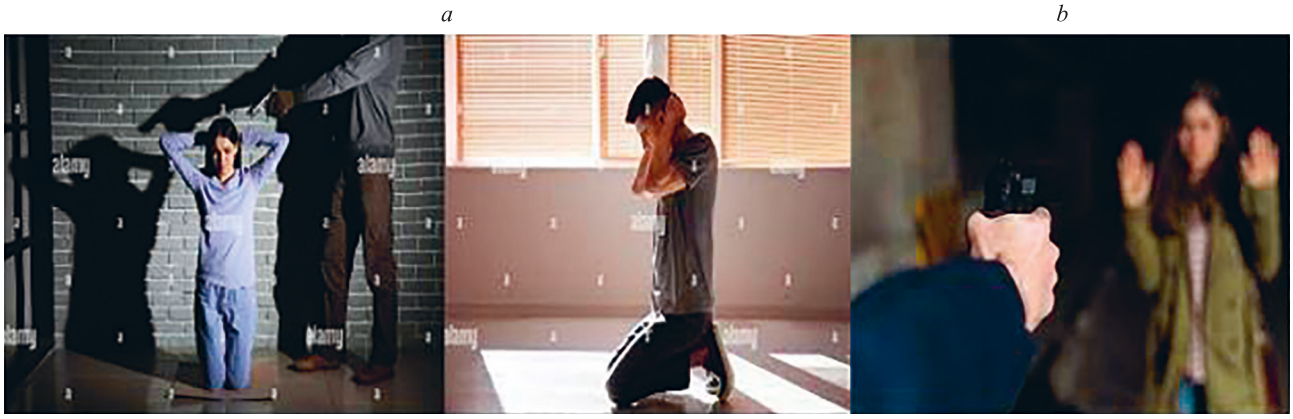


Fig. 3. Sample dataset for Robbery activity class (kneeling (*a*) and hand-up (*b*))

The step by step procedures of dataset creation are the followings.

Step 1: Collect the images from our own camera, internet, and YouTube CCTV videos.

Step 2: Annotate each image using the LabelMe annotation tool in Anaconda3 2021 in order to create ground truth labels with bounding boxes and corresponding class labels.

Step 3: Perform image preprocessing and augmentation.

Step 4: Construct our own custom dataset.

The flowchart of dataset preparation is also depicted in Fig. 4. Firstly, raw data images are collected. Later, the LabelMe annotation tool was utilized to annotate each image as shown in Fig. 5, *a*, and annotated images with JSON files are achieved as shown in Fig. 5, *b*.

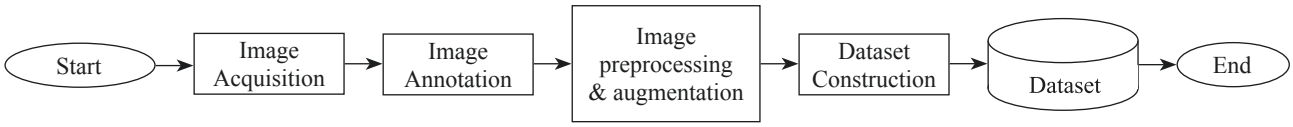


Fig. 4. Flowchart of dataset creation



Fig. 5. Image annotation and labeling (*a*), annotated images with JSON files (*b*)

After annotating images, image preprocessing and augmentation are performed by using Roboflow as illustrated in Fig. 6. In the image preprocessing, auto orientation and resizing images to 416 × 416 pixels were performed to improve the model performance. In data augmentation, random rotation between –5° and +5°, random shear of between –10° and +10° horizontally and vertically, and adding salt and pepper noise to 3 % of pixels were also employed to enhance the performance of deep neural networks. To facilitate data manipulation, the JSON files containing the annotations for each image were converted into a CSV file by using Roboflow. Finally, these images were constructed as the standard dataset. During the system evaluation, 80 % of the images were used for training the model while the remaining 20 % were used for validation and testing.

Training

Then, the RetinaNet model is trained with the different backbones (ResNet50, VGG16, VGG19). It starts with defining a problem, finding the required dataset, applying pre-processing methods, and then finally training and evaluating the dataset. The training process for different three models is shown in Fig. 7.

Classification and Detection

Features are extracted from input images using different backbones (ResNet50, VGG16 and VGG19). First, input image is achieved from real-time camera. The image passes through the detection model. The ImageNet collection is used to train the ResNet50 model. The first layer input image is a 416 × 416 RGB image. For RetinaNet with a backbone (ResNet50, VGG16 and VGG19), 3 batch-sizes require 139 MB of RAM. The flow chart of detection and classification processes with different models is also shown in Fig. 7. After the model training is finished, its AP, recall, and mAP are validated using a test image dataset at this stage. In the detection and classification stage, the value of confidence and Intersection-Over-Union (IoU) threshold is 0.5. The system then shows the output of the model. An alarm message is received by the organization if the system detects a gun. The alarm message will be checked by the police station or the owner of the organization, and they will act against it.



Fig. 6. Image pre-processing and augmentation

System Evaluation

This model was trained on a 64-bit desktop computer with NVIDIA GeForce RTX 2080 graphic card (Compute Capability = 7.5). TensorFlow, Keras, and OpenCV are used to train a gun detection system. Using the gun detection method, a model is trained and tested on a desktop computer achieving accuracy of 0.92 during training and testing. The complete process to evaluate the proposed system is shown in the following Fig. 8. Firstly, the input video frames from CCTV camera are collected. Then, images are pre-processed to improve the quality of operation. This pre-processing step generally includes resizing frame images in order to achieve more accurate deep learning model. After the pre-processing process, the training model is built using Transfer Learning. Then, the prediction of the input resized images is performed. After that, the output for label and boundary box with confidence score is displayed if the confidence value is greater than the threshold value. In this system, the confidence value is considered with the IoU score and the classification score. Finally, the alert message is sent to the user. The step by step procedures to analyze the proposed system are the followings.

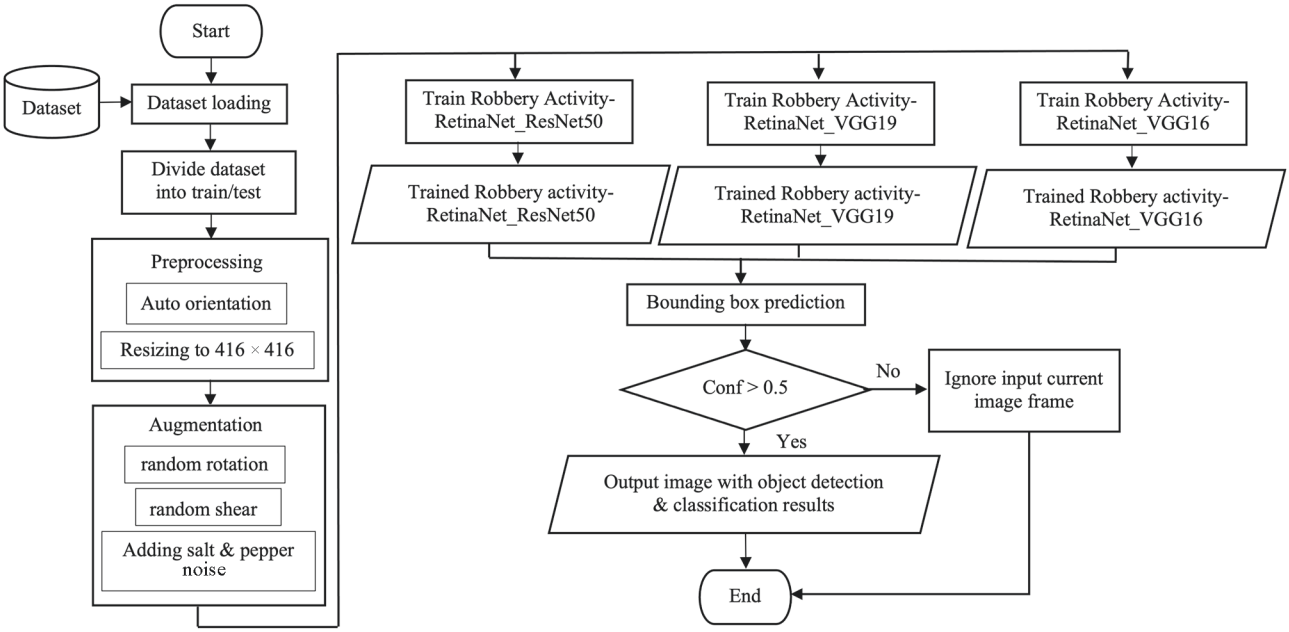


Fig. 7. Training and detection processes with three models

Step 1: Get input CCTV frame or image.

Step 2: Perform preprocessing step to resize the input image.

Step 3: Predict input resized image using the pre-trained model.

Step 4: If confidence value is greater than threshold value, then print output label and bounding box with confidence score and send an alert.

The key metrics of performance measures (AP, Recall, mAP, and Training Time) are evaluated for this proposed system analysis. This system can accurately detect the guns with low resolution and light. RetinaNet is known

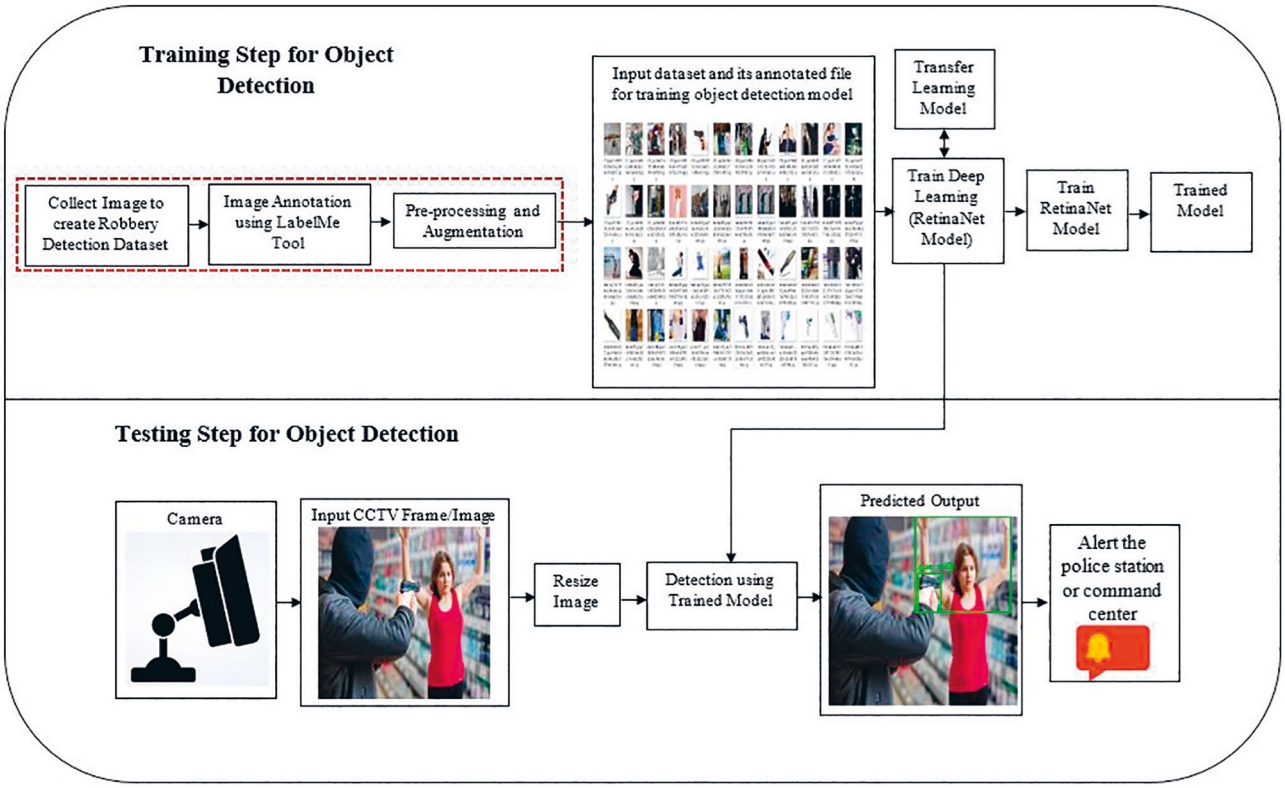


Fig. 8. Proposed system design

Table 1. Splitting Our Custom Dataset

Dataset	Pistol	Rifle	Kneeling	Hands up	No gun_phone	No gun_thermal gun	No gun_metal detector	Total images
Training	1602	1602	1861	1866	1378	1831	1860	12,000
Validation	193	193	242	234	188	228	222	1500
Testing	193	193	242	234	188	228	222	1500

Table 2. AP Results of Three Models for Each Class

Model	AP (gun)	AP (robbery activity)	AP (no gun thermal gun)	AP (no gun_metal detector	AP (no gun_phone)
RetinaNet_ResNet50	0.79	0.98	0.98	0.94	0.92
RetinaNet_VGG19	0.53	0.71	0.87	0.67	0.35
RetinaNet_VGG16	0.42	0.62	0.77	0.50	0.32

for its strong object detection performance. In a hand-gun detection system, RetinaNet capacity to detect objects across various angles and orientations is particularly beneficial. It can detect handguns in both well and poor environments. This adaptability is crucial for ensuring that handguns are identified regardless of the lighting conditions in surveillance or security scenarios. It tends to have a low false positive rate minimizing the likelihood of false alarms in a handgun detection system. This is critical for avoiding unnecessary security interventions. It directly predicts object bounding boxes and class labels without the need for a separate region proposal network. This simplicity can lead to faster inference times and simplified system architecture.

Analysis and Discussion

The system develops the gun detection system using RetinaNet_ResNet50, RetinaNet_VGG19, and RetinaNet_VGG16 models. The self-annotated dataset includes 15,000 images in total: 12,000 images for training, 1500 images for validation and 1500 images for testing as shown in Table 1. To demonstrate system evaluation using three different models, 40 epochs are used because the value of mAP started to stabilize at approximately 0.92. The comparison of training time for three models is described in Fig. 9. The AP results for each class on our annotated dataset with RetinaNet_ResNet50, RetinaNet_VGG19, and RetinaNet_VGG16 models are depicted in Table 2. The recall and mAP on our own custom dataset using three different models are also shown in Table 3. According to the evaluation results, the selection of ResNet50 can be considered the best for target detection although the RetinaNet_ResNet50 takes a little more training time than the other two models.

Table 3. Recall and mAP Results of Three Models

Model	Recall	mAP
RetinaNet_ResNet50	0.74	0.92
RetinaNet_VGG19	0.62	0.63
RetinaNet_VGG16	0.58	0.53

Detection Results: Gun Class in CCTV Records and Robbery Videos

The detection results of Gun class with CCTV records and robbery videos for RetinaNet_ResNet50 model are also illustrate in Fig. 10.

Detection Results: Robbery Activity, Thermal Gun, Cell Phone and Metal Detector Classes

The detection results of robbery activity, thermal gun, cell phone and metal detector classes for RetinaNet_ResNet50 model are also illustrated in Fig. 11. According to the detection results of Fig. 11, it can be seen that our model accurately detects multiple classes in an image.

Misdetections

The misdetection results of RetinaNet_ResNet50 model are also depicted in Fig. 12. Regarding to the detection results, it is seen that other things not being guns are mistakenly detected as guns by our model. So, we will maintain our focus on the continued reduction of false positives recognizing the persistent need for improvement. Additionally, we may explore the possibility of expanding the number of classes or objects in our work.

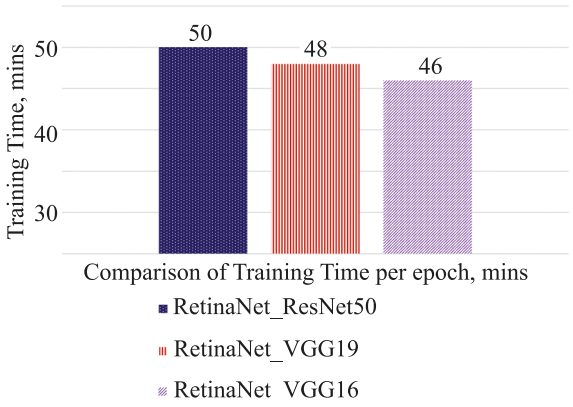


Fig. 9. Training time results of three models

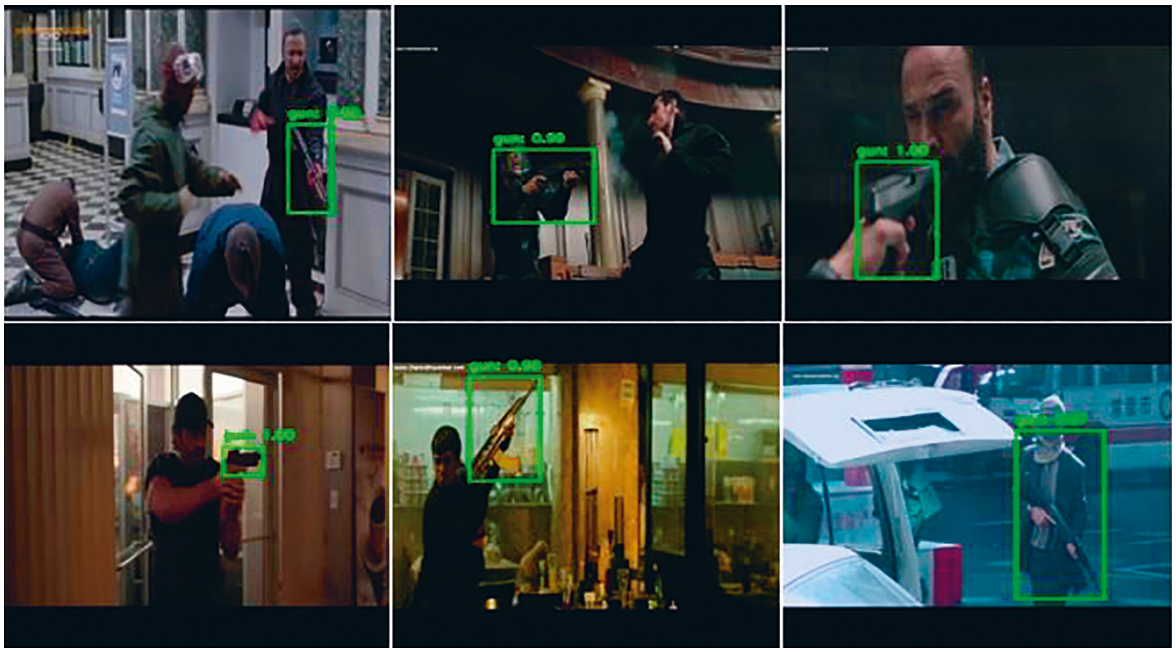


Fig. 10. Detection results of Gun class with robbery videos



Fig. 11. Detection results of robbery activity, thermal gun, cell phone and metal detector classes



Fig. 12. Misdetections: False positives

Conclusion

The main objective of the research conducted was to develop an accurate gun detection system aimed at addressing security concerns and reducing or preventing robbery incidents. The focus was on creating an anti-robbery device that utilizes gun detection capabilities. To achieve this, RetinaNet_ResNet50, RetinaNet_VGG19, and RetinaNet_VGG16 were simulated for the purpose of gun detection. These models were trained and evaluated using an annotated dataset specifically designed for this research. The performance of the models is compared using AP, recall and mAP on the self-annotated dataset. According to the AP results, the RetinaNet_ResNet50 model has shown outstanding AP for each class in our dataset than the other models. In this evaluation, RetinaNet_ResNet50 model achieves the highest detection accuracy and outperforms accurately in detecting gun, robbery activity, thermal gun, metal detector and phone. As mentioned by the evaluation, the RetinaNet_ResNet50 model achieves the best recall and mAP values among three models. It gave 0.92 mAP and 0.74 (recall) on all types of images. Overall, it may be said that RetinaNet_ResNet50 exhibited the highest detection accuracy among the three models. So, RetinaNet_ResNet50 model was deemed the most suitable choice for target detection, considering its superior performance in terms of detection accuracy.

Based on the results of this study, we will further improve our developed dataset for detecting robbery cases precisely. Regarding the classification approaches, other deep learning models, such as EfficientDet and YOLO, would be used for detecting guns. Also, it is recommended the presented results in the paper to be compared with other deep learning models of ensemble classifiers to achieve more accurate detection model. We suggest that researchers should use ensemble method by combining object detection models to detect robbery cases precisely and enhance the mAP score of the robbery detection surveillance system.

References

1. Dever J., da Vitoria Lobo N., Shah M. Automatic visual recognition of armed robbery. *Proc. of the 16th International Conference on Pattern Recognition. V. 1*, 2002, pp. 451–455. https://doi.org/10.1109/ICPR.2002.1044755
2. Ahmed S., Bhatti M.T., Khan M.G., Lövström B., Shahid M. Development and optimization of deep learning models for weapon detection in surveillance videos. *Applied Sciences*, 2022, vol. 12, no. 12, pp. 5772. https://doi.org/10.3390/app12125772
3. Kakadiya R., Lemos R., Mangalan S., Pillai M. AI based automatic robbery/theft detection using smart surveillance in banks. *Proc. of the 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 2019, pp. 201–204. https://doi.org/10.1109/iceca.2019.8822186
4. Narejo S., Pandey B., Vargas D.E., Rodriguez C., Anjum M.R. Weapon detection using YOLO V3 for smart surveillance system. *Mathematical Problems in Engineering*, 2021, vol. 2021, pp. 9975700. https://doi.org/10.1155/2021/9975700
5. Salido J., Lomas V., Ruiz-Santaquiteria J., Deniz O. Automatic handgun detection with deep learning in video surveillance images. *Applied Sciences*, 2021, vol. 11, no. 13, pp. 6085. https://doi.org/10.3390/app11136085
6. Zahrawi M., Shaalan K. Improving video surveillance systems in banks using deep learning techniques. *Scientific Reports*, 2023, vol. 13, pp. 7911. https://doi.org/10.1038/s41598-023-35190-9
7. Ineneji C., Kusaf M. Hybrid weapon detection algorithm, using material test and fuzzy logic system. *Computers & Electrical Engineering*, 2019, vol. 78, pp. 437–448. https://doi.org/10.1016/j.compeleceng.2019.08.005
8. Warsi A., Abdullah M., Husen M.N., Yahya M., Jawaid N. Gun detection system using YOLOv3. *Proc. of the IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA)*, 2019, pp. 1–4. https://doi.org/10.1109/ICSIMA47653.2019.9057329
9. Bhatti M.T., Khan M.G., Aslam M., Fiaz M.J. Weapon detection in real-time CCTV videos using deep learning. *IEEE Access*, 2021, vol. 9, pp. 34366–4382. https://doi.org/10.1109/ACCESS.2021.3059170
10. Hashmi T.S.S., Haq N.U., Fraz M.M., Shahzad M. Application of deep learning for weapons detection in surveillance videos. *Proc. of the 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2)*, 2021. https://doi.org/10.1109/ICoDT252288.2021.9441523
11. Alaqil R.M., Alsuhaibani J.A., Alhumaidi B.A., Alnasser R.A., Alotaibi R.D., Benhidour H. Automatic gun detection from images using Faster R-CNN. *Proc. of the IEEE International Conference of Smart Systems and Emerging Technologies (SMARTTECH)*, 2020,

Литература

1. Dever J., da Vitoria Lobo N., Shah M. Automatic visual recognition of armed robbery // Proc. of the 16th International Conference on Pattern Recognition. V. 1. 2002. P. 451–455. https://doi.org/10.1109/ICPR.2002.1044755
2. Ahmed S., Bhatti M.T., Khan M.G., Lövström B., Shahid M. Development and optimization of deep learning models for weapon detection in surveillance videos // Applied Sciences. 2022. V. 12. N 12. P. 5772. https://doi.org/10.3390/app12125772
3. Kakadiya R., Lemos R., Mangalan S., Pillai M. AI based automatic robbery/theft detection using smart surveillance in banks // Proc. of the 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA). 2019. P. 201–204. https://doi.org/10.1109/iceca.2019.8822186
4. Narejo S., Pandey B., Vargas D.E., Rodriguez C., Anjum M.R. Weapon detection using YOLO V3 for smart surveillance system // Mathematical Problems in Engineering. 2021. V. 2021. P. 9975700. https://doi.org/10.1155/2021/9975700
5. Salido J., Lomas V., Ruiz-Santaquiteria J., Deniz O. Automatic handgun detection with deep learning in video surveillance images // Applied Sciences. 2021. V. 11. N 13. P. 6085. https://doi.org/10.3390/app11136085
6. Zahrawi M., Shaalan K. Improving video surveillance systems in banks using deep learning techniques // Scientific Reports. 2023. V. 13. P. 7911. https://doi.org/10.1038/s41598-023-35190-9
7. Ineneji C., Kusaf M. Hybrid weapon detection algorithm, using material test and fuzzy logic system // Computers & Electrical Engineering. 2019. V. 78. P. 437–448. https://doi.org/10.1016/j.compeleceng.2019.08.005
8. Warsi A., Abdullah M., Husen M.N., Yahya M., Jawaid N. Gun detection system using YOLOv3 // Proc. of the IEEE International Conference on Smart Instrumentation, Measurement and Application (ICSIMA). 2019. P. 1–4. https://doi.org/10.1109/ICSIMA47653.2019.9057329
9. Bhatti M.T., Khan M.G., Aslam M., Fiaz M.J. Weapon detection in real-time CCTV videos using deep learning // IEEE Access. 2021. V. 9. P. 34366–4382. https://doi.org/10.1109/ACCESS.2021.3059170
10. Hashmi T.S.S., Haq N.U., Fraz M.M., Shahzad M. Application of deep learning for weapons detection in surveillance videos // Proc. of the 2021 International Conference on Digital Futures and Transformative Technologies (ICoDT2). 2021. https://doi.org/10.1109/ICoDT252288.2021.9441523
11. Alaqil R.M., Alsuhaibani J.A., Alhumaidi B.A., Alnasser R.A., Alotaibi R.D., Benhidour H. Automatic gun detection from images using Faster R-CNN // Proc. of the IEEE International Conference of Smart Systems and Emerging Technologies (SMARTTECH). 2020. P. 149–154. https://doi.org/10.1109/SMART-TECH49988.2020.00045

pp. 149–154. https://doi.org/10.1109/SMART-TECH49988.2020.00045
12. Yang M., Xiao X., Liu Z., Sun L., Guo W., Cui L., Sun D., Zhang P., Yang G. Deep RetinaNet for dynamic left ventricle detection in multiview echocardiography classification. *Scientific Programming*, 2020, vol. 2020, pp. 7025403. https://doi.org/10.1155/2020/7025403
13. Bhabad D., Kadam S., Malode T., Shinde G., Bage D. Object detection for night vision using deep learning algorithms. *International Journal of Computer Trends and Technology*, 2023, vol. 71, no. 2, pp. 87–92. https://doi.org/10.14445/22312803/ijctt-v71i2p113
14. Rani E.E., Baulkani S. Construction of deep learning model using ResNet 50 for schizophrenia prediction from rsFMRI images. *Research Square*, 2022. https://doi.org/10.21203/rs.3.rs-2106170/v1
15. Guan Q., Wang Y., Ping B., Li D., Du J., Qin Y., Lu H., Wan X., Xiang J. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study. *Journal of Cancer*, 2019, vol. 10, no. 20, pp. 4876–4882. https://doi.org/10.7150/jca.28769

12. Yang M., Xiao X., Liu Z., Sun L., Guo W., Cui L., Sun D., Zhang P., Yang G. Deep RetinaNet for dynamic left ventricle detection in multiview echocardiography classification // Scientific Programming. 2020. V. 2020. P. 7025403. https://doi.org/10.1155/2020/7025403
13. Bhabad D., Kadam S., Malode T., Shinde G., Bage D. Object detection for night vision using deep learning algorithms // International Journal of Computer Trends and Technology. 2023. V. 71. N 2. P. 87–92. https://doi.org/10.14445/22312803/ijctt-v71i2p113
14. Rani E.E., Baulkani S. Construction of deep learning model using ResNet 50 for schizophrenia prediction from rsFMRI images // Research Square. 2022. https://doi.org/10.21203/rs.3.rs-2106170/v1
15. Guan Q., Wang Y., Ping B., Li D., Du J., Qin Y., Lu H., Wan X., Xiang J. Deep convolutional neural network VGG-16 model for differential diagnosing of papillary thyroid carcinomas in cytological images: A pilot study // Journal of Cancer. 2019. V. 10. N 20. P. 4876–4882. https://doi.org/10.7150/jca.28769

Authors

Pyone Pyone Khin — Master, PhD Student, Lecturer, Mandalay Technological University, Mandalay, 05072, Myanmar, https://orcid.org/0009-0002-0512-6414, pyonekhin.ppk@gmail.com

Nay Min Htaik — PhD, Professor, Mandalay Technological University, Mandalay, 05072, Myanmar, https://orcid.org/0009-0009-8295-6914, nayminhtaik@gmail.com

Авторы

Кхин Пьоне Пьоне — магистр, аспирант, преподаватель, Мандалайский технологический университет, Мандалай, 05072, Мья́нма, https://orcid.org/0009-0002-0512-6414, pyonekhin.ppk@gmail.com

Хтайк Най Мин — PhD, профессор, Мандалайский технологический университет, Мандалай, 05072, Мья́нма, https://orcid.org/0009-0009-8295-6914, nayminhtaik@gmail.com

Received 08.08.2023
Approved after reviewing 17.11.2023
Accepted 13.01.2024

Статья поступила в редакцию 08.08.2023
Одобрена после рецензирования 17.11.2023
Принята к печати 13.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-62-69
УДК 004.383:621.391

Решение задачи предварительного разбиения разнородных данных на классы в условиях ограниченного объема

Андрей Владимирович Шарамет✉

ОАО «КБ Радар» — управляющая компания холдинга «Системы радиолокации», Минск, 220026, Республика Беларусь
Белорусский государственный университет информатики и радиоэлектроники, Минск, 220062, Республика Беларусь

a.sharamet@kbradar.by✉, https://orcid.org/0000-0003-0950-8700

Аннотация

Введение. В условиях формирования существенно отличающихся по своей природе разнородных данных даже небольшого объема возникает необходимость их анализа для принятия решений. Это характерно для многих высокотехнологичных промышленных областей деятельности человека. Задача может быть решена путем приведения разнородных данных к единому виду с последующим разбиением на кластеры. Вместо поиска решения по каждому элементу данных предлагается использовать разбиение всей совокупности нормированных данных на кластеры, и тем самым упростить процесс вычленения кластера и принятия по нему решения. **Метод.** Сущность предлагаемого решения состоит в автоматической группировке объектов со схожими данными в кластеры. Это позволяет снизить объем анализируемой информации за счет объединения множества данных и выполнить математические операции уже для кластера. При разбиении предлагается использовать теорию нечеткой логики. Возможность такого подхода связана с тем, что различные объекты всегда имеют несколько признаков, по которым они могут быть объединены. Эти признаки чаще всего не являются явными и плохо формализуются. **Основные результаты.** Предложена иерархическая модификация метода распределения по нечетким кластерам, основанного на операции (max-min) нечеткого отношения сходства. Рассмотрены основные понятия и определения предлагаемого метода автоматического разбиения совокупности входных данных, поэтапная схема соответствующей кластер-процедуры. Работоспособность предложенного метода продемонстрирована на примере решения задачи формирования потока летательных аппаратов. В качестве исходных данных использована информация, которая имеется у лиц, принимающих решение, и анализируется вручную. Численный эксперимент показал, что разработанный алгоритм позволяет автоматически провести анализ информации и корректно сформировать поток. **Обсуждение.** Применение предложенной модификации позволяет провести предварительное разбиение данных на кластеры и снизить в дальнейшем объем анализируемых данных. При этом отсутствует необходимость рассматривать объекты в каждом случае отдельно.

Ключевые слова

снижение объема вычислений, автоматическое разбиение на классы, ограниченный объем данных, иерархический метод, устойчивость алгоритма, порог сходства, транспортный поток

Благодарности

Данная статья подготовлена в память о Д.А. Вятченине, всю жизнь посвятившему себя теории нечеткого кластерного анализа.

Ссылка для цитирования: Шарамет А.В. Решение задачи предварительного разбиения разнородных данных на классы в условиях ограниченного объема // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 62–69. doi: 10.17586/2226-1494-2024-24-1-62-69

© Шарамет А.В., 2024

Solving the problem of preliminary partitioning of heterogeneous data into classes in conditions of limited volume

Andrei V. Sharamet✉

JSC "KB Radar" — Managing Company of "Radar Systems" Holding, Minsk, 220026, Republic of Belarus
Belarusian State University of Informatics and Radioelectronics, Minsk, 220062, Republic of Belarus
a.sharamet@kbradar.by✉, https://orcid.org/0000-0003-0950-8700

Abstract

In the context of the formation of heterogeneous data that differ significantly in nature, even of a small volume, it becomes necessary to analyze them for decision-making. This is typical for many high-tech industrial fields of human activity. The problem can be solved by bringing heterogeneous data to a single view and then dividing it into clusters. Instead of searching for a solution for each data element, it is proposed to use the division of the entire set of normalized data into clusters, and thereby simplify the process of isolating the cluster and making a decision on it. The essence of the proposed solution is the automatic grouping of objects with similar data into clusters. This allows you to reduce the amount of analyzed information by combining a lot of data and perform mathematical operations already for the cluster. When splitting, it is proposed to use the theory of fuzzy logic. The possibility of such an approach is due to the fact that different objects always have several characteristics by which they can be combined. These signs are often not obvious and are poorly formalized. A hierarchical modification of the AFC fuzzy clustering method based on the operation (max-min) of the fuzzy similarity ratio is proposed. The basic concepts and definitions of the proposed method of automatic partitioning of a set of input data, a step-by-step scheme of the corresponding cluster procedure are considered. The efficiency of the proposed method is demonstrated by the example of solving the problem of forming a traffic flow. A numerical experiment has shown that the developed algorithm allows you to automatically analyze heterogeneous data and stably divide them into classes. The application of the proposed modification allows for the preliminary partitioning of data into clusters and allows reducing the volume of analyzed data in the future. There is no need to consider the objects in each case separately.

Keywords

reduction of the amount of calculations, automatic division into classes, limited amount of data, hierarchical method, algorithm stability, similarity threshold, traffic flow

Acknowledgements

This article is prepared in memory of Viatchenin D.A. who devoted his whole life to the theory of fuzzy cluster analysis.

For citation: Sharamet A.V. Solving the problem of preliminary partitioning of heterogeneous data into classes in conditions of limited volume. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 62–69 (in Russian). doi: 10.17586/2226-1494-2024-24-1-62-69

Введение

При разработке систем управления и поддержки принятия решений особую роль играют методы разбиения объектов на группы [1], в частности, методы автоматической кластеризации [2], именуемые также методами численной таксономии, методами распознавания образов с самообучением [3] или методами кластерного анализа [4–7]. Группа методов, образующих оптимизационный подход к решению нечеткой модификации задачи автоматической кластеризации, является наиболее многочисленной [4, 8, 9], однако иерархические кластер-процедуры в ряде случаев позволяют провести более детальный анализ структуры классифицируемого множества [10, 11]. В работе [12] предложен эвристический метод автоматической кластеризации, в основе которого лежит операция «max-min» нечеткого отношения T, описывающая исходные данные в виде матрицы близости $\mathbf{r}_{n\times n} = [\mu_T(x_i, x_j)]$, $i, j = 1, \ldots, n$ объектов анализируемой совокупности данных — $X = \{x_1, \ldots, x_n\}$.

В настоящей работе предложена иерархическая взвешенная модификация эвристического метода, использующая в качестве основы понятие нечеткого отношения уровня α, введенное в работе [13], и аппарат распределения по нечетким кластерам (Allotment among Fuzzy Clusters, AFC) — AFC-метод нечеткой кластеризации [6, 14, 15].

Методика иерархической кластеризации данных

Иерархические кластер-процедуры позволяют наглядно представить структуру разбиваемой на кластеры совокупности X. Вместе с тем общим недостатком для всех иерархических кластер-процедур является то, что, как отмечается в работе [4], «при программной реализации методов иерархической кластеризации требования к объему оперативной памяти вычислительной платформы и времени выполнения алгоритма быстро растут с ростом числа элементов в множестве X». При этом для проведения предварительного анализа данных задача кластеризации заключается в разбиении исследуемой совокупности объектов X на некоторое «естественное» число кластеров. Методы кластеризации, выделяющие кластеры, именуются методами прямой кластеризации [7].

Рассмотрим работу предлагаемого метода. При этом исходные данные являются физическими величинами и не могут иметь отрицательные значения.

Схему соответствующей нечеткой иерархической кластер-процедуры представим в виде следующей последовательности шагов.

Шаг 1. Выполнение нормировки входных данных $\rho_{n\times n} = [\mu_I(x_i, x_j)]$, $i, j = \overline{1, n}$ нечеткого отношения несходства позволяет сформировать матрицу $\mathbf{X}_{m\times n} = [x_i^t]$, $i = \overline{1, n}$, $t = \overline{1, m}$:

$$\mathbf{X}_{m\times n} = \begin{bmatrix} 0{,}0213 & 0{,}7500 & 0{,}1563 & 0{,}0562 & 0{,}3604 & 0{,}2801 & \dots & 0{,}4942 & 0{,}1942 \\ 0{,}0204 & 0{,}7500 & 0{,}1103 & 0{,}0704 & 0{,}3878 & 0{,}2840 & \dots & 0{,}4713 & 0{,}2000 \\ 0{,}0319 & 0{,}7500 & 0{,}0828 & 0{,}0651 & 0{,}0394 & 0{,}0639 & \dots & 0{,}0481 & 0{,}0284 \\ 0{,}0272 & 0{,}7500 & 0{,}0724 & 0{,}0595 & 0{,}0223 & 0{,}0433 & \dots & 0{,}0474 & 0{,}0319 \\ 0{,}0329 & 0{,}7500 & 0{,}1617 & 0{,}0479 & 0{,}0632 & 0{,}0495 & \dots & 0{,}0446 & 0{,}0633 \\ 0{,}0512 & 0{,}5000 & 0{,}1148 & 0{,}0478 & 0{,}0553 & 0{,}0431 & \dots & 0{,}0371 & 0{,}0175 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0{,}0233 & 0{,}5000 & 0{,}0267 & 0{,}0419 & 0{,}0268 & 0{,}0497 & \dots & 0{,}0351 & 0{,}0238 \\ 0{,}0237 & 0{,}2500 & 0{,}0260 & 0{,}0400 & 0{,}0260 & 0{,}0498 & \dots & 0{,}0335 & 0{,}0223 \end{bmatrix}.$$

Далее сформируем матрицу расстояния:

$$\mathbf{d}(x_i, x_j) = \begin{bmatrix} 0 & 0{,}3501 & 0{,}5898 & 0{,}5854 & 0{,}5720 & 0{,}5768 & \dots & 0{,}6046 & 0{,}6176 \\ 0{,}3501 & 0 & 0{,}5535 & 0{,}5633 & 0{,}5602 & 0{,}5788 & \dots & 0{,}6043 & 0{,}6148 \\ 0{,}5898 & 0{,}5535 & 0 & 0{,}0677 & 0{,}2535 & 0{,}2099 & \dots & 0{,}2568 & 0{,}2788 \\ 0{,}5854 & 0{,}5633 & 0{,}0677 & 0 & 0{,}2516 & 0{,}2137 & \dots & 0{,}2699 & 0{,}2914 \\ 0{,}5720 & 0{,}5602 & 0{,}2535 & 0{,}2516 & 0 & 0{,}2047 & \dots & 0{,}3773 & 0{,}3912 \\ 0{,}5768 & 0{,}5788 & 0{,}2099 & 0{,}2137 & 0{,}2047 & 0 & \dots & 0{,}3016 & 0{,}3063 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0{,}6046 & 0{,}6043 & 0{,}2568 & 0{,}2699 & 0{,}3773 & 0{,}3016 & \dots & 0 & 0{,}0652 \\ 0{,}6176 & 0{,}6148 & 0{,}2788 & 0{,}2914 & 0{,}3912 & 0{,}3063 & \dots & 0{,}0652 & 0 \end{bmatrix}. \quad (1)$$

Применение операции дополнения нечеткого отношения к (1), позволяет сформировать матрицу попарных коэффициентов сходства объектов $\boldsymbol{\mu}_I(x_i, x_j) = 1 - \mu_I(x_i, x_j)$, $i, j = \overline{1, n}$ разбиваемой совокупности X, которая задает нечеткую толерантность T:

$$\boldsymbol{\mu}_I(x_i, x_j) = \begin{bmatrix} 1 & 0{,}6498 & 0{,}4101 & 0{,}4145 & 0{,}4279 & 0{,}4231 & \dots & 0{,}3953 & 0{,}3823 \\ 0{,}6498 & 1 & 0{,}4464 & 0{,}4366 & 0{,}4397 & 0{,}4212 & \dots & 0{,}3956 & 0{,}3851 \\ 0{,}4101 & 0{,}4464 & 1 & 0{,}9322 & 0{,}7465 & 0{,}7900 & \dots & 0{,}7431 & 0{,}7211 \\ 0{,}4145 & 0{,}4366 & 0{,}9322 & 1 & 0{,}7483 & 0{,}7862 & \dots & 0{,}7300 & 0{,}7085 \\ 0{,}4279 & 0{,}4397 & 0{,}7465 & 0{,}7483 & 1 & 0{,}7952 & \dots & 0{,}6226 & 0{,}6087 \\ 0{,}4231 & 0{,}4212 & 0{,}7900 & 0{,}7862 & 0{,}7952 & 1 & \dots & 0{,}6983 & 0{,}6936 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0{,}3953 & 0{,}3956 & 0{,}7431 & 0{,}7300 & 0{,}6226 & 0{,}6983 & \dots & 1 & 0{,}9347 \\ 0{,}3823 & 0{,}3851 & 0{,}7211 & 0{,}7085 & 0{,}6087 & 0{,}6936 & \dots & 0{,}9347 & 1 \end{bmatrix}. \quad (2)$$

Шаг 2. К полученной матрице (2) применим «max-min» операцию (рис. 1).

В результате сформируем матрицу толерантности:

$$\mathbf{T} = \begin{bmatrix} 1 & 0{,}6498 & 0{,}4464 & 0{,}4464 & 0{,}4464 & 0{,}4464 & \dots & 0{,}4464 & 0{,}4464 \\ 0{,}6498 & 1 & 0{,}4464 & 0{,}4464 & 0{,}4464 & 0{,}4464 & \dots & 0{,}4464 & 0{,}4464 \\ 0{,}4464 & 0{,}4464 & 1 & 0{,}9322 & 0{,}7900 & 0{,}7900 & \dots & 0{,}7951 & 0{,}7951 \\ 0{,}4464 & 0{,}4464 & 0{,}9322 & 1 & 0{,}7900 & 0{,}7900 & \dots & 0{,}7951 & 0{,}7951 \\ 0{,}4464 & 0{,}4464 & 0{,}7900 & 0{,}7900 & 1 & 0{,}7952 & \dots & 0{,}7900 & 0{,}7900 \\ 0{,}4464 & 0{,}4464 & 0{,}7900 & 0{,}7900 & 0{,}7952 & 1 & \dots & 0{,}7900 & 0{,}7900 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0{,}4464 & 0{,}4464 & 0{,}7951 & 0{,}7951 & 0{,}7900 & 0{,}7900 & \dots & 1 & 0{,}9347 \\ 0{,}4464 & 0{,}4464 & 0{,}7951 & 0{,}7951 & 0{,}7900 & 0{,}7900 & \dots & 0{,}9347 & 1 \end{bmatrix}. \quad (3)$$

Шаг 3. Проведем декомпозицию полученной на шаге 2 матрицы (3) нечеткой толерантности $\mathbf{T}$ в ранжированном виде $0 \le \alpha_1 < \dots < \alpha_l < \dots < \alpha_z \le 1$, что позволит найти значение порога сходства α_l:

$$\alpha_l = \{1{,}0000 \quad 0{,}9347 \quad 0{,}9322 \quad 0{,}7951 \quad \dots \quad 0{,}9000 \quad 0{,}6498 \quad 0{,}4464\}.$$

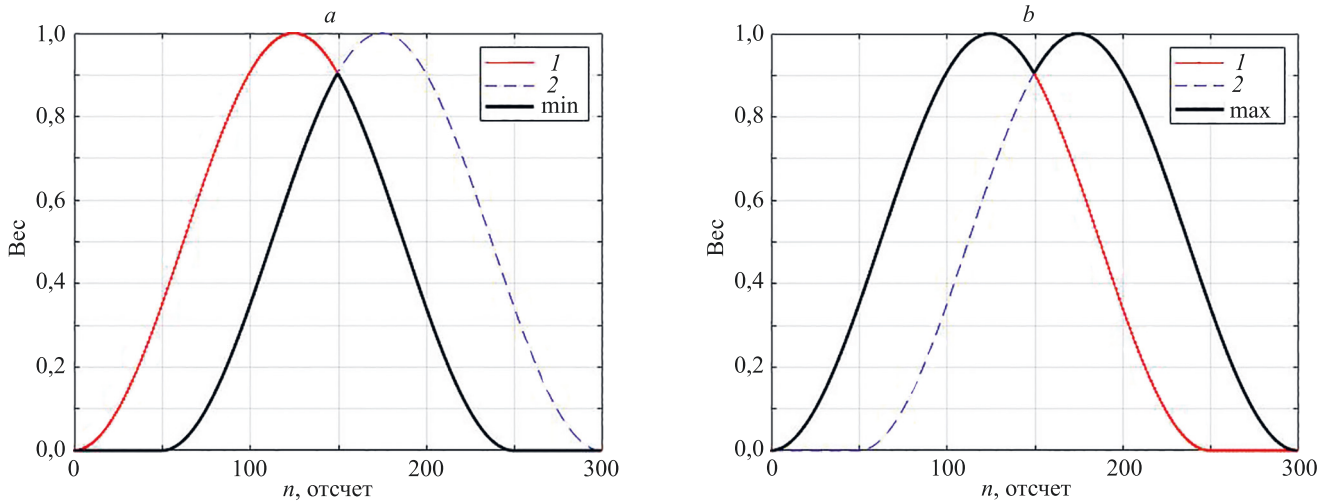


Рис. 1. Операция взятия min (*a*) и max (*b*)
1 — нижняя граница; *2* — верхняя граница

Fig. 1. Operation of taking the min (*a*) and max (*b*)

Шаг 4. Для вычисленного на шаге 3 значения α_l из последовательности $0 \le \alpha_1 < \ldots < \alpha_l < \ldots < \alpha_z \le 1$ построим нечеткое отношение:

$$\mathbf{T}_{(\alpha_l)} = \begin{bmatrix} 1 & 0{,}6498 & 0{,}4464 & 0{,}4464 & 0{,}4464 & 0{,}4464 & \ldots & 0{,}4464 & 0{,}4464 \\ 0{,}6498 & 1 & 0{,}4464 & 0{,}4464 & 0{,}4464 & 0{,}4464 & \ldots & 0{,}4464 & 0{,}4464 \\ 0{,}4464 & 0{,}4464 & 1 & 0{,}9322 & 0{,}7900 & 0{,}7900 & \ldots & 0 & 0 \\ 0{,}4464 & 0{,}4464 & 0{,}9322 & 1 & 0{,}7900 & 0{,}7900 & \ldots & 0 & 0 \\ 0{,}4464 & 0{,}4464 & 0{,}7900 & 0{,}7900 & 1 & 0{,}7952 & \ldots & 0{,}7900 & 0{,}7900 \\ 0{,}4464 & 0{,}4464 & 0{,}7900 & 0{,}7900 & 0{,}7952 & 1 & \ldots & 0{,}7900 & 0{,}7900 \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ 0{,}4464 & 0{,}4464 & 0 & 0 & 0{,}7900 & 0{,}7900 & \ldots & 1 & 0{,}9347 \\ 0{,}4464 & 0{,}4464 & 0 & 0 & 0{,}7900 & 0{,}7900 & \ldots & 0{,}9347 & 1 \end{bmatrix}. \quad (4)$$

Шаг 5. Для нечеткого отношения (4) построим начальное распределение по нечетким α-кластерам $R_l^{\alpha}(X) = \{A_{\alpha}^{l} | l = \overline{1, n}\}$, из элементов которого сформируем распределение $R_z^{\alpha_l}(X)$ и множество $B(\alpha_l) = \{R_z^{\alpha_l}(X)\}$.

Шаг 6. Для каждого $R_z^{\alpha_l}(X) \in B(\alpha_l)$ вычислим значение критерия $F(R_z^{\alpha}(X), \alpha)$.

Шаг 7. Произведем проверку выполнения критерия (шаг 6). Если критерий выполняется для некоторого единственного $R_z^{\alpha_l}(X) \in B(\alpha_l)$, то это распределение является искомым решением $R(X)$ для уровня α_l и алгоритм прекращает работу. Если критерий не выполняется, то построим множество распределений $B'(\alpha_l) \subseteq B(\alpha_l)$ и осуществим переход на шаг 8.

Шаг 8. Для каждого распределения $\{R_z^{\alpha_l}(X) \in B'(\alpha_l)$ выполним следующую последовательность операций.

Шаг 8.1. На начальном шаге предположим, что $l = 1$.

Шаг 8.2. Для нечеткого α-кластера $A_{\alpha_l}^{l} \in R_z^{\alpha_l}(X)$ выделим его носитель $Supp(A_{\alpha_l}^{l}) \in A_{\alpha_l}^{l}$ и путем вычеркивания соответствующих объектам $x_i \notin A_{\alpha_2}^{l}$, $i = \overline{1, n}$ строк в матрице $\mathbf{X}_{m \times n} = [x_i^t]$, $i = \overline{1, n}$, $t = \overline{1, m}$ исходных данных построим матрицу $\mathbf{X}_{m \times n_l} = [x_i^t]$, $x_i \in A_{\alpha_2}^{l}$, $t = \overline{1, m}$ «объект-свойство» кластера $A_{\alpha_l}^{l}$, где $n_l = card(A_{\alpha_l}^{l})$.

Шаг 8.3. Вычислим координаты геометрического центра $\bar{\tau}^l = \{\bar{x}^1, \ldots, \bar{x}^m\}$ кластера $A_{\alpha_l}^{l}$ в соответствии с выражением $\bar{x}_i^t = \frac{1}{n_l} \sum_{x_i \in A_{\alpha_l}^{l}} x_i^t$, $t = \overline{1, m}$ (рис. 2, *a*).

Шаг 8.4. Вычисляется расстояние $d(\tau^l, \bar{\tau}^l)$ между объектом τ^l, являющимся типичной точкой нечеткого α-кластера $A_{\alpha_l}^{l}$, и его геометрическим центром $\bar{\tau}^l$.

Шаг 8.5. Если просмотрены не все нечеткие α-кластеры $A_{\alpha_l}^{l} \in R_z^{\alpha_l}(X)$, то следует $l = l + 1$, и осуществим переход на шаг 8.2, иначе — переход на шаг 9.

Шаг 9. Произведем попарное сравнение нечетких α-кластеров $A_{\alpha_l}^{l}$, являющихся элементами различных распределений $R_z^{\alpha_l}(X) \in B'(\alpha_l)$, и в качестве решения $R(X)$ для уровня α_l выберем распределение, в котором расстояние $d(\tau^l, \bar{\tau}^l)$ является наименьшим для всех нечетких α-кластеров $A_{\alpha_l}^{l}$, и алгоритм прекращает работу.

Для проверки сходимости метода сформируем группу наборов данных, которые кластеризуем (рис. 2, *b*). Анализ полученных результатов показал, что разработанный метод устойчиво позволяет кластеризовать данные объектов для последующего принятия решения.

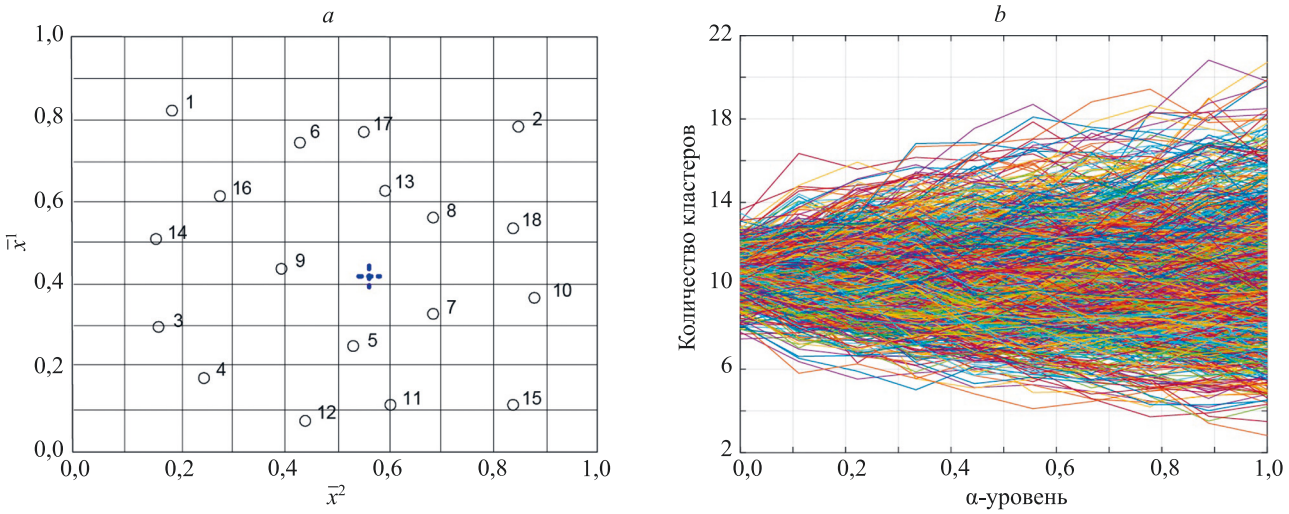


Рис. 2. Пояснения к реализованной методике: поиск геометрического центра кластера (*a*); сходимость методики (*b*)

Fig. 2. Explanations for the implemented method: search for the geometric center of the cluster (*a*); convergence of the technique (*b*)

Таким образом, предложенный метод основан на двух методах и представляет собой, с одной стороны, основанную на понятии нечеткого отношения уровня взвешенную модификацию кластер-процедуры [13], а с другой — иерархическую версию AFC-метода нечеткой кластеризации, отличительная особенность которой от изложенного метода в работе [12] заключается в применении операции нечеткой толерантности. Главным отличием предложенного алгоритма от обоих методов кластеризации является то, что матрицей исходных данных является матрица «объект-свойство» $\mathbf{X}_{m\times n} = [x_i^t]$, $i = \overline{1, n}$, $t = \overline{1, m}$ и для решения задачи кластеризации используется некоторая метрика $\mathbf{d}(x_i, x_j)$. С физической точки зрения сущность разработанного алгоритма заключается в том, что решением задачи кластеризации является иерархия распределений $R(X)$ по α-уровням, типичные точки которой находятся наиболее близко к геометрическим центрам соответствующих нечетких α-кластеров. Отметим, что ввиду применения операции «max-min» предложенная нечеткая кластер-процедура обладает более высокой скоростью по сравнению со стандартным AFC-методом и строит распределения $R(X)$ по полностью разделенным нечетким α-кластерам.

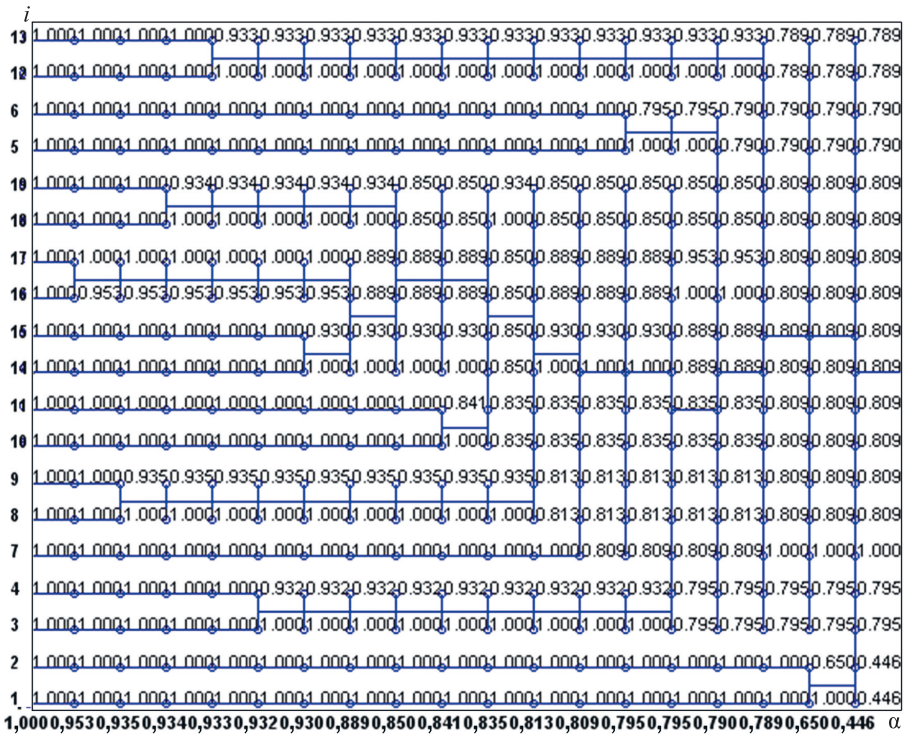


Рис. 3. Графическое представление иерархического распределения по α-уровням

Fig. 3. Graphical representation of hierarchical distribution by α-levels

Результаты численного эксперимента

При проведении вычислительного эксперимента в качестве метрики $\mathbf{d}(x_i, x_j)$ использовано, как и в работе [5], относительное евклидово расстояние. Результат применения предложенного алгоритма к исследуемой совокупности разнородных данных небольшого размера $X = \{x_1, \ldots, x_{19}\}$ анализируемых объектов в виде иерархии распределений по нечетким α-кластерам показан на рис. 3. В результате видно, что при наличии 19 объектов, предложенное решение позволило разбить объекты на 10 кластеров.

Распределение $R(X)$ при соответствующем значении уровня α можно представить в виде компонент взвешенного графа, рис. 4. Типичные точки данного графа нечетких α-кластеров (элементов распределения

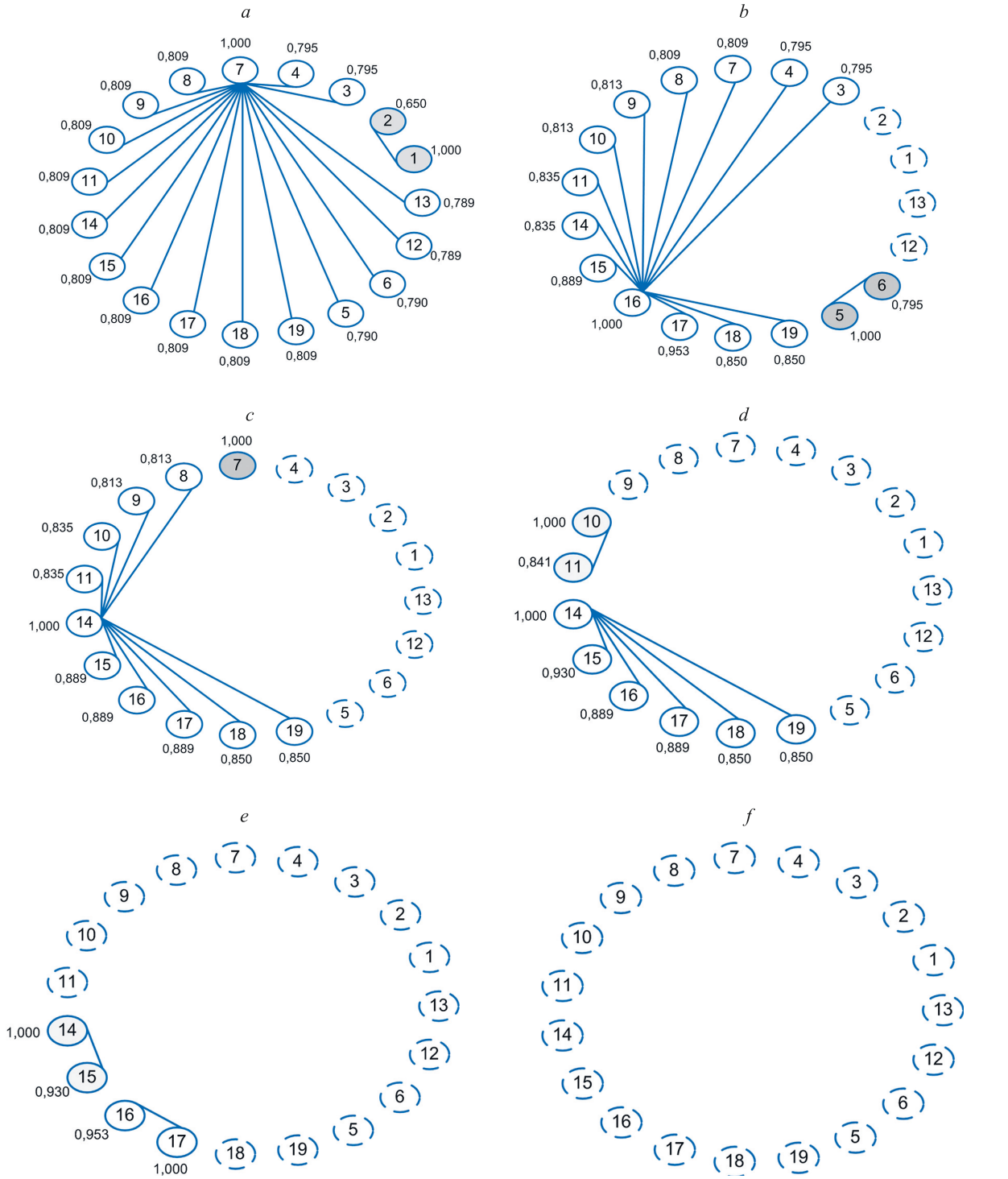


Рис. 4. Выборочное представление распределения в зависимости от α-уровня: α = 0,650 (*a*); α = 0,795 (*b*); α = 0,809 (*c*); α = 0,841 (*d*); α = 0,930 (*e*); α = 0,953 (*f*)

Fig. 4. Selective representation of the distribution depending on the α-level: α = 0.650 (*a*); α = 0.795 (*b*); α = 0.809 (*c*); α = 0.841 (*d*); α = 0.930 (*e*); α = 0.953 (*f*)

соответствующего значения уровня) имеют значения принадлежности, равные единице, а весовые коэффициенты дуг обозначают степень подобия объекта с типичной точкой соответствующего нечеткого α-кластера.

Типичные точки полученного распределения при соответствующем значении уровня могут рассматриваться, как эталоны и использоваться для дальнейшего сравнения с целью их идентификации [16].

Заключение

В ходе решения задачи предварительного разбиения совокупностей, существенно отличающихся по своей природе и имеющих ограниченный объем данных, предложен метод иерархической кластеризации, который позволяет существенно сократить время их разбиения. Выполнен анализ работы метода, который показал устойчивость разработанного алгоритма. Научной новизной метода является применение операции «max-min» нечеткого отношения сходства в составе иерархического алгоритма.

Полученные результаты наглядно показали, что метод может быть использован в качестве методологической основы перспективных систем идентификации, а также систем поддержки принятия решений.

Литература

1. Светашов А.К. Использование искусственных нейронных сетей для их применения в существующих и перспективных радиосистемах: тематическое исследование // Молодой ученый. 2023. № 22(469). С. 52–58.
2. Алимов Х.Т., Дзамихова Ф.Х., Паровик Р.И. Дробная математическая модель Макшерри // Вестник КРАУНЦ. Физико-математические науки. 2023. Т. 42. № 1. С. 164–179. https://doi.org/10.26117/2079-6641-2023-42-1-164-179
3. Salimian F., Damiri M., Ramezankhani M., Fariman S.K. Developing a new interval type-2 hesitant fuzzy TOPSIS-based fuzzy best-worst multicriteria decision-making method for competitive pricing in supply chain // Journal of Mathematics. 2022. V. 2022. P. 7879028. https://doi.org/10.1155/2022/7879028
4. Sinitsyn A.V., Lisay N.Yu., Selivanov S.A., Sinitsyn A.A. Development the information systems with fuzzy logic algorithms and network optimization // Информация и инновации. 2023. Т. 18. № 2. С. 33–47. https://doi.org/10.31432/1994-2443-2023-18-2-33-47
5. Gogoi S., Gohain B., Chutia R. Distance measures on intuitionistic fuzzy sets based on cross-information dissimilarity and their diverse applications // Artificial Intelligence Review. 2023. V. 56. Suppl. 3. P. 3471–3514. https://doi.org/10.1007/s10462-023-10608-y
6. Вятченин Д.А. Нечеткие методы автоматической классификации. Минск: Технопринт, 2004. 219 с.
7. Белов М.А., Гришко С.И., Живетьев А.В., Подгорный С.А., Токарева Н.А. Применение методов нечеткой логики для формирования адаптивной индивидуальной траектории обучения на основе динамического управления сложностью курса // Моделирование, оптимизация и информационные технологии. 2022. Т. 10. № 4. С. 7–8. https://doi.org/10.26102/2310-6018/2022.39.4.018
8. Эдгулова Е.К., Тхабисимова М.М., Бозиева А.М. Особенности построения баз знаний в интеллектуальных системах // Информационные технологии в экологии, образовании и бизнесе: материалы конференции. Нальчик: КБГУ, 2021. С. 168–174.
9. Zhan Q., Jin L., Yager R.R., Mesiar R.. A novel three-way decision method for interval-valued hesitant fuzzy environment // Soft Computing. 2023. V. 27. N 17. P. 12289–12307. https://doi.org/10.1007/s00500-023-08259-w
10. Zhao S., Wang D., Changyong L., Lu W. Induced choquet integral aggregation operators with single-valued neutrosophic uncertain linguistic numbers and their application in multiple attribute group decision-making // Mathematical Problems in Engineering. 2019. V. 2019. P. 9143624. https://doi.org/10.1155/2019/9143624
11. Zhang Y., Li P., Wang Y., Ma P., Su X. Multiattribute decision making based on entropy under interval-valued intuitionistic fuzzy environment // Mathematical Problems in Engineering. 2013. V. 2013. P. 526871. https://doi.org/10.1155/2013/526871
12. Tamura S., Higuchi S., Tanaka K. Pattern classification based on fuzzy relations // IEEE Transactions on Systems, Man, and Cybernetics. 1971. V. SMC-1. N 1. P. 61–66. https://doi.org/0.1109/TSMC.1971.5408605

References

1. Svetashov A.K. Using artificial neural networks for their application in the existing and future radio systems: a case study. *Young Scientis*, 2023, no. 22(469), pp. 52–58. (in Russian)
2. Alimov Kh.T., Dzamikhova F.Kh., Parovik R.I. Fractional mathematical model McSherry. Bulletin KRASEC. *Physical and Mathematical Sciences*, 2023, vol. 42, no. 1, pp. 164–179. (in Russian). https://doi.org/10.26117/2079-6641-2023-42-1-164-179
3. Salimian F., Damiri M., Ramezankhani M., Fariman S.K. Developing a new interval type-2 hesitant fuzzy TOPSIS-based fuzzy best-worst multicriteria decision-making method for competitive pricing in supply chain. *Journal of Mathematics*, 2022, vol. 2022, pp. 7879028. https://doi.org/10.1155/2022/7879028
4. Sinitsyn A.V., Lisay N.Yu., Selivanov S.A., Sinitsyn A.A. Development the information systems with fuzzy logic algorithms and network optimization. *Information and Innovations*, 2023, vol. 18, no. 2, pp. 33–47. https://doi.org/10.31432/1994-2443-2023-18-2-33-47
5. Gogoi S., Gohain B., Chutia R. Distance measures on intuitionistic fuzzy sets based on cross-information dissimilarity and their diverse applications. *Artificial Intelligence Review*, 2023, vol. 56, suppl. 3. pp. 3471–3514. https://doi.org/10.1007/s10462-023-10608-y
6. Viatchenin D.A. *Fuzzy Methods for Automatic Classification*. Minsk, Tehnoprint Publ., 2004, 219 p. (in Russian)
7. Belov M.A., Grishko S.I., Zhivetyev A.V., Podgorny S.A., Tokareva N.A. Use of fuzzy logic to create an adaptive individual learning path based on dynamic course complexity management. *Modeling, Optimization and Information Technology*, 2022, vol. 10, no. 4, pp. 7–8. (in Russian). https://doi.org/10.26102/2310-6018/2022.39.4.018
8. Edgulova E.K., Tkhabisimova M.M., Bozieva A.M. Features of building knowledge bases in intelligent systems. *Information Technologies in Ecology, Education and Business: Conference Proceedings*. Nalchik, KBSU, 2021, pp. 168–174. (in Russian)
9. Zhan Q., Jin L., Yager R.R., Mesiar R.. A novel three-way decision method for interval-valued hesitant fuzzy environment. *Soft Computing*, 2023, vol. 27, no. 17, pp. 12289–12307. https://doi.org/10.1007/s00500-023-08259-w
10. Zhao S., Wang D., Changyong L., Lu W. Induced choquet integral aggregation operators with single-valued neutrosophic uncertain linguistic numbers and their application in multiple attribute group decision-making. *Mathematical Problems in Engineering*, 2019, vol. 2019, pp. 9143624. https://doi.org/10.1155/2019/9143624
11. Zhang Y., Li P., Wang Y., Ma P., Su X. Multiattribute decision making based on entropy under interval-valued intuitionistic fuzzy environment. *Mathematical Problems in Engineering*, 2013, vol. 2013, pp. 526871. https://doi.org/10.1155/2013/526871
12. Tamura S., Higuchi S., Tanaka K. Pattern classification based on fuzzy relations. *IEEE Transactions on Systems, Man, and Cybernetics*, 1971, vol. SMC-1, no. 1, pp. 61–66. https://doi.org/0.1109/TSMC.1971.5408605
13. Sidorov D., Belov M. Design of hardware-software systems in the educational process with the use of Virtual Computer Lab. *System*

13. Сидоров Д.С., Белов М.А. Проектирование аппаратно-программных комплексов в учебном процессе с применением виртуальной компьютерной лаборатории // Системный анализ в науке и образовании. 2020. № 2. С. 70–82. https://doi.org/10.37005/2071-9612-2020-2-70-82
14. Viattchenin D.A. A new heuristic algorithm of fuzzy clustering // Control & Cybernetics. 2004. V. 33. N 2. P. 323–340.
15. Вятченин Д.А. Параметры AFC-метода нечеткой кластеризации // Вестник ВА РБ. 2004. № 4. С. 51–55.
16. Han Y., Huang Y., Jia S., Liu J. An interval-parameter fuzzy linear programming with stochastic vertices model for water resources management under uncertainty // Mathematical Problems in Engineering. 2013. V. 2013. P. 942343. https://doi.org/10.1155/2013/942343
17. Миловидова А.А., Черемисина Е.Н., Добрынин В.Н. Алгоритм определения типа и параметров функции принадлежности нечёткого измерителя // Современная наука: актуальные проблемы теории и практики. Серия: Естественные и технические науки. 2019. № 9. С. 69–74.

Analysis in Science and Education, 2020, no. 2, pp. 70–82. (in Russian). https://doi.org/10.37005/2071-9612-2020-2-70-82
14. Viattchenin D.A. A new heuristic algorithm of fuzzy clustering. *Control & Cybernetics*, 2004, vol. 33, no. 2, pp. 323–340.
15. Viatchenin D.A. Parameters of the AFC fuzzy clustering method. *Vestnik VA RB*, 2004, no. 4, pp. 51–55. (in Russian)
16. Han Y., Huang Y., Jia S., Liu J. An interval-parameter fuzzy linear programming with stochastic vertices model for water resources management under uncertainty. *Mathematical Problems in Engineering*, 2013, vol. 2013, pp. 942343. https://doi.org/10.1155/2013/942343
17. Milovidova A.A., Cheremisina E.N., Dobrynin V.N. Definition algorithm type and parameters of membership function of fuzzy measuring. *Modern Science: actual problems of theory and practice", a series "Natural and Technical Sciences*, 2019, no. 9, pp. 69–74. (in Russian)

Автор

Шарамет Андрей Владимирович — кандидат технических наук, доцент, начальник отдела ОАО «КБ Радар» — управляющая компания холдинга «Системы радиолокации», Минск, 220026, Республика Беларусь; докторант, Белорусский государственный университет информатики и радиоэлектроники, Минск, 220062, Республика Беларусь, https://orcid.org/0000-0003-0950-8700, a.sharamet@kbradar.by

Author

Andrei V. Sharamet — PhD, Associate Professor, Head of Department, JSC "KB Radar" — Managing Company of "Radar Systems" Holding, Minsk, 220026, Republic of Belarus; Doctoral Student, Belarusian State University of Informatics and Radioelectronics, Minsk, 220062, Republic of Belarus, https://orcid.org/0000-0003-0950-8700, a.sharamet@kbradar.by

Статья поступила в редакцию 19.10.2023
Одобрена после рецензирования 05.12.2023
Принята к печати 20.01.2024

Received 19.10.2023
Approved after reviewing 05.12.2023
Accepted 20.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-70-80
УДК 004.056.4

Исправление одиночных пакетов ошибок за пределами корректирующей способности кода с использованием информационных совокупностей

Мария Николаевна Исаева[1]✉, Андрей Анатольевич Овчинников[2]

[1,2] Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербург, 190008, Российская Федерация

[1] misaeva@hse.ru✉, https://orcid.org/0009-0007-6228-0617
[2] a.ovchinnikov@hse.ru, https://orcid.org/0000-0002-8523-9429

Аннотация

Введение. Исправление ошибок, возникающих при хранении, обработке, передаче информации является важнейшим методом обеспечения целостности данных. Для борьбы с возникающими ошибками используются методы помехоустойчивого кодирования. В реальных системах шумовые процессы обычно являются коррелированными, однако традиционные методы кодирования и декодирования используют декорреляцию, при этом известно, что эта процедура снижает предельно достижимые характеристики кодирования. Таким образом, актуальной является задача построения вычислительно эффективных методов декодирования, которые позволяли бы бороться с группирующимися ошибками при использовании широкого класса кодов. **Метод.** Для борьбы с одиночными пакетами ошибок использован подход, основанный на декодировании по информационным совокупностям. Несмотря на то, что при исправлении независимых ошибок данный метод имеет экспоненциальную сложность, предложенный подход применяет количество информационных совокупностей, линейно растущее с длиной кода, и обеспечивает, таким образом, полиномиальную сложность декодирования. Дальнейшее уменьшение числа информационных совокупностей возможно с помощью предложенного метода использования плотных информационных совокупностей. Выполнен анализ векторов ошибки, корректно исправляемых предложенными методами. Анализ проведен для кодов небольшой длины на основе стандартной расстановки, позволяющей оценить как множество ошибок, потенциально исправляемых кодом, так и характеристики декодера. **Основные результаты.** Предложен метод декодирования одиночных пакетов на основе выбора линейного числа информационных совокупностей. Описано улучшение метода декодирования с помощью использования счетчика векторов ошибки, позволяющее в ряде случаев увеличить число исправляемых векторов ошибки. Представлен метод декодирования на основании плотных информационных совокупностей, который позволяет значительно снизить количество информационных совокупностей или повысить количество исправляемых векторов ошибок по критерию минимальной длины пакета. Выполненный анализ рассмотренных декодеров с помощью стандартной расстановки показал, что предложенные алгоритмы позволяют исправлять значительное число векторов ошибки сверх гарантированно исправляемой длины пакета. **Обсуждение.** Предложенные декодеры позволяют исправлять одиночные пакеты ошибок за полиномиальное время для произвольных линейных кодов, при этом результаты экспериментов продемонстрировали, что декодеры не только исправляют все ошибки в пределах корректирующей способности кода, но и значительное количество векторов ошибки сверх нее. Направлениями дальнейших исследований возможен анализ предложенных алгоритмов декодирования для длинных кодов, где метод анализа на основе стандартной расстановки неприменим. Также могут быть осуществлены разработка и анализ методов декодирования для множественных пакетов и совместного исправления пакетирующихся и независимых ошибок.

Ключевые слова

информационные совокупности, корректирующая способность, низкоплотностные коды, каналы с памятью, пакеты ошибок

Благодарности

Статья подготовлена в результате проведения исследования в рамках Программы фундаментальных исследований Национального исследовательского университета «Высшая школа экономики» (НИУ ВШЭ), лаборатория Интернета вещей и киберфизических систем НИУ ВШЭ в Санкт-Петербурге.

© Исаева М.Н., Овчинников А.А., 2024

Ссылка для цитирования: Исаева М.Н., Овчинников А.А. Исправление одиночных пакетов ошибок за пределами корректирующей способности кода с использованием информационных совокупностей // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 70–80. doi: 10.17586/2226-1494-2024-24-1-70-80

Correction of single error bursts beyond the code correction capability using information sets

Maria N. Isaeva[1]✉, Andrei A. Ovchinnikov[2]

[1,2] HSE University, Saint Petersburg, 190008, Russian Federation

[1] misaeva@hse.ru✉, https://orcid.org/0009-0007-6228-0617

[2] a.ovchinnikov@hse.ru, https://orcid.org/0000-0002-8523-9429

Abstract

The most important method of ensuring data integrity is correcting errors that occur during information storage, processing or transmission. The error-correcting coding methods are used to correct errors. In real systems, noise processes are correlated. However, traditional coding and decoding methods use decorrelation, and it is known that this procedure reduces the maximum achievable characteristics of coding. Thus, constructing computationally efficient decoding methods that would correct grouped errors for a wide class of codes is an actual problem. In this paper the decoding by information sets is used to correct single bursts. This method has exponential complexity when correcting independent errors. The proposed approach uses a number of information sets linearly growing with code length, which provides polynomial decoding complexity. A further reduction of the number of information sets is possible with the proposed method of using dense information sets. It allows evaluating both the set of errors potentially corrected by the code and the characteristics of the decoder. An improvement of the decoding method using an error vector counter is proposed, which allows in some cases to increase the number of corrected error vectors. This method allows significantly reducing the number of information sets or increasing the number of corrected error vectors according to the minimum burst length criterion. The proposed decoders allow correction of single error bursts in polynomial time for arbitrary linear codes. The results of experiments based on standard array show that decoders not only correct all errors within the burst correcting capability of the code, but also a significant number of error vectors beyond of it. Possible directions of further research are the analysis of the proposed decoding algorithms for long codes where the method of analysis based on the standard array is not applicable; the development and analysis of decoding methods for multiple bursts and the joint correction of grouped and random errors.

Keywords

information sets, error correcting capability, low-density parity-check codes, channels with memory, error bursts

Acknowledgements

The article was prepared within the framework of the Basic Research Program at HSE University, Internet of Things and Cyber-Physical Systems Laboratory, St. Petersburg School of Physics, Mathematics, and Computer Science.

For citation: Isaeva M.N., Ovchinnikov A.A. Correction of single error bursts beyond the code correction capability using information sets. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 70–80 (in Russian). doi: 10.17586/2226-1494-2024-24-1-70-80

Введение

С развитием цифровых технологий передача и обработка информации нашла применение во многих сферах деятельности человека. В современных сетях и инфокоммуникационных системах необходимо передавать огромные объемы информации на большие расстояния, обеспечивая требования по скорости передачи, надежности и достоверности. Передача информации осуществляется по каналам связи, особенности которых приводят к появлению ошибок. Вопросы надежной доставки сообщений могут решаться с помощью внесения избыточности и управления ею с помощью резервирования [1], другим подходом является использование методов помехоустойчивого кодирования [2, 3].

Корректирующая способность кодов может быть формально определена, например, как исправление любой комбинации из заданного числа ошибок (декодирование в сфере Хэмминга), либо декодирование всех лидеров смежных классов (полное декодирование) [3]. Вместе с тем для большинства методов декодирования современных кодов (таких как турбо-коды, полярные коды, коды с малой плотностью проверок на четность) множество исправляемых ошибок не имеет простого описания, поэтому вероятность ошибки, обеспечиваемая такими декодерами, определяется путем интенсивного компьютерного моделирования, и актуальным является вопрос о возможности как кодов исправлять ошибки сверх корректирующей способности, так и декодеров реализовывать эти возможности.

В реальных каналах ошибки имеют тенденцию группироваться, образуя так называемые пакеты. Эффект памяти может вызываться многолучевостью распространения данных, архитектурой систем хранения, в том числе распределенных, медленной флюктуацией параметров канала и т. п. [4–6]. Для борьбы с данными пакетами может быть применено перемежение, приводящее к росту задержки обработки информации на передатчике и приемнике, а также ухудшающие предельные возможности кодирования. Другим подходом является построение специальных кодов для таких каналов и декодеров, исправляющих пакеты ошибок. Коды и декодирование для каналов с памятью исследованы крайне мало, что не позволяет говорить не только

о приближении к предельным характеристикам, но и о превышении характеристик традиционных кодов для независимых ошибок в сочетании с перемежением.

В теории кодирования одним из классических способов борьбы с группированием ошибок являются коды Рида–Соломона, в том числе с использованием информации о надежностях принятых символов [7], однако декодирование таких кодов достаточно вычислительно затратно по сравнению с современными методами. Способность исправлять ошибки в каналах с памятью исследовалась для кодов с малой плотностью проверок на четность [8–10], в том числе для моделей каналов с двумя состояниями [11, 12], а также для полярных кодов [13]. Отметим, что в большинстве научных работ часто рассматриваются очень специфические конфигурации ошибок [8] или проводится оптимизация кодовых конструкций, к которым затем применяются стандартные методы декодирования для независимых ошибок. В работе [14] предложен декодер для исправления одиночных пакетов, теоретически применимый к любым линейным кодам, однако он вычислительно эффективен только для кодов с малой плотностью проверок на четность.

Целью данной работы является разработка алгоритма декодирования для исправления пакетов ошибок, который может быть применен к любым линейным кодам, и проведение анализа способности исправлять группирующиеся ошибки сверх корректирующей способности. Характеристики декодирования оцениваются с помощью анализа стандартной расстановки для кодов небольшой длины, для экспериментов выбраны блочно-перестановочные (БП) коды с малой плотностью проверок на четность, широко используемые в современных стандартах связи, а также случайные линейные коды, чья корректирующая способность близка к предельной при исправлении одиночных пакетов ошибок.

Модель ошибок и стандартная расстановка

Рассмотрим модель дискретного двоичного канала, ошибки в котором описываются не с помощью переходных вероятностей, а комбинаторно. Таким образом, изучим не вероятности ошибки декодирования, а способность декодера исправлять конкретные комбинации ошибок. При этом рассмотрим передачу информации блоками по n бит. Введем обозначение одиночных пакетов длиной b как число позиций от первой до последней ненулевой компоненты в векторе из n элементов. Корректирующей способностью кода при исправлении одиночных пакетов обозначим величиной $b_{\max}$, при которой любой вектор ошибки, представляющий собой пакет длиной $b \le b_{\max}$, будет корректно исправлен. Заметим, что определение количества исправляемых независимых ошибок является NP-трудной задачей [15], в то время как максимальная длина исправляемого одиночного пакета может быть вычислена с помощью полиномиальной процедуры [14].

В общем случае множество ошибок, исправляемых линейным кодом при передаче по дискретным каналам связи, может быть описано (хотя и с экспоненциальной сложностью) с помощью стандартной расстановки, представляющей собой таблицу всех возможных двоичных векторов длиной n, в которой первая строка состоит из 2^k кодовых слов, все остальные строки называются смежными классами и представляют собой сумму первой строки с некоторым вектором, не являющимся кодовым словом.

Линейный (n,k)-код — k-мерное подпространство n-мерного пространства двоичных векторов. Базис $\mathbf{G}$ n-мерного пространства называется порождающей матрицей, а базис $\mathbf{H}$ ортогонального пространства размерности $r = n - k$ — проверочной матрицей кода. Для любого вектора $\mathbf{y}$ длиной n вектор $\mathbf{S} = \mathbf{y}\mathbf{H}^T$ называется синдромом.

Все векторы в смежном классе имеют одинаковый синдром, у любых векторов из разных смежных классов синдром различен. Из каждого смежного класса может быть выбран один представитель, который называется лидером смежного класса, таким образом, существует взаимно однозначное соответствие между синдромом и лидером смежного класса, и синдром может рассматриваться как номер смежного класса в стандартной расстановке. Лидер смежного класса считается вектором ошибки для данного синдрома при декодировании по стандартной расстановке [3], другими словами, если $\mathbf{y}$ — принятое из канала слово, синдром которого равен $\mathbf{S}$, то результатом декодирования считается кодовое слово $\mathbf{a} = \mathbf{y} - \mathbf{e}_{\mathbf{S}}$, где $\mathbf{e}_{\mathbf{S}}$ — лидер смежного класса с номером $\mathbf{S}$. Таким образом, если расположить все лидеры в начале соответствующих смежных классов, то первый столбец стандартной расстановки представляет собой список исправляемых векторов ошибок для данного кода. Это означает, что если корректирующая способность кода при исправлении пакетов ошибок равна $b_{\max}$, то все векторы ошибок, представляющие собой пакеты длиной $b \le b_{\max}$, могут быть выбраны в качестве лидеров своих смежных классов (т. е. никакие два таких вектора не лежат в одном классе).

Исправление векторов ошибок в рамках корректирующей способности кода называют декодированием в сфере. Вместе с тем код может исправлять значительное количество векторов ошибок за пределами сферы (т. е. с длиной пакетов, превышающей $b_{\max}$), так как код может исправить все ошибки, являющиеся лидерами смежных классов. Декодирование, при котором любой лидер смежного класса может быть исправлен, называется полным.

В качестве лидеров смежных классов выбирают самые вероятные векторы в канале (тогда полное декодирование называется декодированием по максимуму правдоподобия) или по минимуму веса в некоторой метрике (что соответствует декодированию по минимальному расстоянию). В настоящей работе предположим, что в качестве лидеров выбраны векторы смежного класса, образующие одиночные пакеты наименьшей длины (заметим, что длина пакета не является метрикой, так как для нее не выполняется неравенство треугольника, что не позволяет связать длину исправляемого пакета с расстоянием между кодовыми словами, как это происходит в случае независимых ошибок и метрики Хемминга).

Отметим, что практически все используемые на сегодняшний день декодеры не обеспечивают полного декодирования, таким образом, не реализуя потенциальные возможности кодов. Некоторые декодеры могут гарантировать декодирование в сфере (например, алгебраические декодеры циклических кодов) или не иметь простого описания исправляемых векторов ошибок. Оценка множества исправляемых декодером ошибок является важной задачей, позволяющей как получить более точное понимание свойств декодера, так и проводить аналитические оценки вероятностей ошибки без интенсивного компьютерного моделирования.

Декодирование по информационным совокупностям при исправлении пакетов ошибок

Декодирование по информационным совокупностям является лучшим (по сложности) из известных способов декодирования случайных линейных кодов [16]. Тем не менее, при исправлении независимых ошибок сложность такого декодирования экспоненциальна и определяется числом информационных совокупностей, необходимых для достижения малых вероятностей ошибки. Вместе с тем сложность такого декодирования может быть существенно меньше (полиномиальна) при исправлении пакетов ошибок. Это связано с тем, что для такого декодирования требуется существенно меньше информационных совокупностей.

Множество вида $\gamma = \{1 \le j_1 < j_2 < \ldots < j_k \le n\}$ называется информационной совокупностью, если компоненты кодового слова с номерами из γ однозначно определяют все кодовое слово [5]. Информационная совокупность называется свободной от ошибок, если на ее позициях не произошли ошибки (в векторе ошибок нет ненулевых элементов). В таком случае, принятое слово может быть продекодировано корректно, однако проблемой является определение того, что информационная совокупность свободна от ошибок. Таким образом, декодирование по информационной совокупности состоит в генерировании достаточно большого набора совокупностей и перебора по этому набору в поиске информационной совокупности, свободной от ошибок.

При декодировании пакетов ошибок необходимо учитывать их расположение: генерировать информационную совокупность на позициях, на которых нет позиций вектора ошибки. Пусть код с k информационными символами задан порождающей матрицей $\mathbf{G}$ размера $(k \times n)$. Приведем алгоритм декодирования по информационным совокупностям при исправлении пакетов ошибок с использованием порождающей матрицы $\mathbf{G}$.

Этап 1. Инициализация.

Шаг 1.1. Генерируется множество информационной совокупности $\Gamma = \{\gamma_1, \ldots, \gamma_N\}$.

Шаг 1.2. Формируется набор матриц $\mathbf{G}_{\gamma_i} = \mathbf{M}_{\gamma_i}^{-1}\mathbf{G}$, где $\mathbf{M}_{\gamma_i}$ — это подматрица матрицы $\mathbf{G}$, составленная из столбцов с номерами из множества γ_i.

Этап 2. Декодирование принятого слова $\mathbf{b} = \mathbf{a} \oplus \mathbf{e}$, где $\mathbf{a}$ — кодовое слово; $\mathbf{e}$ — пакет ошибок.

Шаг 2.1. Установить $b_{\min} = n$ и $\mathbf{z}_{\min} = \mathbf{b}$.

Шаг 2.2. Для всех $i = 1 \ldots N$:

— вычислить $\mathbf{z}_i = \mathbf{b}(\gamma_i)\mathbf{G}_{\gamma_i}$ где $\mathbf{b}(\gamma_i)$ — элементы из принятого слова $\mathbf{b}$ на позициях γ_i;

— вычислить $\mathbf{e}_i = \mathbf{b} \oplus \mathbf{z}_i$ и длину b_i пакета в векторе $\mathbf{e}_i$;

— если $b_i < b_{\min}$, то $\mathbf{z}_{\min} = \mathbf{z}_i$ и $b_{\min} = b_i$.

Этап 3. Принятие решения о декодированном слове: $\hat{\mathbf{a}} = \mathbf{z}_{\min}$.

Отметим, что в представленном алгоритме декодирования решение принимается в пользу кодового слова, для которого минимальна длина произошедшего одиночного пакета ошибок. Это обусловлено тем, что несмотря на то, что не была введена вероятностная модель ошибок, наиболее вероятны более короткие пакеты, чем более длинные — данное соображение согласуется со многими вероятностными моделями каналов с памятью, например моделями с конечным числом состояний (Гилберта или Гилберта–Эллиотта).

Сложность декодирования в рассмотренном алгоритме определяется числом информационных совокупностей N. Чтобы для каждого возможного расположения пакета длиной b (т. е. позиций начала пакета от 1 до $n - b + 1$) нашлась информационная совокупность, свободная от ошибок, необходимо сформировать множество Γ, состоящее из информационных совокупностей на позициях, не вошедших в пакет. Тогда $N = n - b + 1$ и сложность алгоритма декодирования является полиномиальной (кроме перебора по линейному от n количеству информационных совокупностей, необходимо хранить множество Γ, матрицы $\mathbf{G}_{\gamma_i}$ и вычислить $\mathbf{z}_i$). Более точно, вычислительная сложность этапа 1 не учитывается при оценке сложности декодирования, так как для выбранного кода выполняется один раз. При этом при случайном выборе информационной совокупности в случайном коде вероятность ее нахождения составляет около 0,3 [17]. Таким образом, шаг 1.1 выполняется ожидаемо за $3n$ попыток, при $n \approx N$. Шаг 1.2 требует $O(n)$ обращений матриц и столько же матричных умножений. Результатом этапа 1 является хранение $O(n)$ порождающих матриц размером $k \times n$, т. е. $O(kn^2)$ бит, а также kn ячеек памяти на хранение множества Γ. Цикл на этапе 2 шага 2.1 выполняется $O(n)$ раз, каждая итерация цикла требует $k + 1$ операций «исключающее ИЛИ» векторов длиной n, а также определения длины пакета со сложностью $O(n)$, что дает вычислительную сложность $O(kn)$ битовых операций. Таким образом, можно без применения оптимизации оценить сложность декодера как $O(kn^2)$ памяти и операций. Для сравнения, асимптотическое время выполнения декодера кода Рида–Соломона составляет $O(n^2)$, однако операции проводятся в конечном поле $GF(q)$, что более затратно при практической реализации.

Параметр b наиболее естественно выбрать, исходя из корректирующей способности кода, $b \le b_{\max}$, однако при выборе больших значений b алгоритм будет пытаться продекодировать сверх корректирующей способности.

Заметим, что в рассмотренном алгоритме в процессе декодирования формируется список возможных векторов ошибки для каждой информационной совокупности, и выбор происходит в пользу вектора ошиб-

ки, представляющего собой пакет наименьшей длины. Если пакет с такой длиной один, то решение декодера однозначно, однако пакетов ошибок с одинаковой минимальной длиной может быть несколько. В этом случае декодер может выбрать произвольный пакет ошибок; сигнализировать о невозможности принять однозначное решение или воспользоваться каким-либо дополнительным критерием выбора.

Рассмотрим следующую модификацию декодера, установив дополнительный критерий отбора вектора ошибки при декодировании, кроме длины пакета — счетчик встречаемости вектора ошибки в результирующем списке пакетов минимальной длины. Разные информационные совокупности, на которых отсутствуют ошибки, будут восстанавливать одно и то же кодовое слово, и следовательно — один и тот же вектор ошибки. Тогда можно ожидать, что верный вектор будет встречаться чаще других. Таким образом, если после декодирования получилось несколько векторов ошибки с одинаковой минимальной длиной пакета, выберем среди них тот, который встречается в списке чаще всего. При этом в случае, если есть несколько векторов с одинаковым значением счетчика, снова будет получена неоднозначность результата декодирования.

Рассмотрим подход к уменьшению количества используемых информационных совокупностей. Пусть b_1 и b_2 — номера начала и конца пакета в некотором векторе ошибок. Предположим, что происходит циклическое движение вправо по позициям вектора ошибки, начиная от позиции $b_2 + 1$ (при достижении конца вектора движение продолжится с начала вектора). Первую встреченную при таком обходе позицию информационной совокупности обозначим j_1, последнюю — j_k. При такой нумерации назовем плотной информационную совокупность, у которой минимизируется разность $(j_k - j_1)\mathrm{mod}n$, т. е. минимизируется разброс от первой до последней позиций информационной совокупности с учетом цикличности номеров позиций.

Использование таких плотных информационных совокупностей позволяет сократить количество информационных совокупностей при декодировании, так как одна информационная совокупность может оказаться свободной от ошибок для большего количества расположений пакетов ошибок.

При таком подходе можно значительно уменьшить количество информационных совокупностей (эксперименты показывают, что в ряде случаев — до двух или трех, в зависимости от параметров кода и максимальной исправляемой длины пакета). При использовании плотных информационных совокупностей можно считать $N = \mathrm{const}$, тогда декодирование требует $O(kn)$ памяти и битовых операций.

Отметим, что декодирование с дополнительным использованием счетчика может не дать заметного эффекта для плотных информационных совокупностей, так как значение счетчика может просто не накопиться из-за малого количества плотных информационных совокупностей. В связи с этим рассмотрим дополнительный режим декодирования, при котором построим плотные информационные совокупности для каждого возможного расположения пакета. В этом случае цель использования плотных информационных совокупностей — не минимизация числа информационных совокупностей, а накопление значения счетчика для правильного слова при декодировании по информационным совокупностям, свободным от ошибок.

Описание экспериментов и классификация результатов декодирования

В разделе «Модель ошибок и стандартная расстановка» полное декодирование определено как исправление всех лидеров смежных классов, выбранных по какому-либо критерию. Однако декодер, обеспечивающий полное декодирование, можно определить при другом условии: пусть для каждого возможного синдрома декодер верно декодирует какой-либо вектор ошибки из соответствующего смежного класса. Другими словами, это означает, что существует такой способ выбора списка лидеров смежного класса, что рассматриваемый декодер будет их верно декодировать, если они произошли в канале. При этом, если лидеры смежных классов выбраны традиционно в соответствии с некоторым критерием (например, минимальным весом Хемминга, минимальной длиной пакета, максимальной вероятностью и т. д.), такое декодирование, являясь полным в смысле данного определения, может не обеспечивать полное декодирование по заданному критерию.

Оценим множество векторов ошибок, исправляемых описанным декодером по информационным совокупностям. Выполним оценку экспериментально, на примере кодов небольшой длины, для которых можно результаты декодирования сравнить с расположением векторов в стандартной расстановке. Для проведения экспериментов выберем следующие коды.

БП-коды с малой плотностью проверок на четность БП(γ, ρ) задаются $(m\gamma \times m\rho)$-проверочной матрицей, состоящей из блоков — степеней матрицы циклической перестановки [18, 19], т. е. имеющей вид:

$$\mathbf{H} = \begin{bmatrix} \mathbf{C}^{t_{11}} & \dots & \mathbf{C}^{t_{1\rho}} \\ \dots & \dots & \dots \\ \mathbf{C}^{t_{\gamma 1}} & \dots & \mathbf{C}^{t_{\gamma\rho}} \end{bmatrix},$$

где $\mathbf{C}$ — матрица циклической перестановки; ρ — количество блоков в строке; γ — количество блоков в столбце; m — размер блока; степени t_{ij} — целые неотрицательные числа. Для задания такой матрицы достаточно указать (γ, ρ)-матрицу степеней t_{ij}, называемую базовой матрицей.

Для исследования выбраны пять кодов с параметрами, приведенными в табл. 1.

Для БП-кодов максимальная длина исправляемого пакета $b_{\max} \le m - 1$ [14]. Выбран код 1 с базовой матрицей

$$\mathbf{H}_{\mathbf{base}} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 2 \end{bmatrix} \qquad (1)$$

с размером блоков $m = 7$. Получена для кода 1 скорость $R = 0{,}38$ и $b_{\max} = 6$, и для которого $N_{\min} = 2$ плотные информационные совокупности покрывают все возможные расположения пакета. Код, задаваемый матрицей (1), является кодом Гилберта, его способность

Таблица 1. Параметры выбранных кодов
Table 1. Parameters of selected codes

Номер кода	Конструкция	n	k	R	$N_{\min}$	$b_{\max}$	Граница Рейгера
1	БП(2,3)	21	8	0,38	2	6	6
2	БП(2,4)	20	11	0,55	3	4	4
3	случайный	20	7	0,35	3	5	6
4	случайный	20	10	0,5	4	3	5
5	случайный	20	13	0,65	5	2	3

исправлять одиночные пакеты ошибок рассмотрены в работе [20].

Рассмотрим более высокоскоростной код 2 со скоростью $R = 0{,}55$, для которого базовая матрица имеет вид

$$\mathbf{H_{base}} = \begin{bmatrix} 0 & 2 & 1 & 1 \\ 2 & 3 & 4 & 1 \end{bmatrix} \qquad (2)$$

и размер блоков $m = 5$. Для этого кода $b_{\max} = 4$, $N_{\min} = 3$. Особенности нахождения информационных совокупностей в БП-кодах приведены в [17].

Выберем три случайных линейных кода, сгенерировав порождающие матрицы с вероятностью элементов в них 0,5: код 3 ($R = 0{,}35$, $b_{\max} = 5$, $N_{\min} = 3$), код 4 ($R = 0{,}5$, $b_{\max} = 3$, $N_{\min} = 4$) и код 5 ($R = 0{,}65$, $b_{\max} = 2$, $N_{\min} = 5$).

Отметим, что в соответствии с границей Рейгера для линейных (n, k)-кодов, максимальная длина одиночного исправляемого пакета не превышает $b_{\max} \le \lfloor (n-k)/2 \rfloor$ [3]. При этом для выбранных кодов их корректирующая способность лежит на границе или близка к ней.

Для проведения эксперимента построим стандартную расстановку для выбранного кода, отбирая в лидеры смежных классов пакеты наименьшей длины. Поставим на вход декодера все возможные векторы ошибок (не только лидеры смежных классов), т. е. будем декодировать каждый вектор стандартной расстановки. Предположим, что декодер в качестве результата выдает вектор ошибки, учитывая, что результатом декодирования также может быть список таких векторов (в случае, если несколько векторов ошибки имеют одинаковую минимальную длину пакета), приводящий к неоднозначности принятия решения. Пусть «*F*» — обозначение того, что декодером получен однозначный ответ, тем не менее, не являющийся правильным вектором ошибки (т. е. однозначное и неверное декодирование); «*T*» — получен однозначный ответ, совпадающий с правильным вектором ошибки (однозначно верно); «*T**» — получено несколько векторов ошибки с минимальной длиной пакета, но среди них есть верный; «*F**» — получено несколько векторов минимальной длины пакета, и среди них нет верного.

Выполним анализ декодирования с точки зрения способности декодера исправлять пакеты в пределах корректирующей способности, а также за ее пределами. Для этого введем следующую классификацию результатов декодирования.

Назовем смежный класс декодируемым (однозначно декодируемым), если в нем есть единственный вектор ошибок, однозначно исправляемый данным декодером. Назовем смежный класс вероятностно декодируемым, если исправляемых векторов ошибки в нем нет, но в некоторых случаях декодер возвращает список, содержащий правильный вектор ошибки. Для классификации декодера выделим следующие результаты декодирования для каждого смежного класса.

Случай I (декодируемый смежный класс по выбранному критерию): смежный класс декодируется верно и однозначно в пакет минимальной длины. Это означает, что для векторов ошибки из данного смежного класса один раз декодер вернул результат «*T*», и этот результат был получен для лидера.

Продекодированный вектор ошибок может не находиться в лидерах смежного класса, выбранных по некоторому принципу, в рассматриваемом случае — по минимальной длине пакета. При этом возможны два варианта:

вариант 1 — в смежном классе есть несколько пакетов одинаковой минимальной длины (это возможно только для пакетов длиной, большей $b_{\max}$), и продекодирован пакет минимальной длины, но не тот, который был выбран в лидеры;

вариант 2 — верно и однозначно продекодирован пакет, более длинный, чем лидер смежного класса.

При варианте 1 выполним переупорядочение стандартной расстановки (смежного класса) по результатам декодирования, и выберем верно продекодированный пакет в лидеры, при этом не меняется распределение длин исправляемых пакетов, находящихся в лидерах. Этот вариант не будем отличать от случая I, когда верно продекодированным вектором ошибок сразу оказался лидер. При варианте 2 получим следующий случай.

Случай I* (декодируемый смежный класс): при однозначном и верном декодировании смежного класса правильно продекодированный вектор ошибок не является лидером и не является пакетом минимальной длины. Это означает, что для данного смежного класса (и соответствующего синдрома) декодер не обеспечивает декодирование по выбранному критерию, однако смежный класс имеет единственный верно декодируемый вектор ошибки.

Случай II (вероятностно декодируемый смежный класс по выбранному критерию): лидеру смежного класса соответствует результат декодирования «*T**», т. е. декодер возвратил список векторов ошибки, среди которых есть и правильный, и при случайном выборе есть вероятность верного декодирования, зависящая от размера списка (можно ожидать, что он не будет боль-

шим). При этом предполагается, что результатов «*T*» в смежном классе нет.

Случай II* (вероятностно декодируемый смежный класс): аналогично случаю I*, смежный класс содержит результаты декодирования «*T**», однако они не соответствуют векторам с минимальной для данного смежного класса длиной пакета.

Случай III (не декодируемый смежный класс): результаты декодирования векторов смежного класса равны «*F*» или «*F**», т. е. вероятность правильного декодирования любого вектора ошибки из данного смежного класса равна нулю, вне зависимости от того, принял декодер однозначное решение либо нет.

Схематически описанные случаи представлены на рис. 1. Основываясь на выделенных результатах декодирования смежных классов, декодирование может быть классифицировано следующим образом:

— полное декодирование по выбранному критерию: декодирование каждого смежного класса относится к случаю I;
— полное декодирование: декодирование каждого смежного класса относится к случаям I или I*;
— вероятностное полное декодирование по выбранному критерию: декодирование каждого смежного класса относится к случаям I или II;
— вероятностное полное декодирование: декодирование каждого смежного класса относится к случаям I, I*, II или II*;
— декодирование в сфере по выбранному критерию: декодирование смежных классов с лидерами, имеющими длину пакета, не превышающую b_{max}, соответствует случаю I;
— декодирование в сфере: декодирование смежных классов с лидерами, имеющими длину пакета, не превышающую b_{max}, соответствует случаям I или II;
— вероятностное декодирование в сфере по выбранному критерию: декодирование смежных классов с лидерами, имеющими длину пакета, не превышающую b_{max}, соответствует случаям I или I*;
— вероятностное декодирование в сфере: декодирование смежных классов с лидерами, имеющими длину пакета, не превышающую b_{max}, соответствует случаям I, I*, II или II*.

Заметим, что декодирование по информационным совокупностям является вероятностным в том смысле, что при одинаковых параметрах (числе информационных совокупностей *N*) результаты декодирования могут зависеть от конкретного множества информационных совокупностей, которое строится случайно.

При использовании декодера с дополнительным принятием решения по счетчику может быть достигнуто улучшение полученных характеристик, так как повысится доля декодируемых смежных классов. Для этого декодера введем дополнительные обозначения результатов декодирования: «T_c» (*True with counter*) — однозначное и верное декодирование после использования значения счетчика (т. е. результат «*T**» стал результатом «*T*»); «F_c» — однозначное неверное декодирование после выбора по счетчику; «T_c*» — неоднозначность не была разрешена (среди векторов ошибки с одинаковой минимальной длиной пакета есть несколько, имеющих одинаковое значение счетчика), но в списке присутствует верный вектор; «F_c*» — неоднозначность не была разрешена, и верный вектор в списке отсутствует. Теперь однозначно, верно, декодированным векторам соответствуют результаты декодирования «*T*» или «T_c». В табл. 2 представлены параметры декодирования, которые будут использованы в ходе экспериментов.

Результаты экспериментов

С учетом введенных в разделе «Описание экспериментов и классификация результатов декодирования» обозначений проведем эксперименты с кодом с базовой матрицей (1) (код 1, табл. 1). Так как код 1 задается (21,8)-матрицей, то его стандартная расстановка содержит 8191 смежных класса (не считая самого множества

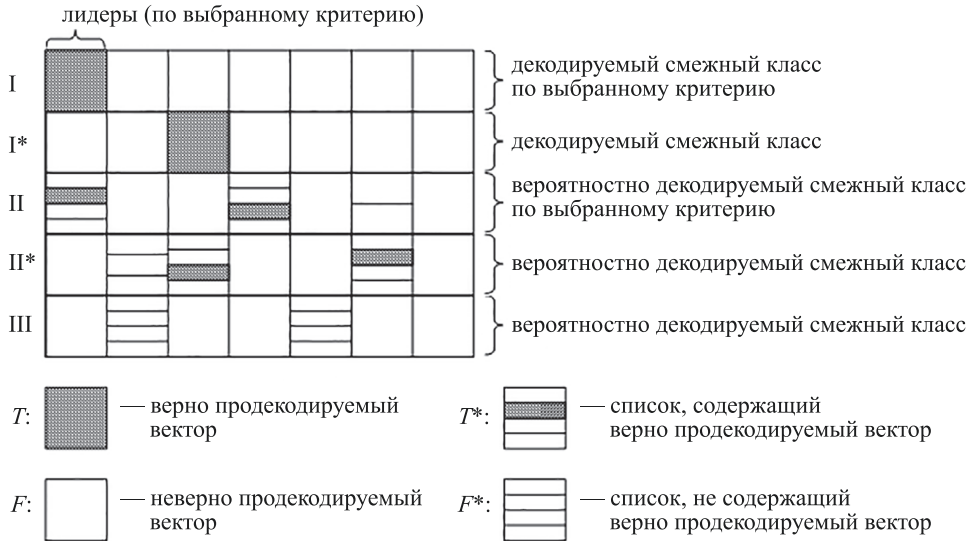


Рис. 1. Классификация результатов декодирования по смежным классам

Fig. 1. Classification of decoding results by cosets

Таблица 2. Параметры декодирования
Table 2. Decoding parameters

Обозначение	Параметры декодирования
A	$N = n - b + 1$ информационная совокупность для каждого возможного начала пакета;
B	$N = n - b + 1$ информационная совокупность для каждого возможного начала пакета и использование счетчика;
C	$N_{\min}$ плотных информационных совокупностей;
D	$N_{\min}$ плотных информационных совокупностей и использование счетчика;
E	$N = n - b + 1$ плотных информационных совокупностей для каждого возможного начала пакета;
F	$N = n - b + 1$ плотных информационных совокупностей для каждого возможного начала пакета и использование счетчика.

кодовых слов), каждый смежный класс содержит 256 векторов.

На рис. 2 представлены результаты декодирования векторов из стандартной расстановки. Базовый алгоритм декодирования помечен как (1), с использованием счетчика — как (2). Представлена доля декодируемых смежных классов (с выделением случая декодирования в рамках корректирующей способности), вероятностно декодируемых и не декодируемых.

Заметим, что во всех случаях в смежных классах результаты декодирования содержат одну «*T*» и остальные «*F*» (т. е. соответствуют случаям I и I*) или содержат некоторое количество «*T**» и остальные «*F**» (что соответствует случаям II и II*). Случая III ни для какого декодера получено не было. Также отметим, что во всех случаях рис. 2 гарантированно исправляются все лидеры, представляющие собой пакеты длиной до $b_{\max}$. Таким образом, можно сделать вывод, что для

Таблица 3. Результаты декодирования для выбранных кодов, %
Table 3. Decoding results for selected codes, %

Номер кода	Параметры декодирования	Выделенные классы				
		случай I, вариант 1 ($b \le b_{\max}$)	случай I, вариант 2 ($b > b_{\max}$)	случай I*	случай II	случай II*
2	A	28,13	44,73	13,09	9,96	4,10
	B	28,13	51,96	14,47	2,34	3,13
	C	28,13	46,68	8,59	9,77	6,84
	D	28,13	46,68	8,59	8,98	7,62
	E	28,13	47,66	—	24,23	—
	F	28,13	47,66	24,23	—	—
3	A	3,32	70,21	7,50	13,88	5,09
	B	3,32	77,78	11,03	4,76	3,11
	C	3,32	58,82	36,11	1,04	0,71
	D	3,32	58,87	36,11	0,43	1,26
	E	3,32	71,37	—	25,31	—
	F	3,32	91,12	—	5,55	—
4	A	7,42	54,1	21,58	5,96	10,97
	B	7,42	57,52	27,44	2,15	5,47
	C	7,42	49,61	29,79	3,42	9,77
	D	7,42	49,61	29,78	3,22	9,96
	E	7,42	66,02	—	26,56	—
	F	7,42	86,04	—	6,54	—
5	A	31,25	34,38	19,53	3,91	10,94
	B	31,25	36,72	28,12	1,56	2,34
	C	31,25	42,97	7,81	11,72	6,25
	D	31,25	42,97	7,81	10,16	7,81
	E	31,25	42,97	—	25,78	—
	F	31,25	62,5	—	6,25	—

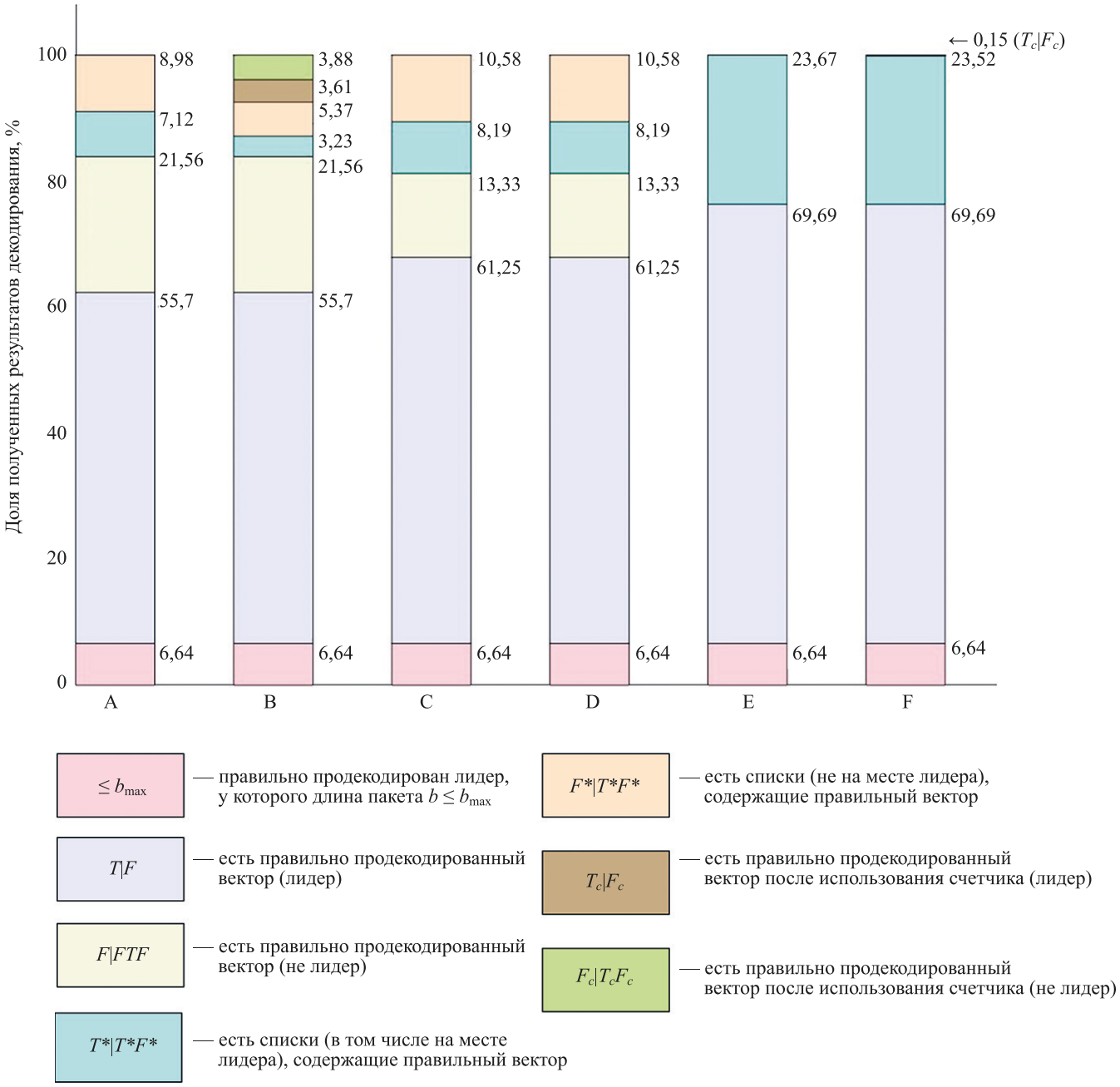


Рис. 2. Результаты декодирования по информационным совокупностям по минимальной длине пакета для блочно-перестановочного кода 1 (табл. 1) с использованием параметров декодирования из табл. 2

Fig. 2. Results of information set decoding by minimal burst length for block-permutation code 1 (Table 1) using decoding parameters from Table 2

всех рассмотренных декодеров обеспечивается декодирование в сфере по критерию минимальной длины пакета, а также вероятностное полное декодирование. Видно, что большая доля пакетов, имеющих длину сверх корректирующей способности, также могут быть однозначно исправлены.

В целом можно сделать вывод, что использование счетчика повышает долю однозначно декодируемых смежных классов для алгоритма с распределенными информационными совокупностями и практически не имеет эффекта при использовании плотных информационных совокупностей. При этом использование плотных информационных совокупностей имеет большую долю исправляемых лидеров смежных классов (еще большую при использовании плотных информационных совокупностей для всех расположений пакетов), но меньшую долю общего числа декодируемых смежных классов.

В табл. 3 представлены результаты декодирования для БП-кода и случайных кодов, номера и параметры которых представлены в табл. 1. Для БП-кода 2 (табл. 1) можно сделать выводы, аналогичные результатам для БП-кода 1 на рис. 2. Вместе с тем отметим, что при использовании плотных информационных совокупностей для каждого расположения пакета получено полное декодирование.

Рассмотрим теперь случайные коды. В табл. 3 представлены результаты декодирования для случайного кода 3 (табл. 1) с параметрами $n = 20$, $k = 7$ и $b_{\max} = 5$. По сравнению с БП-кодами, использование плотных информационных совокупностей снижает долю декодирований лидеров, однако существенно увеличивает количество декодируемых смежных классов, в которых исправляемый вектор не является лидером. Рассмотрим результаты для случайного кода 4 (табл. 1) с параметрами $n = 20$, $k = 10$, $b_{\max} = 3$ и результаты для случайного кода 5 $n = 20$, $k = 7$ с максимальной длиной исправляемого пакета $b_{\max} = 2$.

Подводя итог, можно сделать вывод, что для всех рассмотренных кодов все декодеры обеспечивают декодирование в рамках корректирующей способности. Также во всех случаях отсутствуют не декодируемые смежные классы, что с учетом используемой терминологии описано как вероятностное полное декодирование. Во всех случаях доля декодируемых векторов ошибки значительно превосходит корректирующую способность кода. Декодирование по критерию минимальной длины пакета лучше всего обеспечивается при использовании плотных информационных совокупностей для всех расположений пакета.

Заключение

В работе рассмотрено применение декодирования по информационным совокупностям для исправления одиночных пакетов ошибок. Введено понятие плотной информационной совокупности, которое может быть использовано как для сокращения их числа при декодировании в рамках корректирующей способности $b_{\max}$, так и для увеличения доли исправляемых пакетов сверх $b_{\max}$. Представлена модификация декодера с применением счетчика векторов ошибок, что позволяет уменьшить количество неоднозначных результатов декодирования. Приведена классификация декодирования, обобщающая понятие полного декодирования.

Выполнены эксперименты для кодов с небольшими параметрами: блочно-перестановочных кодов с малой плотностью проверок на четность, а также случайных линейных кодов, показывающие, что все рассмотренные декодеры обеспечивают для выбранных кодов декодирование в рамках корректирующей способности, а также исправление значительной доли векторов ошибок сверх корректирующей способности. Кроме этого, для рассмотренных кодов и декодеров не существует смежных классов, вероятность правильного декодирования в которых равнялась бы нулю, таким образом, данные декодеры обеспечивают ненулевую вероятность полного декодирования.

Возможными направлениями дальнейших исследований должен быть анализ предложенных алгоритмов декодирования для длинных кодов, где эффект памяти канала проявляется более выраженно, а также разработка и анализ методов декодирования для множественных пакетов и совместного исправления пакетирующихся и независимых ошибок. Необходимо принять во внимание, что это приведет к увеличению сложности декодирования.

Литература

1. Богатырев В.А., Богатырев С.В., Богатырев А.В. Оценка готовности компьютерной системы к своевременному обслуживанию запросов при его совмещении с информационным восстановлением памяти после отказов // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23. № 3. С. 608–617. https://doi.org/10.17586/2226-1494-2023-23-3-608-617
2. Moon T.K. Error Correction Coding: Mathematical Methods and Algorithms. Wiley, 2021. 992 p.
3. MacWilliams F.J., Sloane N.J.A. The Theory of Error-Correcting Codes. North-Holland Publishing Company, 1983. 762 p. (North-Holland Mathematical Library; Vol. 16).
4. Krouk E., Ovchinnikov A., Poikonen J. Channel models and reliable communication // Modulation and coding techniques in Wireless communications. Wiley, 2011. P. 1–20. https://doi.org/10.1002/9780470976777.ch1
5. Bogatyrev V.A., Bogatyrev A.V., Bogatyrev S.V. Multipath transmission of heterogeneous traffic in acceptable delays with packet replication and destruction of expired replicas in the nodes that make up the path // Communications in Computer and Information Science. 2023. V. 1748. P. 104–121. https://doi.org/10.1007/978-3-031-30648-8_9
6. Bogatyrev V., Bogatyrev S., Bogatyrev A. Timeliness of multipath redundant transmissions when all paths are not accessible for some request sources // Studies in Systems, Decision and Control. 2023. V. 457. P. 671–681. https://doi.org/10.1007/978-3-031-22938-1_46
7. Kulhandjian M., Kulhandjian H., D'Amours C. Improved soft decoding of Reed-Solomon codes on Gilbert-Elliott channels // Proc. of the IEEE International Symposium on Information Theory (ISIT). 2019. P. 1072–1076. https://doi.org/10.1109/ISIT.2019.8849456
8. Xie N., Zhang T., Haratsch E.F. Improving burst error tolerance of LDPC-centric coding systems in read channel // IEEE Transactions on Magnetics. 2010. V. 46. N 3. P. 933–941. https://doi.org/10.1109/TMAG.2009.2034012

References

1. Bogatyrev V.A., Bogatyrev S.V., Bogatyrev A.V. Assessment of the readiness of a computer system for timely servicing of requests when combined with information recovery of memory after failures. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2023, vol. 23, no. 3, pp. 608–617. (in Russian). https://doi.org/10.17586/2226-1494-2023-23-3-608-617
2. Moon T.K. *Error Correction Coding: Mathematical Methods and Algorithms*. Wiley, 2021, 992 p.
3. MacWilliams F.J., Sloane N.J.A. *The Theory of Error-Correcting Codes*. North-Holland Publishing Company, 1983, 762 p. North-Holland Mathematical Library, Vol. 16.
4. Krouk E., Ovchinnikov A., Poikonen J. Channel models and reliable communication. *Modulation and Coding Techniques in Wireless Communications*. Wiley, 2011, pp. 1–20. https://doi.org/10.1002/9780470976777.ch1
5. Bogatyrev V.A., Bogatyrev A.V., Bogatyrev S.V. Multipath transmission of heterogeneous traffic in acceptable delays with packet replication and destruction of expired replicas in the nodes that make up the path. *Communications in Computer and Information Science*, 2023, vol. 1748, pp. 104–121. https://doi.org/10.1007/978-3-031-30648-8_9
6. Bogatyrev V., Bogatyrev S., Bogatyrev A. Timeliness of multipath redundant transmissions when all paths are not accessible for some request sources. *Studies in Systems, Decision and Control*, 2023, vol. 457, pp. 671–681. https://doi.org/10.1007/978-3-031-22938-1_46
7. Kulhandjian M., Kulhandjian H., D'Amours C. Improved soft decoding of Reed-Solomon codes on Gilbert-Elliott channels. *Proc. of the IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1072–1076. https://doi.org/10.1109/ISIT.2019.8849456
8. Xie N., Zhang T., Haratsch E.F. Improving burst error tolerance of LDPC-centric coding systems in read channel. *IEEE Transactions on Magnetics*, 2010, vol. 46, no. 3, pp. 933–941. https://doi.org/10.1109/TMAG.2009.2034012

9. Yang M., Pan Z., Djordjevic I.B. FPGA-based burst-error performance analysis and optimization of regular and irregular SD-LDPC codes for 50G-PON and beyond // Optics Express. 2023. V. 31. N 6. P. 10936–10946. https://doi.org/10.1364/OE.477546
10. Xiao X., Vasić B., Lin S., Li J., Abdel-Ghaffar K. Quasi-cyclic LDPC codes with parity-check matrices of column weight two or more for correcting phased bursts of erasures // IEEE Transactions on Communications. 2021. V. 69. N 5. P. 2812–2823. https://doi.org/10.1109/TCOMM.2021.3059001
11. Eckford A., Kschischang F., Pasupathy S. Analysis of low-density parity-check codes for the gilbert–elliott channel // IEEE Transactions on Information Theory. 2005. V. 51. N 11. P. 3872–3889. https://doi.org/10.1109/TIT.2005.856934
12. Ovchinnikov A., Veresova A., Fominykh A. Usage of LDPC codes in a gilbert channel // Труды учебных заведений связи. 2022. Т. 8. № 4. С. 55–63. https://doi.org/10.31854/1813-324X-2022-8-4-55-63
13. Ghaddar N., Kim Y.-H., Milstein L.B., Ma L., Yi B.K. Joint channel estimation and coding over channels with memory using polar codes // IEEE Transactions on Communications. 2021. V. 69. N 10. P. 6575–6589. https://doi.org/10.1109/TCOMM.2021.3098822
14. Ovchinnikov A.A., Veresova A.M., Fominykh A.A. Decoding of linear codes for single error bursts correction based on the determination of certain events // Информационно-управляющие системы. 2022. № 6. С. 41–52. https://doi.org/10.31799/1684-8853-2022-6-41-52
15. Barg A., Krouk E., van Tilborg H.C.A. On the complexity of minimum distance decoding of long linear codes // IEEE Transactions on Information Theory. 1999. V. 45. N 5. P. 1392–1405. https://doi.org/10.1109/18.771141
16. Both L., May A. Optimizing BJMM with nearest neighbors: full decoding in 22/21n and McEliece security // WCC Workshop on Coding and Cryptography. 2017. V. 214.
17. Исаева М.Н. Поиск информационных совокупностей при исправлении пакетов ошибок квазициклическими кодами // T-Comm: Телекоммуникации и транспорт. 2023. Т. 17. № 7. С. 4–12. https://doi.org/10.36724/2072-8735-2023-17-7-4-12
18. Fossorier M.P.C. Quasicyclic low-density parity-check codes from circulant permutation matrices // IEEE Transactions on Information Theory. 2004. V. 50. N 8. P. 1788–1793. https://doi.org/10.1109/TIT.2004.831841
19. Xiao X., Vasić B., Lin S., Abdel-Ghaffar K., Ryan W.E. Reed-Solomon based quasi-cyclic LDPC codes: designs, girth, cycle structure, and reduction of short cycles // IEEE Transactions on Communications. 2019. V. 67. N 8. P. 5275–5286. https://doi.org/10.1109/TCOMM.2019.2916605
20. Крук Е.А., Овчинников А.А. Точная корректирующая способность кодов Гилберта при исправлении пакетов ошибок // Информационно-управляющие системы. 2016. № 1. С. 80–87. https://doi.org/10.15217/issn1684-8853.2016.1.80

9. Yang M., Pan Z., Djordjevic I.B. FPGA-based burst-error performance analysis and optimization of regular and irregular SD-LDPC codes for 50G-PON and beyond. *Optics Express*, 2023, vol. 31, no. 6, pp. 10936–10946. https://doi.org/10.1364/OE.477546
10. Xiao X., Vasić B., Lin S., Li J., Abdel-Ghaffar K. Quasi-cyclic LDPC codes with parity-check matrices of column weight two or more for correcting phased bursts of erasures. *IEEE Transactions on Communications*, 2021, vol. 69, no. 5, pp. 2812–2823. https://doi.org/10.1109/TCOMM.2021.3059001
11. Eckford A., Kschischang F., Pasupathy S. Analysis of low-density parity-check codes for the gilbert–elliott channel. *IEEE Transactions on Information Theory*, 2005, vol. 51, no. 11, pp. 3872–3889. https://doi.org/10.1109/TIT.2005.856934
12. Ovchinnikov A., Veresova A., Fominykh A. Usage of LDPC codes in a gilbert channel. *Proceedings of Telecommunication Universities*, 2022, vol. 8, no. 4, pp. 55–63. https://doi.org/10.31854/1813-324X-2022-8-4-55-63
13. Ghaddar N., Kim Y.-H., Milstein L.B., Ma L., Yi B.K. Joint channel estimation and coding over channels with memory using polar codes. *IEEE Transactions on Communications*, 2021, vol. 69, no. 10, pp. 6575–6589. https://doi.org/10.1109/TCOMM.2021.3098822
14. Ovchinnikov A.A., Veresova A.M., Fominykh A.A. Decoding of linear codes for single error bursts correction based on the determination of certain events. *Information and Control Systems*, 2022, no. 6, pp. 41–52. https://doi.org/10.31799/1684-8853-2022-6-41-52
15. Barg A., Krouk E., van Tilborg H.C.A. On the complexity of minimum distance decoding of long linear codes. *IEEE Transactions on Information Theory*, 1999, vol. 45, no. 5, pp. 1392–1405. https://doi.org/10.1109/18.771141
16. Both L., May A. Optimizing BJMM with nearest neighbors: full decoding in 22/21n and McEliece security. *WCC Workshop on Coding and Cryptography*, 2017, vol. 214.
17. Isaeva M.N. Finding information sets when correcting error bursts with quasi-cyclic codes. *T-Comm*, 2023, vol. 17, no. 7, pp. 4–12. (in Russian). https://doi.org/10.36724/2072-8735-2023-17-7-4-12
18. Fossorier M.P.C. Quasicyclic low-density parity-check codes from circulant permutation matrices. *IEEE Transactions on Information Theory*, 2004, vol. 50, no. 8, pp. 1788–1793. https://doi.org/10.1109/TIT.2004.831841
19. Xiao X., Vasić B., Lin S., Abdel-Ghaffar K., Ryan W.E. Reed-Solomon based quasi-cyclic LDPC codes: designs, girth, cycle structure, and reduction of short cycles. *IEEE Transactions on Communications*, 2019, vol. 67, no. 8, pp. 5275–5286. https://doi.org/10.1109/TCOMM.2019.2916605
20. Krouk E.A., Ovchinnikov A.A. Exact burst-correction capability of gilbert codes. *Information and Control Systems*, 2016, no. 1, pp. 80–87. (in Russian). https://doi.org/10.15217/issn1684-8853.2016.1.80

Авторы

Исаева Мария Николаевна — младший научный сотрудник, Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербург, 190008, Российская Федерация, sc 57243599200, https://orcid.org/0009-0007-6228-0617, misaeva@hse.ru

Овчинников Андрей Анатольевич — кандидат технических наук, доцент, ведущий научный сотрудник, Национальный исследовательский университет «Высшая школа экономики», Санкт-Петербург, 190008, Российская Федерация, sc 57207711162, https://orcid.org/0000-0002-8523-9429, a.ovchinnikov@hse.ru

Authors

Maria N. Isaeva — Junior Researcher, HSE University, Saint Petersburg, 190008, Russian Federation, sc 57243599200, https://orcid.org/0009-0007-6228-0617, misaeva@hse.ru

Andrei A. Ovchinnikov — PhD, Associate Professor, Leading Researcher, HSE University, Saint Petersburg, 190008, Russian Federation, sc 57207711162, https://orcid.org/0000-0002-8523-9429, a.ovchinnikov@hse.ru

Статья поступила в редакцию 23.10.2023
Одобрена после рецензирования 01.12.2023
Принята к печати 13.01.2024

Received 23.10.2023
Approved after reviewing 01.12.2023
Accepted 13.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-81-89

A novel strategic trajectory-based protocol for enhancing efficiency in wireless sensor networks

Rangaraj Gopalakrishnan[1]✉, Angamuthu Senthil Kumar[2]

[1] Periyar University, Salem, 636011, India
[2] Thiruvalluvar Government Arts College, Rasipuram, 637401, India

[1] gopaldharun1978@gmail.com✉, https://orcid.org/0000-0003-2456-265X
[2] senthilkumarmca76@gmail.com, https://orcid.org/0000-0001-5131-7428

Abstract

This research presents a comprehensive approach to enhance the efficiency and performance of Wireless Sensor Networks (WSNs) by addressing critical challenges, such as race conditions, reservation problems, and redundant data. A novel protocol combining Self-Adaptive Redundancy Elimination Clustering and Distributed Load Bandwidth Management is proposed to mitigate these challenges. The work intelligently extracts transmission hops and any-cast transmission features from diversity traffic information obtained through trace files, to eliminate nodes harboring redundant data. To optimize network organization, the number of clusters is dynamically adjusted according to the node density using the affinity propagation technique. Furthermore, load balancing is achieved by reallocating available bandwidth through bandwidth re-segmentation. The research also delves into the Proposed Network Infrastructure and Channel Coordination. The architecture encompasses cooperative clustering of nodes, strategic access point selection, data compression, and channel migration. By fostering collaboration among nodes within clusters, selecting access points judiciously, and employing efficient data compression techniques, the network overall efficiency is significantly improved. Channel migration strategies further bolster the network agility and responsiveness. The integration of Channel Sensing enriches the approach by collecting channel state information, enriched with spatial and temporal node information. This added insight empowers the network to make more informed decisions regarding channel allocation and coordination contributing to reduced interference and optimized data transmission. As a result of the work, the proposed methodology achieves remarkable results, including an average Packet Delivery Ratio of 99.1 % and an average reduction of packet loss by 4.3 % compared to existing studies. Additionally, the proposed protocol exhibits an average throughput improvement of 4.7 % and reduces average network delay to 52 milliseconds highlighting its significant contributions to the enhancement of WSN performance.

Keywords

wireless sensor networks, redundancy elimination clustering, security

For citation: Gopalakrishnan R., Sentil Kumar A. A novel strategic trajectory-based protocol for enhancing efficiency in wireless sensor networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 81–89. doi: 10.17586/2226-1494-2024-24-1-81-89

УДК 004.72

Новый стратегический траекторно-базированный протокол для повышения эффективности беспроводных сенсорных сетей

Рангарадж Гопалакришнан[1]✉, Ангамуту Сентил Кумар[2]

[1] Университет Перияр, Салем, 636011, Индия
[2] Государственный колледж искусств Тируваллувара, Расипурам, 637401, Индия

[1] gopaldharun1978@gmail.com✉, https://orcid.org/0000-0003-2456-265X
[2] senthilkumarmca76@gmail.com, https://orcid.org/0000-0001-5131-7428

Аннотация

Исследован комплексный подход к повышению эффективности и производительности беспроводных сенсорных сетей путем решения критических проблем, таких как условия гонки, проблемы резервирования и избыточные

© Gopalakrishnan R., Sentil Kumar A., 2024

данные. Для решения этих проблем предложен новый протокол, сочетающий в себе самоадаптирующуюся кластеризацию с устранением избыточности и управление полосой пропускания распределенной нагрузки. Используется разумный способ для извлечения функции передачи сообщений с «перескоком» через узлы (их пропуском) и передачи сообщений любому из ближайших узлов. Для этого используется информация о разнесенном трафике, получаемая на основе файлов трассировки, исключая узлы с избыточными данными. Количество кластеров динамически регулируется в зависимости от плотности узлов, используя метод распространения сходства. Балансировка нагрузки достигается посредством перераспределения доступной полосы пропускания и повторной сегментации полосы пропускания. В исследовании представлена предлагаемая сетевая инфраструктура и координация каналов. Архитектура включает в себя совместную кластеризацию узлов, выбор точки доступа, сжатие данных и миграцию каналов. Эффективность сети значительно повышается за счет взаимодействия между узлами внутри кластеров, разумного выбора точек доступа и использования эффективных методов сжатия данных. Стратегия миграции каналов еще больше повышает гибкость и оперативность сети. Интеграция способности обнаружения канала обогащает подход за счет сбора информации о состоянии канала и дополнения пространственной и временной информации об узлах. Эта дополнительная информация позволяет сети принимать более обоснованные решения относительно распределения и координации каналов, способствуя снижению помех и оптимизации передачи данных. Предложенная методология позволила получить средний коэффициент доставки пакетов (PDR) равный 99,1 % и среднее снижение потерь пакетов на 4,3 % по сравнению с существующими исследованиями. Новый протокол продемонстрировал повышение средней пропускной способности на 4,7 % и снижение средней задержки сети до 52 мс, что подчеркивает его значительный вклад в увеличение производительности WSN.

Ключевые слова

беспроводные сенсорные сети, кластеризация с устранением избыточности, безопасность

Ссылка для цитирования: Гопалакришнан Р., Сентил Кумар А. Новый стратегический траекторно-базированный протокол для повышения эффективности беспроводных сенсорных сетей // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 81–89 (на англ. яз.). doi: 10.17586/2226-1494-2024-24-1-81-89

Introduction

Wireless Sensor Networks (WSNs) have emerged as a vital technology for monitoring and collecting data from diverse environments. However, the effective operation of WSNs faces significant challenges, including race conditions, reservation problems, and redundant data storage. These challenges can lead to unpredictable network behavior, inefficient resource utilization, and increased communication overhead [1–4]. To overcome these challenges, this research introduces a novel approach that combines Self-Adaptive Redundancy Elimination Clustering and Distributed Load Bandwidth Management. The objective is to enhance the overall efficiency, reliability, and scalability of WSNs, ultimately improving their utility in various applications, such as environmental monitoring, healthcare, and industrial automation. The primary focus of this research is to devise a solution that dynamically adapts to changing network conditions and optimizes resource utilization. By extracting transmission hops and any-cast transmission features from diversity traffic information sourced from trace files, the proposed approach aims to identify and eliminate nodes that carry redundant data. This reduction in redundant information not only conserves network resources but also reduces data transmission latency. A critical aspect of the proposed solution is its ability to automatically adjust the number of clusters based on the node density. Leveraging the affinity propagation technique, the network self-organizes into clusters, aligning with the density of nodes in specific regions. Additionally, the research tackles the challenge of load balancing by redistributing available bandwidth through a process called bandwidth re-segmentation. This adaptive approach ensures that no single node or cluster becomes a bottleneck enhancing the network responsiveness and resource utilization [5–9]. The architecture of WSN is shown in Fig. 1.

The research further explores the proposed network infrastructure, which encompasses cooperative clustering of nodes, strategic access point selection, data compression, and channel migration. These components collectively contribute to the robustness of the network architecture. By optimizing the way nodes collaborate within clusters, strategically placing access points, employing efficient data compression algorithms, and intelligently migrating channels, the network achieves improved data throughput and reduced congestion.

Furthermore, the inclusion of Channel Sensing enriches the network understanding of the wireless channel environment. By collecting channel state information along with spatial and temporal node characteristics, the network

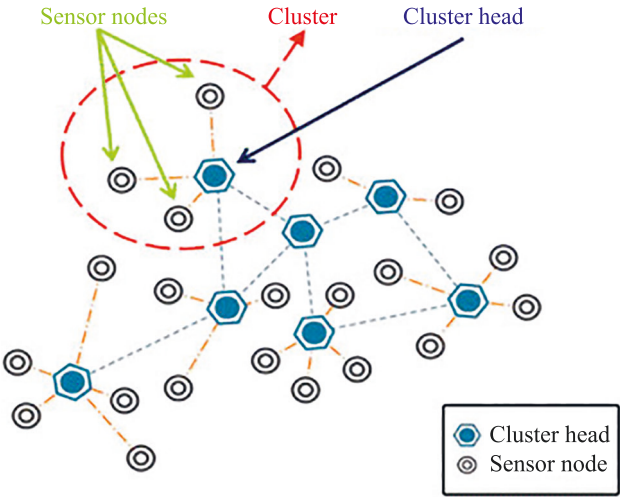


Fig. 1. Architecture of Cluster based WSN

gains valuable insights into the availability and quality of channels. This information is leveraged to make informed decisions about channel allocation and coordination, thereby minimizing interference and maximizing data transmission efficiency. In summary, this research proposes a holistic solution to address critical challenges in WSNs. By synergistically integrating self-adaptive redundancy elimination clustering, distributed load bandwidth management, and sophisticated channel sensing, the proposed approach aims to propel the capabilities of WSNs to new heights. The subsequent sections of this paper delve into the technical details of the proposed methodology, experimental setup, results, and comparisons with existing solutions, thereby validating the efficacy of the approach in improving the performance of WSNs in real-world scenarios. The subsequent sections of this paper delve into the technical intricacies of the proposed methodology, spanning route setup, route discovery, clustering of nodes, access point selection, and redundancy elimination strategies. We then delve into the Cooperative Clustering mechanism and the integration of channel sensing for enhanced trace file generation. Additionally, we present an innovative approach to Access Point Selection and discuss the intricacies of the Redundancy Elimination and Outage Interference Cancellation Model.

Related work

WSNs have gained significant attention due to their wide range of applications. Among the key challenges in the context of WSNs, identifying network layer attacks is a critical concern. The paper [10] delves into the problem of identifying network layer attacks in WSNs with the aim of developing intrusion detection systems. In their work, the authors outline 15 types of behavior, which include 14 states under attack and the normal state, providing a comprehensive framework for characterizing network behavior in WSNs. Additionally, they specify identification features crucial for the detection of network layer attacks. A notable contribution of the paper is the proposed technique for identifying attacks, which combines the Random Forest algorithm and a probability classifier. This innovative approach offers a promising avenue for accurately detecting and classifying network layer attacks in WSNs. Furthermore, the paper discusses a detailed study of the relationship between the accuracy of the probability classifier and the average number of features used. This analysis allows for the reduction of the feature space to a single feature optimizing the detection process. The paper highlights the relative constancy of the dependence values for most attacks in the range of 0 to 0.7 ± 0.15, followed by a decrease indicating the quality of identification. These findings offer valuable insights into the behavior and identification of network layer attacks in WSNs. Race conditions and reservation problems have been acknowledged as major hurdles in WSNs. Alleviating these challenges requires efficient synchronization and resource allocation mechanisms. Previous works in [10] proposed a distributed reservation protocol that mitigates contention and collision issues enhancing the predictability of network behavior. Similarly, [11] introduced a time-slot-based coordination scheme that effectively addresses race conditions ensuring synchronized access to shared resources. Redundancy elimination in WSNs has been a focus to optimize network performance. Research by [12] proposed a data fusion approach to identify and eliminate redundant data in multi-hop WSNs. Their method reduced communication overhead and conserved energy by aggregating similar data packets. In another study [13], a clustering-based technique is presented that removed duplicate data at the cluster head, enhancing overall data integrity and network efficiency. Load bandwidth management plays a vital role in maintaining balanced data traffic [14] devised a load-aware routing algorithm that dynamically adjusts routes based on network load preventing congestion and maximizing throughput. Additionally, [15] introduced a bandwidth reservation scheme that optimizes channel utilization in heavily loaded networks by preemptively allocating resources to critical tasks [16–18]. While prior research has provided valuable insights into addressing individual challenges, there is a gap in the literature concerning the comprehensive integration of redundancy elimination, load balancing, and adaptive clustering in WSNs. This research aims to bridge this gap by proposing a holistic approach that combines these techniques to enhance network performance [19–22]. In the following sections, we present a novel protocol that leverages the strengths of affinity propagation for self-adaptive clustering, dynamic bandwidth re-segmentation for load balancing, and channel sensing for improved channel allocation. This integrated approach builds upon the foundations laid by previous works and strives to overcome limitations inherent to individual solutions. The comparison of the existing works can be found in Table 1.

In response to the issues identified in the literature review, our proposed framework introduces innovative solutions aimed at elevating the efficiency and performance of WSNs. Despite the valuable contributions of previous studies across various WSN aspects, challenges associated with mitigating race conditions, reservation problems, and redundant data persisted. Our framework, which seamlessly integrates Self-Adaptive Redundancy Elimination Clustering and Distributed Load Bandwidth Management, takes a comprehensive approach to address these challenges. The unique amalgamation of these components exhibits intelligent extraction of transmission hops and anycast transmission features from diverse traffic information, as validated through trace file analysis. This intelligent process effectively eliminates nodes harboring redundant data, marking a significant advancement in overcoming persistent challenges observed in earlier research.

Methodology

The methodology of this research encompasses several interconnected phases that collectively constitute the proposed protocol for Self-Adaptive Redundancy Elimination Clustering, Distributed Load Bandwidth Management, and Cooperative Clustering. Each phase aims to address specific challenges and contribute to the overall enhancement of WSNs.

Table 1. Comparison of current solutions

Reference	Method	Main Contributions	Focus Areas
[10]	Random Forest, Probability Classifier	– Identifies 15 types of network layer behaviors including 14 attack states and normal state. – Proposes a technique combining Random Forest and probability classifier for attack identification. – Analyzes the relationship between accuracy and feature count enabling feature reduction to optimize detection. – Reveals relative constancy of dependence values for most attacks within a certain range followed by a decrease corresponding to identification quality.	Network layer attack identification, Behavior analysis in WSNs
[11]	TTDD	Proposes a two-tier data dissemination approach in large-scale WSNs. Provides efficient and reliable data delivery in WSNs with improved scalability and reduced energy consumption	Data dissemination, scalability, energy efficiency
[12]	RE-based cluster-head selection	Introduces a cluster-head selection algorithm based on residual energy levels. Enhances network lifetime and efficiency for Internet of Things (IoT) applications	Cluster-head selection, IoT applications
[13]	QWRP	Presents QWRP, a query-driven virtual wheel-based routing protocol improving data transmission efficiency in WSNs with mobile sinks	Routing protocol, mobile sinks
[14]	LBRR	Proposes LBRR, a load-balanced ring routing protocol for heterogeneous sensor networks. Balances traffic and improves network performance	Routing in heterogeneous WSNs
[15]	Ring routing	Introduces an energy-efficient routing protocol for WSNs with mobile sinks. Enhances energy conservation and prolongs network lifetime	Routing with mobile sinks, energy efficiency
[16]	Vertical and horizontal segregation data protocol	Presents a data dissemination protocol based on vertical and horizontal segregation. Improves data delivery efficiency in WSNs	Data dissemination, efficiency
[17]	Hilbert-chain topology for energy conservation	Proposes the use of Hilbert-chain topology to conserve energy in large-scale WSNs. Enhances energy efficiency and network longevity	Topology for energy conservation
[18]	Optimal location for mobile sink in WSNs	Investigates the optimal placement of mobile sinks in WSNs for efficient data collection. Enhances data gathering efficiency	Mobile sink placement, data collection
[19]	Efficient data gathering with mobile collectors	Introduces efficient data gathering techniques using mobile collectors. Optimizes data collection in WSNs	Data gathering with mobile collectors
[20]	Adaptive data collection strategies in WSNs	Discusses adaptive data collection strategies to prolong the lifetime of WSNs with constrained resources	Data collection strategies in WSNs
[21]	Trust-based on-demand multipath routing in MANET	Proposes a trust-based on-demand multipath routing protocol for Mobile Ad-Hoc Networks (MANETs) to enhance security and reliability	Trust-based routing in MANET
[22]	Efficient Trusted Secure Ad-Hoc On-Demand Multipath	Presents an efficient and trusted secure on-demand multipath routing approach for MANETs, focusing on security and reliability	Secure multipath routing in MANET

Route Setup and Route Discovery

In the Route Setup phase, the process of establishing communication routes within the network is initiated. Sensor nodes play a pivotal role in forwarding packets to their intended destinations. To achieve efficient routing, each node gathers crucial metadata about its neighboring nodes, particularly focusing on the transmission cost (*tc*) associated with these connections.

The *tc* value represents the number of transmissions required for a packet to be successfully received at the destination. This value encapsulates the probabilistic aspects of communication. The probability of forward (p) and backward (q) packet reception over a link forms the basis for *tc* calculation:

$$tc = \frac{1}{p \times q},$$

where $p \times q$ represents the multiplication of the probabilities of forward (p) and backward (q) packet reception over a link.

Nodes then store this transmission cost information contributing to the dynamic routing decision process.

In the Route Discovery phase, a source node that intends to transmit data initiates the process to identify multiple potential paths to the destination. This phase is crucial for optimizing data transmission reliability and efficiency. The proposed multipath routing protocol capitalizes on reliability metrics, such as transmission cost, optimal traffic ratio, and remaining energy.

A Route Request (RR) packet is generated by the source node and disseminated throughout the network. When an intermediate node receives an RR packet, it evaluates the transmission cost, optimal traffic ratio, and remaining energy associated with different paths from the source to the destination. The goal is to identify an optimal route that balances these factors. The node then forwards the RR packet to its neighboring node with the minimum computed cost contributing to route discovery.

Clustering of Nodes

Recognizing the significance of clustering in managing the impact of sink mobility and network scalability, the Clustering of Nodes phase aims to create a hierarchical structure within the network [23, 24]. Clustering allows for efficient management of network overhead in large-scale scenarios. The selection of cluster heads plays a critical role in this process, as they become coordinators for their respective clusters. Clustering Based on Location Similarity. In this approach, nodes are clustered based on their spatial proximity. This mechanism ensures that clusters are composed of direct neighbors. The goal is to mitigate the impact of node mobility on the network structure by containing topological changes within each cluster. Clustering Based on Node Density Similarity: nodes are grouped based on their node density, which is determined by factors such as node degree, speed, and battery power. The clustering process aims to identify clusters with similar densities. Cluster head selection within this approach is influenced by criteria such as high node degree and node density.

Cooperative Clustering and Access Point Selection

Cooperative Clustering constitutes the generation of multiple cluster heads to enhance scalability and resource utilization. Cluster heads serve as access points for channel allocation and data dissemination. This approach aims to optimize spatial resource reuse, reducing interference and optimizing channel efficiency. The Access Point Selection mechanism employs game-theoretic principles to aggregate channels for data transmission. Utility functions are designed considering Quality of Service (QoS) requirements and energy consumption. Channels with varying widths are allocated to access points, with a focus on non-overlapping channels for efficient channel allocation.

Redundancy Elimination and Outage Interference Cancellation

The Redundancy Elimination and Outage Interference Cancellation phase addresses interference and redundancy through advanced scheduling techniques. These techniques encompass CDMA-based systems, multi-carrier frequency-division multiple access, and time division multiple access. The central objective is to optimize packet delay jitter, a key factor in ensuring efficient packet scheduling and resource allocation. This phase relies on network information and employs heuristics to eliminate interference and redundancy ultimately enhancing overall network stability and ensuring the QoS levels.

Results and discussion

In this chapter, we present a comprehensive experimental setup and discuss the evaluation results of the proposed methodology. The performance of the various components and strategies described in earlier sections is assessed through simulations conducted in a controlled environment. We describe the simulation platform, evaluation metrics, experimental scenarios, and provide a detailed analysis of the results.

Simulation Environment

To conduct a rigorous evaluation of the proposed methodology, we utilized the NS2 (Network Simulator 2) platform. NS2 is a widely accepted network simulation tool that allows us to model wireless communication scenarios accurately. It provides a versatile environment for simulating MANETs and assessing protocol performance in a controlled and reproducible manner. Table 2 shows the parameters used in the work.

Evaluation Metrics

We employed a set of well-defined evaluation metrics to measure different aspects of the proposed methodology performance:

Throughput: measures the rate of successful data transmission from source to destination reflecting the network capacity to deliver data efficiently.

Packet Delivery Ratio (PDR): represents the ratio of successfully received data packets to the total number of packets generated indicating the reliability of data delivery.

Routing Overhead: measures the additional data packets generated for routing and control purposes reflecting the efficiency of the routing mechanisms.

End-to-End Delay: quantifies the time taken for data packets to travel from the source to the destination providing insights into network latency.

Packet Loss: indicates the percentage of data packets that were not successfully delivered to their intended destinations.

Energy Utilization: assesses the energy consumed by network nodes during data transmission, highlighting the energy efficiency of the proposed strategies.

Results and Discussion

We conducted extensive simulations for each experimental scenario and analyzed the results comprehensively. The performance metrics were plotted on graphs, numerical values were tabulated, and comparisons were drawn between the proposed methodology and existing solutions. In Fig. 2, we present the throughput performance of the proposed methodology under varying node densities.

Table 2. Simulation parameters

Simulation Parameter	Value
Simulator	NS2
Topology Size, m	1000 × 1000
Number of Nodes	200
Bandwidth of the Network, Mbps	2
Traffic type	CBR
Pause Time, s	10, 20
Data Packet size, bytes	512
Buffer size, packets	30
Simulation Time, min	30

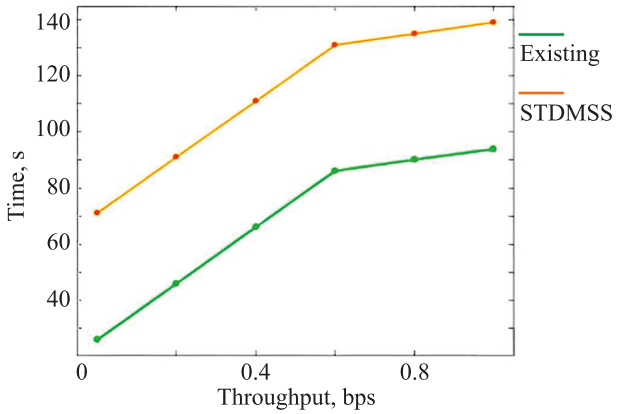


Fig. 2. Throughput analysis

Packet Delivery Ratio

PDR is a crucial metric that indicates the reliability of a network. It quantifies the proportion of data packets successfully received by their intended destinations compared to the total number of packets generated by the source node. In WSNs, a high PDR signifies the network ability to deliver data accurately and efficiently.

A high PDR is desirable as it reflects a network ability to maintain reliable communication under various conditions. It ensures that data generated by source nodes actually reaches the intended recipients without being lost or corrupted during transmission. This metric is particularly important in applications where data accuracy and integrity are paramount, such as environmental monitoring or critical event detection.

Packet Loss

Packet loss is an undesirable event that occurs when data packets do not reach their intended destinations due to factors like collisions, interference, or node failures. These issues can lead to data corruption and hinder the successful transmission of critical information. Packet loss impacts the overall network performance and may result in incomplete or inaccurate data collection. The proposed algorithm aims to mitigate packet loss by employing early detection mechanisms. When the algorithm identifies a lost packet, it takes preventive measures to halt further transmission from the affected node. By doing so, the algorithm prevents the redundant transmission of data that may further contribute to network congestion and resource depletion. Addressing packet loss is crucial for maintaining the reliability of data communication in WSNs. Detecting and handling packet loss promptly helps ensure that the network operates efficiently and effectively, meeting the requirements of various applications.

Delay

Delay, in the context of WSNs, refers to the time it takes for data packets to travel from the source node to the destination node. Delay is influenced by factors, such as network congestion, the number of hops a packet traverses, and the overall traffic load. Excessive delay can lead to delays in data collection and transmission impacting the real-time nature of applications that depend on timely information. Reducing delay is crucial, especially in applications like industrial monitoring or emergency response systems, where rapid data exchange is essential. By optimizing routing paths, managing network traffic, and minimizing the number of hops, delay can be minimized enhancing the network efficiency and responsiveness.

Performance Analysis of Strategic Trajectory for Throughput Maximization

The proposed protocol, based on a strategic trajectory for throughput maximization, focuses on optimizing network performance by improving channel coordination among cluster-based nodes. This strategy enhances data transmission efficiency and throughput, enabling the network to handle more data traffic while maintaining reliability.

In Fig. 3, the analysis of PDR against data traffic is depicted. Higher node mobility leads to increased PDRs. However, it's important to note that this also introduces challenges such as higher path maintenance overhead and network delay. Achieving a balance between node mobility and network reliability is a key consideration in optimizing packet delivery performance.

Fig. 4 illustrates the comparison between the proposed protocol and existing ones in terms of network overhead. The proposed protocol outperforms existing solutions due to its effective packet scheduling algorithm and bandwidth compression constraints. Lower overhead translates to better network efficiency and reduced resource consumption.

Routing latency, as depicted in Fig. 5, is impacted by network traffic and the efficiency of channel allocation. The proposed protocol effectively handles data transfer and

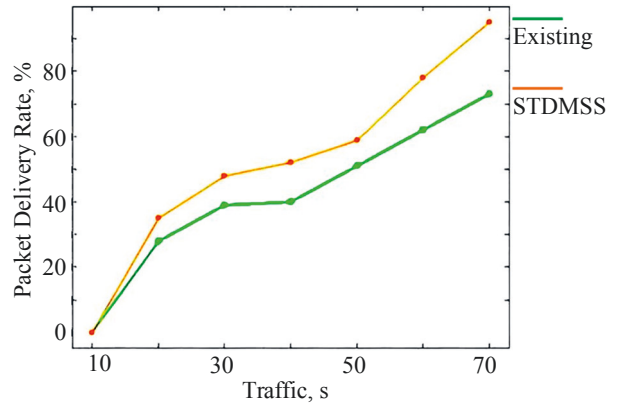


Fig. 3. Packet Delivery Ratio Analysis against Data Traffic

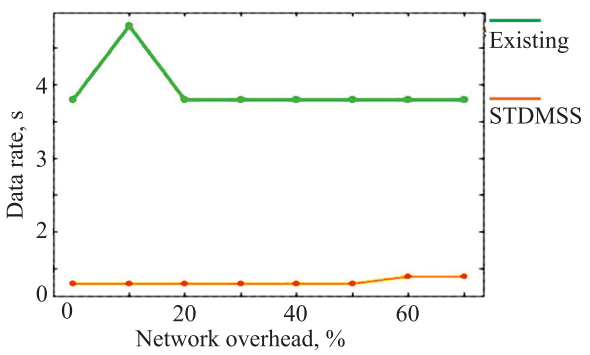


Fig. 4. Overhead Analysis against Data Traffic

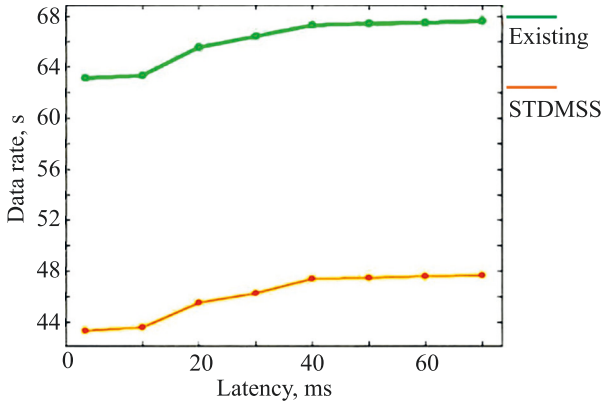


Fig. 5. Performance Analysis of Routing Latency of STDMSS Model

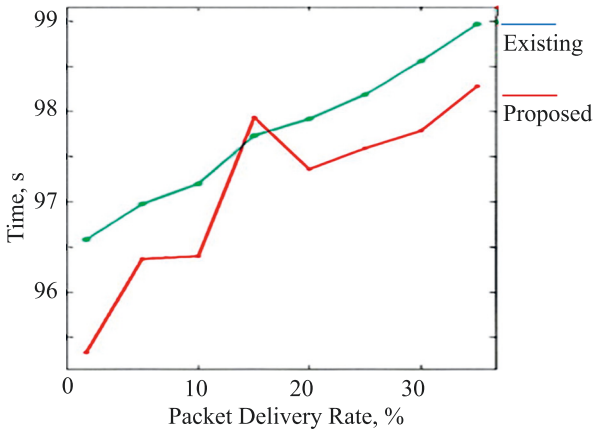


Fig. 7. Performance Analysis of Packet Delivery Ratio of Multiple Mobile Sink

routing complexities, resulting in minimal latency. This is essential for maintaining timely data communication and responsiveness.

Performance Analysis of Protocols with Multiple Energy Efficient Routing

This section focuses on the performance analysis of various energy-efficient routing protocols in the context of WSNs. The aim is to assess the effectiveness of these protocols in terms of critical performance metrics, such as throughput, PDR, network overhead, and packet loss.

The evaluation compares the proposed scheme against existing schemes in terms of PDR and energy utilization. The results demonstrate that the proposed scheme exhibits better energy efficiency and PDR compared to the existing scheme.

Fig. 6 showcases the performance analysis of energy utilization in multiple mobile sink scheduling. The proposed scheme effectively addresses the energy hole problem resulting in a gradual increase in residual energy when compared to the existing scheme.

Fig. 7 displays the performance analysis of PDR in multiple mobile sink scheduling. The proposed scheme achieves a higher PDR due to improved connectivity among nodes, contributing to effective coverage and communication performance.

The proposed clustering algorithm achieves a higher PDR while utilizing significantly less energy compared to the existing approach. The scalability and reliability of the proposed protocol are highlighted in the analysis. The network lifetime and effectiveness of the proposed scheme in optimizing system energy while ensuring communication performance are evident from the results.

Performance Analysis on Self Adaptive Redundancy Elimination

This subsection introduces the concept of Self Adaptive Redundancy Elimination Clustering and Distributed Load Bandwidth Management Constraints. The protocol emphasizes load balancing and redundancy elimination in the network. It also utilizes a routing table containing trace information to handle the distribution of data within the network effectively.

Fig. 8 illustrates the results of the throughput analysis for the Self Adaptive Redundancy Elimination Clustering protocol. The protocol exhibits superior throughput performance compared to existing Quality of Data (QoD) protocols. This improvement is attributed to effective data distribution and load balancing strategies.

Fig. 9 presents the overhead analysis of the Self Adaptive Redundancy Elimination Clustering protocol. The graph demonstrates that the protocol effectively manages overhead by optimizing the transmission of control packets across multiple hops. This efficient overhead management contributes to improved network performance.

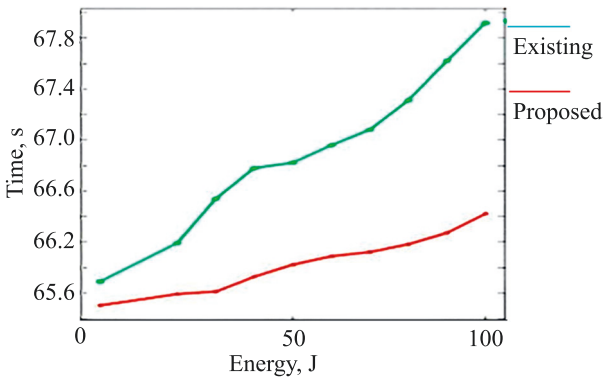


Fig. 6. Performance of Energy Utilization of Multiple Mobile Sink Scheduling

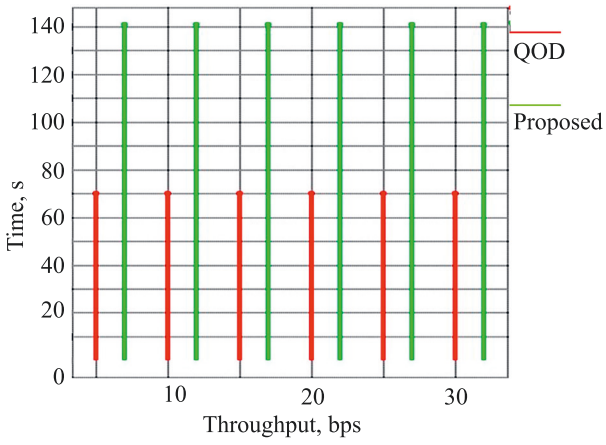


Fig. 8. Throughput Analysis of Self Adaptive Redundancy Elimination Clustering Protocol

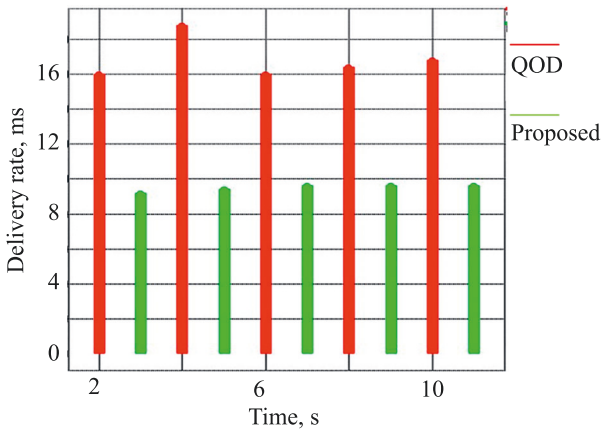


Fig. 9. Overhead Analysis of Self Adaptive Redundancy Elimination Clustering Protocol

The proposed protocol demonstrates better throughput, lower overhead, higher PDR, and reduced packet loss, highlighting its efficiency and effectiveness.

Conclusion

In conclusion, the comprehensive evaluation conducted within the simulated environment has yielded valuable insights into the performance and efficacy of the proposed protocol in a Mobile Ad-Hoc Network. The simulation parameters were thoughtfully selected to closely emulate real-world scenarios, lending credibility to the obtained outcomes. The metrics scrutinized throughout the study collectively offer a comprehensive understanding of the protocol capabilities. Notably, the Packet Delivery Ratio metric demonstrates the protocol proficiency in reliably transmitting data packets. Swift identification of Packet Loss due to collisions or node failures showcases the protocol resilience. Moreover, the analysis of Delay sheds light on latency concerns arising from network congestion and packet traversal.

The Packet Delivery Ratio reveals a notable improvement, registering a delta of 6 % between the initial and final measurements, highlighting the protocol effectiveness in ensuring consistent data packet transmission. The strategic trajectory analysis emphasizes the protocol pivotal function in enhancing channel coordination resulting in a substantial 30 % increase in throughput. Additionally, the evaluation of Multiple Energy Efficient Routing Protocols affirms the dominance of the proposed clustering algorithm, showcasing advancements in both energy efficiency and packet delivery ratio, with a delta of 25 % from the initial values. In summary, the comprehensive performance analysis robustly validates the protocol promise in elevating network dependability, energy efficiency, throughput, and resource management within wireless sensor networks. As an innovative solution, the protocol holds substantial potential to drive network optimization and elevate communication effectiveness.

References

1. Akyildiz F., Su W., Sankarasubramaniam Y., Cayirci E. A survey on sensor networks. *IEEE Communication Magazine*, 2002, vol. 40, no. 8, pp. 102–114. https://doi.org/10.1109/MCOM.2002.1024422
2. Cheng W.C., Chou C., Golubchik L., Khuller S., Wan Y.C. A coordinated data collection approach: design, evaluation, and comparison. *IEEE Journal on Selected Areas in Communication*, 2004, vol. 22, no. 10, pp. 2004–2018. https://doi.org/10.1109/JSAC.2004.836009
3. Xu K., Hassanein H., Takahara G., Wang Q. Relay node deployment strategies in heterogeneous wireless sensor networks. *IEEE Transactions on Mobile Computing*, 2010, vol. 9, no. 2, pp. 145–159. https://doi.org/10.1109/TMC.2009.105
4. Chandrasekaran V., Shanmugam A. A review on hierarchical cluster based routing in wireless sensor networks. *Journal Global Research in Computer Science*, 2012, vol. 3, no. 2, pp. 12–16.
5. Hou Y., Shi Y., Sherali H., Midkiff S. On energy provisioning and relay node placement for wireless sensor networks. *IEEE Transaction on Wireless Communication*, 2005, vol. 4, no. 5, pp. 2579–2590. https://doi.org/10.1109/twc.2005.853969
6. Sanjana S., Shavanthi L., Bhagya R. Analysis of energy aware sleep scheduling routing protocol (EASSR) in wireless sensor networks. *Proc. of the International Conference on Intelligent Computing and Control (I2C2)*, 2017, pp. 1–6. https://doi.org/10.1109/I2C2.2017.8321920
7. Mahdi O.A., Wahab A.W.A., Idris M.Y.I., Znaid A.A., Al-Mayouf Y.R.B., Khan S. WDARS: A weighted data aggregation routing strategy with minimum link cost in event-driven WSNs. *Journal of Sensors*, 2016, vol. 2016, pp. 3428730. https://doi.org/10.1155/2016/3428730
8. Zahedi A., Arghavani M., Parandin F., Arghavani A. Energy efficient reservation-based cluster head selection in WSNs. *Wireless Personal Communication*, 2018, vol. 100, no. 3, pp. 667–679. https://doi.org/10.1007/s11277-017-5189-9
9. Rhee I., Warrier A., Min J., Xu L. DRAND: Distributed randomized TDMA scheduling for wireless ad-hoc networks. *MobiHoc '06: Proc.*

Литература

1. Akyildiz F., Su W., Sankarasubramaniam Y., Cayirci E. A survey on sensor networks // IEEE Communication Magazine. 2002. V. 40. N 8. P. 102–114. https://doi.org/10.1109/MCOM.2002.1024422
2. Cheng W.C., Chou C., Golubchik L., Khuller S., Wan Y.C. A coordinated data collection approach: design, evaluation, and comparison // IEEE Journal on Selected Areas in Communication. 2004. V. 22. N 10. P. 2004–2018. https://doi.org/10.1109/JSAC.2004.836009
3. Xu K., Hassanein H., Takahara G., Wang Q. Relay node deployment strategies in heterogeneous wireless sensor networks // IEEE Transactions on Mobile Computing. 2010. V. 9. N 2. P. 145–159. https://doi.org/10.1109/TMC.2009.105
4. Chandrasekaran V., Shanmugam A. A review on hierarchical cluster based routing in wireless sensor networks // Journal Global Research in Computer Science. 2012. V. 3. N 2. P. 12–16.
5. Hou Y., Shi Y., Sherali H., Midkiff S. On energy provisioning and relay node placement for wireless sensor networks // IEEE Transaction on Wireless Communication. 2005. V. 4. N 5. P. 2579–2590. https://doi.org/10.1109/twc.2005.853969
6. Sanjana S., Shavanthi L., Bhagya R. Analysis of energy aware sleep scheduling routing protocol (EASSR) in wireless sensor networks // Proc. of the International Conference on Intelligent Computing and Control (I2C2). 2017. P. 1–6. https://doi.org/10.1109/I2C2.2017.8321920
7. Mahdi O.A., Wahab A.W.A., Idris M.Y.I., Znaid A.A., Al-Mayouf Y.R.B., Khan S. WDARS: A weighted data aggregation routing strategy with minimum link cost in event-driven WSNs // Journal of Sensors. 2016. V. 2016. P. 3428730. https://doi.org/10.1155/2016/3428730
8. Zahedi A., Arghavani M., Parandin F., Arghavani A. Energy efficient reservation-based cluster head selection in WSNs // Wireless Personal Communication. 2018. V. 100. N 3. P. 667–679. https://doi.org/10.1007/s11277-017-5189-9
9. Rhee I., Warrier A., Min J., Xu L. DRAND: Distributed randomized TDMA scheduling for wireless ad-hoc networks // MobiHoc '06:

of the 7th ACM International Symposium on Mobile ad hoc Networking and Computing, 2006, pp. 190–201. https://doi.org/10.1145/1132905.1132927
10. Korzhuk V., Groznykh A., Menshikov A., Strecker M. Identification of attacks against wireless sensor networks based on behaviour analysis. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, 2019, vol. 10, no. 2, pp. 1–21. https://doi.org/10.22667/JOWUA.2019.06.30.001
11. Tunca C., Isik S., Donmez M., Ersoy C. Distributed mobile sink routing for wireless sensor networks: A survey. *IEEE Communications Surveys & Tutorials*, 2014, vol. 16, no. 2, pp. 877–897. https://doi.org/10.1109/surv.2013.100113.00293
12. Luo H., Ye F., Cheng J., Lu S., Zhang L. TTDD: Two-tier data dissemination in large-scale wireless sensor networks. *Wireless Networks*, 2005, vol. 11, no. 1-2, pp. 161–175. https://doi.org/10.1007/s11276-004-4753-x
13. Behera T.M., Mohapatra S.K., Samal U.C., Khan M.S., Daneshmand M., Gandomi A.H. Residual energy-based cluster-head selection in wsns for iot application. *IEEE Internet of Things Journal*, 2019, vol. 6, no. 3, pp. 5132–5139. https://doi.org/10.1109/jiot.2019.2897119
14. Jain S., Pattanaik K., Shukla A. QWRP: Query-driven virtual wheel based routing protocol for wireless sensor networks with mobile sink. *Journal of Network and Computer Applications*, 2019, vol. 147, pp. 102430. https://doi.org/10.1016/j.jnca.2019.102430
15. Maurya S., Gupta V., Jain V.K. LBRR: Load balanced ring routing protocol for heterogeneous sensor networks with sink mobility. *Proc. of the 2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1–6. https://doi.org/10.1109/wcnc.2017.7925728
16. Tunca C., Isik S., Donmez M.Y., Ersoy C. Ring routing: An energy-efficient routing protocol for wireless sensor networks with a mobile sink. *IEEE Transactions on Mobile Computing*, 2015, vol. 14, no. 9, pp. 1947–1960. https://doi.org/10.1109/TMC.2014.2366776
17. Jain S., Sharma S., Bagga N. A vertical and horizontal segregation based data dissemination protocol. *Emerging Research in Computing, Information, Communication and Applications*, Springer, 2016, pp. 401–412. https://doi.org/10.1007/978-81-322-2553-9_37
18. Lin Y.C., Zhong J.-H. Hilbert-chain topology for energy conservation in large-scale wireless sensor networks. *Proc. of the 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing (UIC/ATC)*, 2012, pp. 225–232. https://doi.org/10.1109/uic-atc.2012.37
19. Khodashahi M.H., Tashtarian F., Moghaddam M.H.Y., Honary M.T. Optimal location for mobile sink in wireless sensor networks. *Proc. of the IEEE Wireless Communications and Networking Conference (WCNC)*, 2010, pp. 1–6. https://doi.org/10.1109/wcnc.2010.5506171
20. Zhao M., Ma M., Yang Y. Efficient data gathering with mobile collectors and space-division multiple access technique in wireless sensor networks. *IEEE Transaction in Computers*, 2011, vol. 60, no. 3, pp. 400–417. https://doi.org/10.1109/tc.2010.140
21. Tang X., Xu J. Adaptive data collection strategies for lifetime-constrained wireless sensor networks. *IEEE Transactions in Parallel and Distributed Systems*, 2008, vol. 19, no. 6, pp. 721–734. https://doi.org/10.1109/tpds.2008.27
22. Li X., Jia Z., Zhang P., Zhang R., Wang H. Trust-based on-demand multipath routing in mobile ad hoc networks. *IET Information Security*, 2010, vol. 4, no. 4, pp. 212–232. https://doi.org/10.1049/iet-ifs.2009.0140
23. Praveena A., Sangeetha R., Prem P.E. Efficient trusted secure ad-hoc on-demand multipath distance vector in MANET. *International Journal of Engineering Development and Research*, 2017, vol. 5, no. 2, pp. 1614–1620.
24. Patel V.H., Zaveri M.A., Rath H.K. Trust based routing in mobile ad-hoc networks. *Lecture Notes on Software Engineering*, 2015, vol. 3, no. 4, pp. 318–324. https://doi.org/10.7763/lnse.2015.v3.212

Proc. of the 7th ACM International Symposium on Mobile ad hoc Networking and Computing. 2006. P. 190–201. https://doi.org/10.1145/1132905.1132927
10. Korzhuk V., Groznykh A., Menshikov A., Strecker M. Identification of attacks against wireless sensor networks based on behaviour analysis // Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA). 2019. V. 10. N 2. P. 1–21. https://doi.org/10.22667/JOWUA.2019.06.30.001
11. Tunca C., Isik S., Donmez M., Ersoy C. Distributed mobile sink routing for wireless sensor networks: A survey // IEEE Communications Surveys & Tutorials. 2014. V. 16. N 2. P. 877–897. https://doi.org/10.1109/surv.2013.100113.00293
12. Luo H., Ye F., Cheng J., Lu S., Zhang L. TTDD: Two-tier data dissemination in large-scale wireless sensor networks // Wireless Networks. 2005. V. 11. N 1-2. P. 161–175. https://doi.org/10.1007/s11276-004-4753-x
13. Behera T.M., Mohapatra S.K., Samal U.C., Khan M.S., Daneshmand M., Gandomi A.H. Residual energy-based cluster-head selection in wsns for iot application // IEEE Internet of Things Journal. 2019. V. 6. N 3. P. 5132–5139. https://doi.org/10.1109/jiot.2019.2897119
14. Jain S., Pattanaik K., Shukla A. QWRP: Query-driven virtual wheel based routing protocol for wireless sensor networks with mobile sink // Journal of Network and Computer Applications. 2019. V. 147. P. 102430. https://doi.org/10.1016/j.jnca.2019.102430
15. Maurya S., Gupta V., Jain V.K. LBRR: Load balanced ring routing protocol for heterogeneous sensor networks with sink mobility // Proc. of the 2017 IEEE Wireless Communications and Networking Conference (WCNC). 2017. P. 1–6. https://doi.org/10.1109/wcnc.2017.7925728
16. Tunca C., Isik S., Donmez M.Y., Ersoy C. Ring routing: An energy-efficient routing protocol for wireless sensor networks with a mobile sink // IEEE Transactions on Mobile Computing. 2015. V. 14. N 9. P. 1947–1960. https://doi.org/10.1109/TMC.2014.2366776
17. Jain S., Sharma S., Bagga N. A vertical and horizontal segregation based data dissemination protocol // Emerging Research in Computing, Information, Communication and Applications. Springer, 2016. P. 401–412. https://doi.org/10.1007/978-81-322-2553-9_37
18. Lin Y.C., Zhong J.-H. Hilbert-chain topology for energy conservation in large-scale wireless sensor networks // Proc. of the 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing (UIC/ATC). 2012. P. 225–232. https://doi.org/10.1109/uic-atc.2012.37
19. Khodashahi M.H., Tashtarian F., Moghaddam M.H.Y., Honary M.T. Optimal location for mobile sink in wireless sensor networks // Proc. of the IEEE Wireless Communications and Networking Conference (WCNC). 2010. P. 1–6. https://doi.org/10.1109/wcnc.2010.5506171
20. Zhao M., Ma M., Yang Y. Efficient data gathering with mobile collectors and space-division multiple access technique in wireless sensor networks // IEEE Transaction in Computers. 2011. V. 60. N 3. P. 400–417. https://doi.org/10.1109/tc.2010.140
21. Tang X., Xu J. Adaptive data collection strategies for lifetime-constrained wireless sensor networks // IEEE Transactions in Parallel and Distributed Systems. 2008. V. 19. N 6. P. 721–734. https://doi.org/10.1109/tpds.2008.27
22. Li X., Jia Z., Zhang P., Zhang R., Wang H. Trust-based on-demand multipath routing in mobile ad hoc networks // IET Information Security. 2010. V. 4. N 4. P. 212–232. https://doi.org/10.1049/iet-ifs.2009.0140
23. Praveena A., Sangeetha R., Prem P.E. Efficient trusted secure ad-hoc on-demand multipath distance vector in MANET // International Journal of Engineering Development and Research. 2017. V. 5. N 2. P. 1614–1620.
24. Patel V.H., Zaveri M.A., Rath H.K. Trust based routing in mobile ad-hoc networks // Lecture Notes on Software Engineering. 2015. V. 3. N 4. P. 318–324. https://doi.org/10.7763/lnse.2015.v3.212

Authors

Rangaraj Gopalakrishnan — PhD, Research Scholar, Periyar University, Salem, 636011, India, https://orcid.org/0000-0003-2456-265X, gopaldharun1978@gmail.com
Angamuthu Senthil Kumar — PhD, Assistant Professor, Thiruvalluvar Government Arts College, Rasipuram, 637401, India, sc 57674299700, https://orcid.org/0000-0001-5131-7428, senthilkumarmca76@gmail.com

Received 27.02.2023
Approved after reviewing 12.12.2023
Accepted 13.01.2024

Авторы

Гопалакришнан Рангарадж — PhD, исследователь, Университет Перияр, Салем, 636011, Индия, https://orcid.org/0000-0003-2456-265X, gopaldharun1978@gmail.com
Сентил Кумар Ангамуту — PhD, доцент, Государственный колледж искусств Тируваллувара, Расипурам, 637401, Индия, sc 57674299700, https://orcid.org/0000-0001-5131-7428, senthilkumarmca76@gmail.com

Статья поступила в редакцию 27.02.2023
Одобрена после рецензирования 12.12.2023
Принята к печати 13.01.2024

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2024-24-1-90-100
УДК 004.8, 004.93

Автоматизация распознавания сложной текстовой CAPTCHA с применением условной генеративно-состязательной нейронной сети

Александр Сергеевич Задорожный[1], Анастасия Андреевна Корепанова[2], Максим Викторович Абрамов[3]✉, Артем Азатович Сабреков[4]

[1] Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация
[2,3,4] Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация

[1] alexander.zadorozhnyy@yandex.ru, https://orcid.org/0009-0004-8903-7244
[2] aak@mail.ru, https://orcid.org/0000-0003-2962-8670
[3] mva@dscs.pro✉, https://orcid.org/0000-0002-5476-3025
[4] aas@dscs.pro, https://orcid.org/0009-0000-8378-563X

Аннотация

Введение. С быстрым развитием интернет-технологий проблемы сетевой безопасности продолжают обостряться. Так одним из наиболее распространенных методов поддержания безопасности и предотвращения вредоносных атак на интернет-ресурсы является полностью автоматизированный публичный тест Тьюринга (Completely Automated Public Turing test to tell Computers and Humans Apart, CAPTCHA). CAPTCHA чаще всего состоит из некоторого защитного кода, для обхода которого необходимо выполнить простую задачу. Однако наиболее широко используемым видом CAPTCHA до сих пор остается текстовый тип. В последние годы развитие компьютерного зрения и, в частности, нейронных сетей позволило снизить устойчивость к взлому текстовых CAPTCHA. Однако безопасность и надежность сложных CAPTCHA, содержащих много шума и искажений, все еще недостаточно изучена. Предметом данного исследования является устойчивость к распознаванию CAPTCHA. Особенность CAPTCHA — использование большого количества разнообразных искажений, причем на каждом отдельном изображении применяется свой набор искажений. При наличии искажений человеческий глаз не всегда может распознать, что изображено на фотографии. Данная работа состоит в тестировании устойчивости сайтов, использующих исследуемую CAPTCHA, к автоматизированному решению. Полученная методика тестирования может быть применена для последующей разработки рекомендаций по повышению эффективности механизмов защиты. **Метод.** В работе продемонстрирован новый метод к распознаванию CAPTCHA с использованием синтетического генератора и дискриминатора условно генеративно-состязательной архитектуры, а также программы-декодера, представляющей собой обученную сверточную нейронную сеть. **Основные результаты.** Результатом работы являются реализованные генератор и дискриминатор совместно с программой-декодером, решающей данный тип CAPTCHA. Точность распознавания построенной модели составила 63 % на изначально очень ограниченном наборе данных, что показывает риски информационной безопасности, которые могут нести сайты, использующие подобный вид CAPTCHA. **Обсуждение.** Несмотря на то, что получена точность распознавания ниже, чем у некоторых существующих методов (70–99 %), данный результат обладает следующей значимостью: продемонстрирован неизученный ранее в подобных работах вид CAPTCHA, а также предложен новый метод по распознаванию текстовых CAPTCHA.

Ключевые слова

текстовая CAPTCHA, глубокое обучение, условная генеративно-состязательная нейронная сеть, CGAN, CNN, информационная безопасность

Благодарности

Работа выполнена в рамках проекта по государственному заданию Санкт-Петербургского Федерального исследовательского центра Российской академии наук «Санкт-Петербургский институт информатики и автоматизации Российской академии наук» № FFZF-2022-0003.

Ссылка для цитирования: Задорожный А.С., Корепанова А.А., Абрамов М.В., Сабреков А.А. Автоматизация распознавания сложной текстовой CAPTCHA с применением условной генеративно-состязательной нейронной сети // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 90–100. doi: 10.17586/2226-1494-2024-24-1-90-100

© Задорожный А.С., Корепанова А.А., Абрамов М.В., Сабреков А.А., 2024

Automation of complex text CAPTCHA recognition using conditional generative adversarial networks

Alexander S. Zadorozhnyy[1], Anastasia A. Korepanova[2], Maxim V. Abramov[3]✉, Artem A. Sabrekov[4]

[1] St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation
[2,3,4] Saint Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation

[1] alexander.zadorozhnyy@yandex.ru, https://orcid.org/0009-0004-8903-7244
[2] aak@mail.ru, https://orcid.org/0000-0003-2962-8670
[3] mva@dscs.pro✉, https://orcid.org/0000-0002-5476-3025
[4] aas@dscs.pro, https://orcid.org/0009-0000-8378-563X

Abstract
With the rapid development of Internet technologies, the problems of network security continue to worsen. So, one of the most common methods of maintaining security and preventing malicious attacks is CAPTCHA (fully automated public Turing test). CAPTCHA most often consists of some kind of security code, to bypass which it is necessary to perform a simple task, such as entering a word displayed in an image, solving a basic arithmetic equation, etc. However, the most widely used type of CAPTCHA is still the text type. In the recent years, the development of computer vision and, in particular, neural networks has contributed to a decrease in the resistance to hacking of text CAPTCHA. However, the security and resistance to recognition of complex CAPTCHA containing a lot of noise and distortion is still insufficiently studied. This study examines CAPTCHA, the distinctive feature of which is the use of a large number of different distortions, and each individual image uses its own different set of distortions, that is why even the human eye cannot always recognize what is depicted in the photo. The purpose of this work is to assess the security of sites using the CAPTCHA text type by testing their resistance to an automated solution. This testing will be used for the subsequent development of recommendations for improving the effectiveness of protection mechanisms. The result of the work is an implemented synthetic generator and discriminator of the CGAN architecture, as well as a decoder program, which is a trained convolutional neural network that solves this type of CAPTCHA. The recognition accuracy of the model constructed in the article was 63 % on an initially very limited data set, which shows the information security risks that sites using a similar type of CAPTCHA can carry.

Keywords
text-based CAPTCHAs, deep learning, conditional generative adversarial network, CGAN, CNN, information security

Acknowledgements
The work was carried out within the framework of the project under the state assignment of SPC RAS SPIIRAS no. FFZF-2022-0003.

For citation: Zadorozhnyy A.S., Korepanova A.A., Abramov M.V., Sabrekov A.A. Automation of complex text CAPTCHA recognition using conditional generative adversarial networks. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 90–100 (in Russian). doi: 10.17586/2226-1494-2024-24-1-90-100

Введение

В области информационной безопасности используется множество методов защиты от автоматизированного сбора данных [1], включая полностью автоматизированный публичный тест Тьюринга (Completely Automated Public Turing Tests to Tell Computers and Humans Apart, CAPTCHA). CAPTCHA представляет собой защитный код, обход которого требует выполнения простого для человека задания, такого как ввод искаженных символов с изображения, решение арифметических уравнений или другие подобные задачи. Текстовая CAPTCHA, содержащая изображение с искаженными символами для ввода, является одним из наиболее распространенных типов.

Однако уровень безопасности сайтов с текстовой CAPTCHA снижается из-за развития технологий компьютерного зрения. Исследования показывают уязвимость многих видов искажений CAPTCHA перед современными системами. Однако устойчивость к распознаванию сложных CAPTCHA с большим числом шумов и различных искажений все еще неисследована до конца [2]. Сложные CAPTCHA состоят из изображений с разнообразными символами, подвергающимися различным искажениям и шумам, но остающимися решаемыми для человека. Рассматриваемый в данном исследовании вид CAPTCHA относится именно к такому типу. Несмотря на то, что она состоит из четырех случайных символов, каждый из которых может быть как одной из букв латинского алфавита, так и цифрой, CAPTCHA обладает большой вариативностью: от расстояния между буквами до наложенного поверх паттерна. К разным изображениям применяется суммарно более 30 видов искажений, включая искажение перспективы, растяжение или другие методы, которые делают CAPTCHA практически невозможной для распознания. Следствием такого разнообразия является не только трудность различить, что представлено на изображении даже для человеческого глаза, но и сложность в обучении модели распознавания CAPTCHA, которая бы одинаково хорошо распознавала все изображения, подвергшиеся разным видам искажений. Вопрос, насколько действительно сложно злоумышленнику построить систему прохождения такой системы защиты, ранее не исследовался.

Цель работы — исследование устойчивости к распознаванию сложной CAPTCHA с большим числом различных искажений. В перспективе полученные ре-

зультаты помогут сформировать рекомендации для повышения защищенности сайтов, использующих текстовые CAPTCHA. В настоящей работе изучен первый этап в достижении данной цели — из исследуемого множества возможных изображений CAPTCHA выбрано подмножество, соответствующее одному типу искажений и исследованы способы обучения решателя CAPTCHA такого вида. В дальнейшем полученные результаты будут использованы для построения решателя для полного набора изображений CAPTCHA. Теоретическая значимость работы состоит в предложении нового метода распознавания символов на изображении CAPTCHA. Например, метод может быть полезен в ситуациях, когда необходимо обеспечить баланс между сложностью распознавания для машины и удобством для человека. Отметим, что значимость работы заключается в потенциале использования полученных результатов для формулирования рекомендаций по повышению защищенности сайтов, использующих текстовые CAPTCHA.

Обзор работ

В ходе работы были изучены ведущие сервисы для решения CAPTCHA: CapMonster[1], Azcaptcha[2], RuCaptcha[3], Captcha.Guru[4], XEvil[5], BestCaptchaSolver[6], SolveCaptcha[7], AntiCaptcha[8]. Данные сервисы разделены на автоматизированные решатели (XEvil, CapMonster, Azcaptcha, Captcha.Guru) с использованием оптического распознавания символов (Optical Character Recognition, OCR) и на привлекающие наемных сотрудников (BestCaptchaSolver, SolveCaptcha, RuCaptcha, AntiCaptcha), тратя до 20 с на одну CAPTCHA. Подробные результаты сравнения сервисов, включая их стоимость и методы, представлены в табл. 1.

Таким образом, как это видно из табл. 1, имеющиеся сервисы либо не способны распознать имеющийся набор данных, либо прибегают к помощи человека и не способны делать это самостоятельно, что показывает защищенность данной CAPTCHA от простейших атак с использованием готовых инструментов.

В научных исследованиях последних лет устоявшимся подходом к разработке инструмента для распознавания символьных CAPTCHA является применение нейронных сетей [3–5]. В работе [3] предложена архитектура и методика обучения сверточной нейронной сети (Convolutional Neural Network, CNN) для распознавания CAPTCHA, сгенерированной с помощью библиотеки Python ImageCaptcha Library. Точность распознавания предложенного решения по метрике accuracy составила 98,9 %. В [2] разработана модель, основанная на долгой краткосрочной памяти, для распознавания моделей CAPTCHA разных видов, точность модели на CAPTCHA с 50 популярных сайтов находится в промежутке от 74,8 % до 97,3 %. Работа [4] посвящена распознаванию CAPTCHA, составленных из символов индийских языков хинди, бенгали и тамильского, с помощью модели CNN и достигнутой точностью 81 % на сгенерированном вручную наборе данных. Также для решения подобной задачи была предложена архитектура CNN с соединениями быстрого доступа (shortcut connections или skip-connections) [5] и итоговой точностью до 98,82 %. В работе [6] была применена функция потерь Focal loss для повышения точности распознавания на популярных примерах CAPTCHA и получена точность до 99 %.

Обучение модели CNN с различными модификациями демонстрирует высокую точность в данной задаче, однако требует большого объема данных для обучения: в работах [3–6] количество изображений CAPTCHA с известными ответами, использованных для обучения моделей, превышает 10 000. В [2] отмечено, что в зависимости от сложности CAPTCHA, для обучения модели, способной показывать точность более 80 %, может потребоваться несколько тысяч размеченных изображений.

Для уменьшения числа требуемых данных, размеченных вручную, в изученных научных работах были предложены следующие подходы: использование генеративно-состязательных сетей (Generative Adversarial Nets, GAN) для увеличения размера набора данных и использование алгоритма neural style transfer [7, 8] для упрощения изображения CAPTCHA. В работе [7] предложена модель, которая с помощью механизма neural style transfer упрощает изображения CAPTCHA, убирая искажения, после чего символы становится гораздо проще распознать с помощью другой модели. Точность распознавания после упрощения изображений с помощью предложенного решения составила 66–89 % на разных типах CAPTCHA. В [9–11] проведены исследования, направленные на изучение возможности применения GAN для решения CAPTCHA. Результаты позволили получить на небольшом обучающем наборе данных в количестве 500 размеченных вручную изображений точность от 40 до 100 % распознавания в зависимости от сложности CAPTCHA.

Таким образом, обзор показал, что существующие аналоги либо требуют ручного решения, что затратно по времени, либо используют OCR и решают только

[1] Capmonster.cloud: сервис для распознавания CAPTCHA [Электронный ресурс]. URL: https://capmonster.cloud (дата обращения: 10.11.2023).

[2] Azcaptcha: сервис для распознавания CAPTCHA [Электронный ресурс]. URL: https://azcaptcha.com (дата обращения: 10.11.2023).

[3] ruCaptcha: сервис для распознавания CAPTCHA [Электронный ресурс]. URL: https://rucaptcha.com (дата обращения: 10.11.2023).

[4] Captcha.Guru: сервис для распознавания CAPTCHA [Электронный ресурс]. URL: https://captcha.guru (дата обращения: 10.11.2023).

[5] XEvil: сервис для распознавания CAPTCHA [Электронный ресурс]. URL: https://xevil.net (дата обращения: 11.11.2023).

[6] BestCaptchaSolver: сервис для распознавания CAPTCHA [Электронный ресурс]. URL: https://bestcaptchasolver.com (дата обращения: 11.11.2023).

[7] SolveCaptcha: сервис для распознавания CAPTCHA [Электронный ресурс]. URL: https://solvecaptcha.com (дата обращения: 11.11.2023).

[8] AntiCaptcha: сервис для распознавания CAPTCHA. URL: https://anti-captcha.com (дата обращения: 11.11.2023).

Таблица 1. Сравнение сервисов, предоставляющих услуги решения CAPTCHA
Table 1. Comparison of services providing CAPTCHA solution

Сервис	Метод	Стоимость, долл. США за 1000 шт.	Решает ли рассматриваемую CAPTCHA
XEvil	OCR	0	нет
CapMonster	OCR	0,3	нет
Azcaptcha	OCR	0,4	нет
Captcha.Guru	OCR	0,1	да
BestCaptchaSolver	человек	0,6	да
SolveCaptcha	человек	0,5–1,0	да
RuCaptcha	человек	0,6	да
AntiCaptcha	человек	0,5	да

простые текстовые CAPTCHA, не решая конкретно задачу, поставленную в настоящей работе. А также, что ни одно из исследований не посвящено данному типу CAPTCHA с его разнообразием шумов и искажений. Исходя из этого, было предложено обучить собственную модель, основанную на существующих решениях, для оценки защищенности этого типа CAPTCHA от специально обученных моделей.

Постановка задачи

Задача, решаемая в данной работе, состоит в оценке защищенности к распознаванию сложной CAPTCHA с большим числом возможных искажений. Цель достигается через разработку модели для автоматического распознавания.

Исследуемая CAPTCHA представляет собой изображение с набором символов, каждый из которых может быть буквой из латинского алфавита или цифрой, с различными искажениями и шумами. Исходный набор данных состоял из 18 020 пар изображений и их расшифровок, собранных вручную (рис. 1). Как видно, набор CAPTCHA обладает существенной вариативностью: перечеркнутые линии, сильный шум, наложенный поверх узор, переменное расстояние между символами, а также разная ширина каждой отдельной буквы. При этом не было обнаружено никакой зависимости между размером символов и расстоянием между ними — все это сильно затрудняет дальнейшую обработку изображений и обучение единой модели, которая могла бы распознать все CAPTCHA.

Для упрощения обучения решателя в рамках исследования использована кластеризация набора данных на группы изображений с однородными шумами. Применена модель VGG16 [12] в сочетании с алгоритмом кластеризации KMeans [13]. Эксперимент определил оптимальное количество кластеров — 30. Во время эксперимента была обучена модель на одном из кластеров, включающем 600 изображений (рис. 2). Кластер выбран как наиболее визуально простой для тестирования обучаемости решателя, но содержащий сложные шумы, такие как нечеткие контуры и слипание символов. В ходе работы было проверено обучение модели на данном типе изображений и выполнена оценка возможности адаптации решения для других кластеров в дальнейшем.

Методы

В настоящей работе, из-за отсутствия готовых инструментов, было решено применить машинное обучение. В большинстве существующих решений для поставленной задачи используются CNN [14–16]. По этой причине, на основании результатов анализа существующих архитектур [14], были выбраны модели ResNet [17, 18] и Inception, обладающие большой рас-

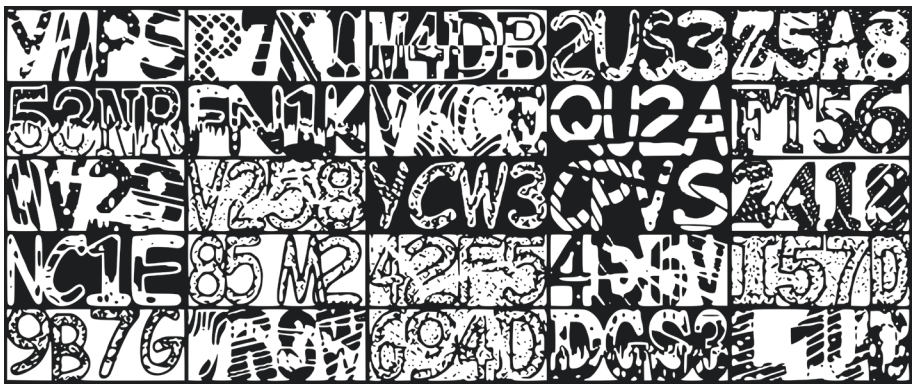


Рис. 1. Примеры CAPTCHA
Fig. 1. CAPTCHA examples



Рис. 2. Кластер, выбранный для обучения
Fig. 2. Cluster selected for training

пространенностью и хорошей точностью для решения различных задач [19, 20].

В работе [2] отмечено, что для корректного и эффективного обучения нейронной сети может потребоваться более 10 000 размеченных данных. А сбор и ручная маркировка такого количества реальных CAPTCHA является очень трудозатратным процессом. Таким образом, несмотря на то что CNN требуют большого количества образцов для обучения, они являются наиболее эффективными как для задач распознавания объектов, так и для решения задачи распознавания CAPTCHA. В то же время, чтобы устранить проблему с нехваткой обучающих данных, было решено использовать архитектуру GAN.

Архитектура GAN [21] представляет собой две ключевые модели: генеративную сеть, создающую синтетические примеры, и дискриминативную сеть, различающую синтезированные и реальные примеры. Однако в отличии от существующих работ, которые также основаны на архитектуре GAN [7, 8], предлагается не просто генерировать похожие изображения, а создавать полностью уникальные примеры из текстовой расшифровки. Тем самым модель становится независимой от разнообразия начальных данных, воссоздавая форму и стиль любых входных данных. Для этого возможно воспользоваться условно генеративно-состязательной архитектурой (CGAN). CGAN позволяет не только генерировать CAPTCHA с заданными искажениями и случайным набором букв, но и указывать требуемую расшифровку. В архитектуре CGAN генератор (Generator) и дискриминатор (Discriminator) настраиваются с использованием дополнительной информации, такой как метка класса (рис. 3). На данной схеме y, y_1, y_2, ..., y_n — некоторая метка, которую нужно преобразовать в изображение, необработанный вход (шум), представлен в виде вектора случайных чисел — **z**. Выходное значение, которое получается в результате работы генератора, — x^*, а $x^*|y$ — x^* с учетом или при условии y, σ обозначена некоторая функция активации, Training dataset — датасет с тренировочными данными, где настоящие примеры с метками показаны как (x, y).

В результате итоговая архитектура решения состоит из двух отдельных компонентов: CGAN, который генерирует реалистичные примеры, неотличимые от оригинальных CAPTCHA, причем с известными расшифровками, и CNN, который, имея достаточное количество пар пример-метка от предыдущего компонента после обучения на искусственных данных, сможет эффективно распознавать уже реальные CAPTCHA.

Программная реализация

Опишем реализацию двух ключевых компонент предлагаемого подхода к распознаванию CAPTCHA. Для этого в качестве основного языка выберем Python. Используем стандартные библиотеки для работы с изображениями Pillow и OpenCV, первая из которых позволяет удобно работать с изображениями, а вторая

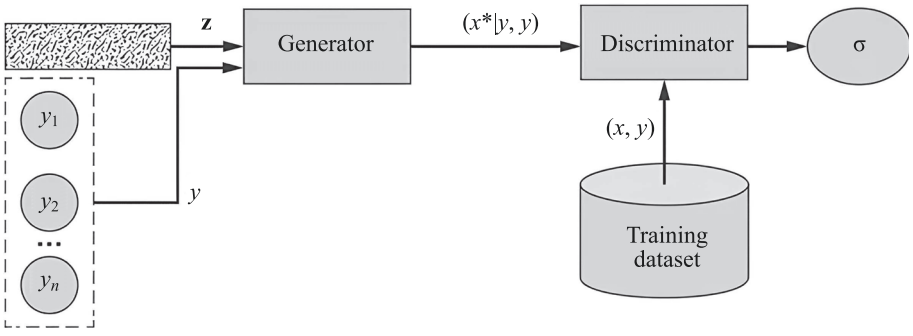


Рис. 3. CGAN архитектура
Fig. 3. CGAN architecture

применяет алгоритмы компьютерного зрения для распознавания и описания объектов на фотографии. Для работы непосредственно с обучением нейронных сетей применим библиотеку TensorFlow.

В первую очередь рассмотрим процесс обучения модели CGAN, которая позволяет генерировать синтетические образцы CAPTCHA с соответствующими расшифровками. Во-вторых, продемонстрируем решатель CAPTCHA на основе архитектуры ResNet-34, адаптированной под ее распознавание.

Реализация CGAN-архитектуры. Рассмотрим процесс обучения предложенной модели генерации новых CAPTCHA. В качестве меток для генерируемых примеров CGAN-архитектурой можно было бы использовать расшифровки CAPTCHA целиком. Однако данный тип CAPTCHA представляет собой последовательность из четырех символов, где каждый символ является буквой латинского алфавита или цифрой. Таким образом, всего возможных последовательностей символов 34^4 или 1 336 336. Чтобы CGAN научилась генерировать все возможные изображения CAPTCHA, набор данных должен содержать несколько примеров каждой метки, т. е. быть в разы больше этого числа. Из чего можно сделать вывод, что обучить CGAN модель на количестве исходных данных в 600 изображений с метками такого типа невозможно. По этой причине было решено научиться отдельно генерировать каждый из 34 символов и далее собрать из них CAPTCHA для обучения модели. Таким образом, вместо 34^4 используем всего 34 метки. Вручную был размечен набор данных, в котором содержится по 32 образца каждого символа.

Процесс построения CGAN сети состоит из трех основных этапов.

Этап 1. Генератор CAPTCHA. Чтобы формализовать определение, назовем меткой текстовый символ, который нужно преобразовать в изображение — y. А необработанный вход (шум), представленный в виде вектора случайных чисел, которые генератор использует в качестве отправной точки для синтеза поддельных примеров — **z**. Выходное значение, которое получается в результате работы генератора, — x^*. Тогда их комбинация дает условный $G(z, y) = x^*|y$ (читается как «x^* с учетом или при условии y»). Цель этого поддельного примера — выглядеть для дискриминатора как можно ближе к реальному примеру для данной метки (в нашем случае в качестве метки выступают символы латинского алфавита или цифры). Наилучшего результата удалось добиться при помощи CNN, состоящей из трех сверточных слоев и функцией активации выходного слоя — Sigmoid. Отметим, что генератор обновляет свои параметры, минимизируя следующую функцию потерь:

$$H_p(q) = -\frac{1}{N}\sum_{i=1}^{N} y_i \log_2(p(y_i)) + (1 - y_i)\log_2(1 - (p(y_i)), \qquad (1)$$

где N — количество данных, на основе которых вычисляется функция потерь; $p(x)$ — функция вероятности; $H_p(q)$ — логистическая функция потерь (binary cross-entropy).

Этап 2. Дискриминатор CAPTCHA. Дискриминатор получает на вход как настоящие примеры с метками (x, y), так и поддельные примеры с меткой, которая использовалась для их генерации $(x^*|y, y)$. На синтетических примерах он учится распознавать фальшивые изображения. Метки y ему необходимы, чтобы научиться проверять, действительно ли сгенерированное изображение соответствует изначальному заданному символу. Таким образом, цель дискриминатора заключается в том, чтобы определить, не только является ли образ, полученный генератором, реальным или сгенерированным, но и в том, совпадает ли полученный результат с указанной меткой. Для этого была использована нейронная сеть, состоящая из трех сверточных слоев с функцией активации LeakyRelu и одного полносвязного слоя с той же функцией активации (1) для BackPropogation, что и для генератора. В качестве ответа дискриминатор выдает одно число — вероятность того, что входные данные являются реальной совпадающей парой, т. е. символом из реального набора данных, соответствующего его метке. Вероятность вычисляется при помощи сигмоидальной функции активации.

Этап 3. Обучение. Был использован оптимизатор Adam с коэффициентом скорости обучения 0,0003 для обучения генератора CAPTCHA. Общая цель обучения соответствует общему подходу GAN [22], однако с некоторыми небольшими изменениями:

$$\min_G \max_D V(D, G) =$$
$$= E_{x\sim p_{data(x)}}[\log_2 D(x, y)] + E_{z\sim p_z(z)}[\log_2(1 - D(G(z, y)))],$$

где E_x представляет собой математическое ожидание; $p_{data(x)}$ — распределение реальных данных, а следовательно $E_{x\sim p_{data(x)}}$ — математическое ожидание логарифма вероятности того, что дискриминатор правильно классифицирует некоторые данные x, где x выбирается из определенного распределения; $V(D, G)$ — функция потерь для обучения CGAN. Таким образом, генератор G пытается минимизировать разницу между сгенерированной парой изображения и метки и реальной, в то время как дискриминатор D пытается максимизировать ее. Во время обновления параметров одного из них параметры другого фиксировались.

Пример. Рассмотрим в качестве примера одно из изображений CAPTCHA, которое было составлено из сгенерированных символов. На рис. 4 видно, что стиль сгенерированной CAPTCHA соответствует оригинальной, при этом начертания букв различаются. Это позволяет воссоздать неограниченное количество похожих друг на друга, но не идентичных экземпляров для дальнейшего обучения, тем самым сократив количество времени на сбор и структуризацию данных.

Создание решателя CAPTCHA. Процесс создания решателя разделим на три этапа.

Этап 1. Подготовка обучающих данных. На том основании, что генерируются отдельные символы, а не вся CAPTCHA целиком, необходимо научиться соединять их обратно в одно изображение. В связи с этим было решено получившиеся изображения символов случайным образом расширять или сужать, а после



Рис. 4. Сравнение оригинальной (*a*) и сгенерированной нейросетью (*b*) CAPTCHA
Fig. 4. Comparison of the original (*a*) and generated by neural network (*b*) CAPTCHA

соединять в одну CAPTCHA. Для этого было применено полиномиальное распределение, которое позволяет выбрать n чисел суммы m. Ширина изображений CAPTCHA исследуемого вида составляет 96 пикселов и содержит 4 символа, вследствие чего, каждый из сгенерированных символов имеет ширину в 24 пиксела. Далее случайным образом выбиралась ширина каждого нового изображения символа таким образом, чтобы в сумме они давали 96, т. е. $n = 4$, а $m = 96$. После применения данного подхода к полученным ранее синтетическим одиночным изображениям $40 \times 24 \times 1$ и последовательного их объединения, были получены изображения размером $40 \times 96 \times 1$, содержащие нужное количество стилистически неотличимых от оригинала символов, при этом находящихся на случайном расстоянии друг от друга и имеющих случайную ширину.

Этап 2. Выбор способа хранения. Существует два варианта работы с данными: можно непрерывно генерировать образцы во время обучения или предварительно создать достаточное количество. В данной работе выбран второй подход из-за значительного увеличения скорости обучения. При этом использовано: для обучения — 5000 оригинальных изображений, полученных путем клонирования нескольких сотен исходных экземпляров, совместно с 40 000 синтетических CAPTCHA, для валидации — 2000 оригинальных и 10 000 сгенерированных изображений.

Этап 3. Поиск оптимальной архитектуры. На данном этапе выбрана архитектура CNN для распознавания полученных CAPTCHA. Сначала были протестированы архитектуры AlexNet [23], VGG19 [12], DenseNet [24], Unet [25]. Выбор данных архитектур обоснован их популярностью в области компьютерного зрения и распознавания образов, высокой точностью распознавания и результатами на известных наборах данных, таких как ImageNet. Однако на данных исследуемого набора вышеперечисленные архитектуры за различное количество эпох не смогли достичь точности выше 10 %, несмотря на то что на синтетических CAPTCHA показатели спустя несколько первых эпох стремились к 100 %.

Для решения поставленной задачи была выбрана архитектура XCeption [26], которая справилась с задачей намного лучше, но наилучшие результаты были получены при помощи архитектуры ResNet-34 [27]. Результаты исследования архитектур приведены в разделе «Вычислительный эксперимент». Приведем процесс обучения более детально. Сеть состоит из шести частей: входного уровня, четырех остаточных блоков и выходного уровня. Внутри остаточных блоков содержится по три, четыре, шесть и три сверточных блока, каждый из которых, в свою очередь, состоит из двух сверточных слоев. Особенность этой сети состоит в том, что у каждого такого блока есть свой «обходной путь», который напрямую связывает вход с выходом. Код реализованной модели выложен в GitHub-репозитории[1].

Вычислительный эксперимент

Выполним анализ результатов как метода синтеза обучающего множества, так и метода распознавания реализованных нейронных сетей на одном из кластеров исследуемой CAPTCHA.

Синтез обучающего множества. В данной работе использованы различные методы для генерации синтетических изображений, однако наилучших результатов удалось добиться при помощи архитектур GAN и CGAN. Синтетические изображения CAPTCHA, полученные с помощью GAN, сохраняют контуры букв, однако, в свою очередь не обладают схожей текстурой: серый фон, инверсия цветов и слияние символов. Все эти особенности подчеркивают нарушение структуры синтезированных образцов по сравнению с оригинальной CAPTCHA (рис. 5).

В настоящем исследовании использованы стандартные параметры обучения архитектуры GAN. Отметим, что применение данной архитектуры и различных экспериментов с варьированием гиперпараметров в разных диапазонах ($50 \leq$ latent space(размерность скрытого пространства) ≤ 200; $3 \leq$ количество слоев ≤ 10; $10^{-6} \leq$ $\leq$ коэффициент скорости обучения $\leq 10^{-2}$), не привели к существенным изменениям результатов. Для генератора изображений и дискриминатора, ответственного за распознавание CAPTCHA, использованы следующие гиперпараметры: размерность входных данных для генератора равна 100, а для дискриминатора — $40 \times 96 \times 1$; количество слоев генератора — 3. Применены следующие функции активации генератора: ReLU для скрытых слоев; Sigmoid для выходного слоя; коэффициент скорости обучения генератора равен $2 \cdot 10^{-5}$.

Как видно на представленных изображениях (рис. 4, *b* и рис. 5), CGAN преуспевает в создании более реалистичных образцов CAPTCHA, которые при этом полностью сохраняют текстуру оригинала. А полученные результаты подчеркивают не только возможность автоматизированного распознавания сложных текстовых CAPTCHA с использованием CGAN, но и полез-

[1] GitHub репозиторий с реализованным приложением [Электронный ресурс]. URL: https://github.com/ticslab/CGAN_CAPTCHA_SOLVER (дата обращения: 20.12.2023).

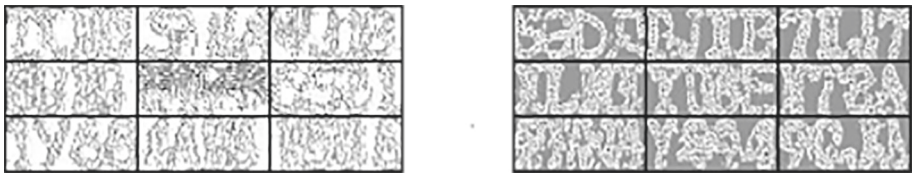


Рис. 5. Примеры CAPTCHA, сгенерированные при помощи GAN, на разных промежуточных этапах обучения
Fig. 5. CAPTCHA examples generated using GAN at various intermediate stages of training

Таблица 2. Наивысшие показатели метрик решателя CAPTCHA в зависимости от выбранной архитектуры
Table 2. The highest metrics of the CAPTCHA solver depending on the chosen architecture

Архитектура	Точность работы генератора, %	Точность работы распознавателя, %	Скорость, с
AlexNet	99,2	3,1	320,4
VGG19	98,7	5,8	301,2
DenseNet	99,3	7,4	334,3
Unet	97,8	9,4	320,8
Xception	98,1	58,1	430,2
ResNet-34	100	63,1	419,1

ность данного метода в контексте создания обучающего множества с высококачественными синтезированными образцами.

Распознавание CAPTCHA

Условия эксперимента. Характеристики машины, на которой проводилось тестирование: процессор Intel(R) Core (TM) i5-10300H, видеокарта NVIDIA GeForce GTX 1650 Ti, размер оперативной памяти 8GB.

Метрики. В рамках эксперимента была замерена точность (accuracy) распознавания 1000 сгенерированных (точность работы генератора CAPTCHA, %) и 1000 оригинальных (точность работы распознавателя CAPTCHA, %) образцов, а также скорость распознавания 5000 произвольных CAPTCHA (скорость, с).

Для вычисления точности распознавания использована следующая формула:

$$\text{Точность (Accuracy)} = \frac{\text{Количество верно распознанных CAPTCHA}}{\text{Общее количество тестовых CAPTCHA}} \cdot 100\ \%.$$

Результаты тестирования разных архитектур нейронных сетей для распознавания CAPTCHA представлены в табл. 2.

Таким образом, наилучшие показатели точности на исследуемых данных получены при использовании архитектуры ResNet-34 за 30 эпох с оптимизатором Adam, шагом обучения 10^{-6} и функцией активации Categorical Cross-Entropy.

Обсуждение

В работе разработано решение для текстовых CAPTCHA на основе технологии CGAN. Результаты исследования доказали потенциал эффективного применения генеративных нейронных сетей в данной области, особенно можно выделить высокую конфигурируемость и гибкость CGAN. Это позволяет настраивать распознавание CAPTCHA для различных изображений и контекстов, а также обеспечивает преимущества в адаптации к разнообразным стилям CAPTCHA. Эффективное использование CGAN также упрощает адаптацию к изменяющимся условиям и стилям CAPTCHA, сокращая время и ресурсы, затрачиваемые на обучение, по сравнению с традиционными методами. Таким образом, применение CGAN в области распознавания текстовых CAPTCHA открывает новые перспективы для автоматизированного анализа и обработки визуальных данных.

Дальнейшие исследования будут направлены на улучшение способности распознавателя CAPTCHA справляться с различными шумами и искажениями. Это включает в себя изучение различных методов обработки изображений и генерации данных. Также будет исследовано совместное обучение генератора синтетических образцов и распознавателя CAPTCHA для улучшения точности и сокращения времени обучения. В будущем планируется использование других моделей машинного обучения, включая алгебраические байесовские сети [28–30]. Реализованный инструмент обеспечивает легкое расширение набора распознаваемых CAPTCHA, независимо от их сложности и зашумленности, что позволяет адаптироваться к постоянно меняющимся схемам CAPTCHA.

Заключение

В работе изучена защищенность компаний, использующих конкретный тип CAPTCHA сложного вида от автоматического решения. В исследуемом типе CAPTCHA есть более 30 возможных искажений, что затрудняет использование готовых инструментов и обучение модели для ее распознавания. Рассмотрен новый подход к решению текстовых CAPTCHA на основе условной генеративно-состязательной и сверточной нейронных сетей. Несмотря на то, что точность рас-

познавания составляет 63 %, что ниже, чем у некоторых существующих методов (70–99 %), предложенное решение имеет ряд преимуществ. Во-первых, оно демонстрирует эффективное использование генеративных нейронных сетей в этой области. Другими словами, в отличие от аналогов, которые основаны на архитектуре генеративно-состязательных сетей, реализованный инструмент позволяет не просто генерировать незначительно отличающиеся внешне от начальных данных группы изображений, а создавать полностью уникальные образцы с заранее заданным стилем и расшифровкой. Во-вторых, новый подход требует значительно меньшего количества реальных CAPTCHA, что снижает необходимость в человеческом участии в процессе обучения нейронной сети и делает подход более практичным и экономически эффективным.

Таким образом, реализованный инструмент, несмотря на текущую точность в 63 %, обладает значительным потенциалом для улучшения и расширения, так как он способен адаптироваться к постоянно меняющимся схемам CAPTCHA, что делает его гибким и универсальным решением. Полученные результаты показывают, что CAPTCHA представленного формата является потенциально уязвимой к автоматическому распознаванию. В связи с этим в рамках дальнейшей работы будет исследоваться, сколько ресурсов требуется для повышения уязвимости и распознавания CAPTCHA со всеми типами искажений, чтобы на основе полученных результатов сформулировать рекомендации для сайтов, использующих текстовые CAPTCHA, для повышения их защищенности.

Литература

1. Корепанова А.А., Бушмелев Ф.В., Сабреков А.А. Технологии парсинга на Node.js в задаче агрегации сведений и оценки параметров грузовых маршрутов посредством извлечения данных из открытых источников // Компьютерные инструменты в образовании. 2021. № 3. С. 41–56. https://doi.org/10.32603/2071-2340-2021-3-41-56
2. Zi Y., Gao H., Cheng Z., Liu Y. An end-to-end attack on text CAPTCHAs // IEEE Transactions on Information Forensics and Security. 2019. V. 15. P. 753–766. https://doi.org/10.1109/TIFS.2019.2928622
3. Noury Z., Rezaei M. Deep-CAPTCHA: A deep learning based CAPTCHA solver for vulnerability assessment // ERN: Neural Networks & Related Topics (Topic). 2020. https://doi.org/10.2139/ssrn.3633354
4. Sahil Ahmed S., Anand K.M. Convolution neural network-based CAPTCHA recognition for indic languages // Advances in Intelligent Systems and Computing. 2021. V. 1407. P. 493–502. https://doi.org/10.1007/978-981-16-0171-2_46
5. Lu S., Huang K., Meraj T., Rauf H.T. A novel CAPTCHA solver framework using deep skipping Convolutional Neural Networks // PeerJ Computer Science. 2022. V. 8. P. e879. https://doi.org/10.7717/peerj-cs.879
6. Wang Z., Shi P. CAPTCHA recognition method based on CNN with focal loss // Complexity. 2021. V. 2021. P. 6641329. https://doi.org/10.1155/2021/6641329
7. Chen J., Luo X., Zhu L., Zhang Q., Gan Y. Handwritten CAPTCHA recognizer: a text CAPTCHA breaking method based on style transfer network // Multimedia Tools and Applications. 2023. V. 82. N 9. P. 13025–13043. https://doi.org/10.1007/s11042-021-11485-9
8. Bostik O., Horak K., Kratochvila L., Zemcik T., Bilik S. Semi-supervised deep learning approach to break common CAPTCHAs // Neural Computing and Applications. 2021. V. 33. N 20. P. 13333–13343. https://doi.org/10.1007/s00521-021-05957-0
9. Le T.A., Baydin A.G., Zinkov R., Wood F. Using synthetic data to train neural networks is model-based reasoning // Proc. of the 2017 International Joint Conference on Neural Networks (IJCNN). 2017. P. 3514–3521. https://doi.org/10.1109/IJCNN.2017.7966298
10. Wang Y., Wei Y., Zhang M., Liu Y., Wang B. Make complex captchas simple: a fast text CAPTCHA solver based on a small number of samples // Information Sciences. 2021. V. 578. P. 181–194. https://doi.org/10.1016/j.ins.2021.07.040
11. Li C., Chen X., Wang H., Wang P., Zhang Y., Wang W. End-to-end attack on text-based CAPTCHAs based on cycle-consistent generative adversarial network // Neurocomputing. 2021. V. 433. P. 223–236. https://doi.org/10.1016/j.neucom.2020.11.057
12. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition // arXiv. 2014. arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556
13. Hartigan J.A., Wong M.A. Algorithm AS 136: A k-means clustering algorithm // Journal of the Royal Statistical Society. Series C (Applied Statistics). 1979. V. 28. N 1. P. 100–108. https://doi.org/10.2307/2346830

References

1. Korepanova A.A., Bushmelev F.V., Sabrekov A.A. Node.js parsing technologies in the task of aggregating information and evaluating the parameters of cargo routes by extracting data from open sources. *Computer Tools in Education Journal*, 2021, no. 3, pp. 41–56. (in Russian). https://doi.org/10.32603/2071-2340-2021-3-41-56
2. Zi Y., Gao H., Cheng Z., Liu Y. An end-to-end attack on text CAPTCHAs. *IEEE Transactions on Information Forensics and Security*, 2019, vol. 15, pp. 753–766. https://doi.org/10.1109/TIFS.2019.2928622
3. Noury Z., Rezaei M. Deep-CAPTCHA: A deep learning based CAPTCHA solver for vulnerability assessment. *ERN: Neural Networks & Related Topics (Topic)*, 2020. https://doi.org/10.2139/ssrn.3633354
4. Sahil Ahmed S., Anand K.M. Convolution neural network-based CAPTCHA recognition for indic languages. *Advances in Intelligent Systems and Computing*, 2021, vol. 1407, pp. 493–502. https://doi.org/10.1007/978-981-16-0171-2_46
5. Lu S., Huang K., Meraj T., Rauf H.T. A novel CAPTCHA solver framework using deep skipping Convolutional Neural Networks. *PeerJ Computer Science*, 2022, vol. 8, pp. e879. https://doi.org/10.7717/peerj-cs.879
6. Wang Z., Shi P. CAPTCHA recognition method based on CNN with focal loss. *Complexity*, 2021, vol. 2021, pp. 6641329. https://doi.org/10.1155/2021/6641329
7. Chen J., Luo X., Zhu L., Zhang Q., Gan Y. Handwritten CAPTCHA recognizer: a text CAPTCHA breaking method based on style transfer network. *Multimedia Tools and Applications*, 2023, vol. 82, no. 9, pp. 13025–13043. https://doi.org/10.1007/s11042-021-11485-9
8. Bostik O., Horak K., Kratochvila L., Zemcik T., Bilik S. Semi-supervised deep learning approach to break common CAPTCHAs. *Neural Computing and Applications*, 2021, vol. 33, no. 20, pp. 13333–13343. https://doi.org/10.1007/s00521-021-05957-0
9. Le T.A., Baydin A.G., Zinkov R., Wood F. Using synthetic data to train neural networks is model-based reasoning. *Proc. of the 2017 International Joint Conference on Neural Networks (IJCNN)*, 2017, pp. 3514–3521. https://doi.org/10.1109/IJCNN.2017.7966298
10. Wang Y., Wei Y., Zhang M., Liu Y., Wang B. Make complex captchas simple: a fast text CAPTCHA solver based on a small number of samples. *Information Sciences*, 2021, vol. 578, pp. 181–194. https://doi.org/10.1016/j.ins.2021.07.040
11. Li C., Chen X., Wang H., Wang P., Zhang Y., Wang W. End-to-end attack on text-based CAPTCHAs based on cycle-consistent generative adversarial network. *Neurocomputing*, 2021, vol. 433, pp. 223–236. https://doi.org/10.1016/j.neucom.2020.11.057
12. Simonyan K., Zisserman A. Very deep convolutional networks for large-scale image recognition. *arXiv*, 2014, arXiv:1409.1556. https://doi.org/10.48550/arXiv.1409.1556
13. Hartigan J.A., Wong M.A. Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 1979, vol. 28, no. 1, pp. 100–108. https://doi.org/10.2307/2346830

14. Khan A., Sohail A., Zahoora U., Qureshi A.S. A survey of the recent architectures of deep convolutional neural networks // Artificial Intelligence Review. 2020. V. 53. N 8. P. 5455–5516. https://doi.org/10.1007/s10462-020-09825-6
15. Oliseenko V., Abramov M. Identification of user profiles in online social networks: a combined approach with face recognition // Journal of Physics: Conference Series. 2021. V. 1864. P. 012119. https://doi.org/10.1088/1742-6596/1864/1/012119
16. Bushmelev F., Khlobystova A., Abramov M., Livshits L. Deep machine learning techniques in the problem of estimating the expression of psychological characteristics of a social media user // Studies in Systems, Decision and Control. 2023. V. 457. P. 315–324. https://doi.org/10.1007/978-3-031-22938-1_22
17. Shafiq M., Gu Z. Deep residual learning for image recognition: a survey // Applied Sciences. 2022. V. 12. N 18. P. 8972. https://doi.org/10.3390/app12188972
18. Hossen M.I., Hei X. A low-cost attack against the hcaptcha system // Proc. of the 2021 IEEE Security and Privacy Workshops (SPW). 2021. P. 422–431. https://doi.org/10.1109/SPW53761.2021.00061
19. Kapoor A., Shah R., Bhuva R., Pandit T. Understanding inception network architecture for image classification: Technical Report. 2020. https://doi.org/10.13140/RG.2.2.16212.35204
20. Mittal S., Kaushik P., Hashmi S., Kumar K. Robust real time breaking of image CAPTCHAs using inception v3 model // Proc. of the 2018 Eleventh International Conference on Contemporary Computing (IC3). 2018. P. 1–5. https://doi.org/10.1109/IC3.2018.8530607
21. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair Sh., Courville A., Bengio Y. Generative Adversarial Networks // Communications of the ACM. 2020. V. 63. N 11. P. 139–144. https://doi.org/10.1145/3422622
22. Mirza M., Osindero S. Conditional generative Adversarial Nets // arXiv. 2014. arXiv:1411.1784. https://doi.org/10.48550/arXiv.1411.1784
23. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // Communications of the ACM. 2017. V. 60. N 6. P. 84–90. https://doi.org/10.1145/3065386
24. Huang G., Liu Z., Van Der Maaten L., Weinberger K.Q. Densely connected convolutional networks // Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 2261–2269. https://doi.org/10.1109/CVPR.2017.243
25. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation // Lecture Notes in Computer Science. 2015. V. 9351. P. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
26. Chollet F. Xception: Deep learning with depthwise separable convolutions // Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2017. P. 1800–1807. https://doi.org/10.1109/CVPR.2017.195
27. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // Proc. of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2016. P. 770–778. https://doi.org/10.1109/CVPR.2016.90
28. Вяткин А.А., Тулупьев А.Л. Автоматизация проверки непротиворечивости идеалов конъюнктов с оценками вероятности истинности // Информационная безопасность регионов России (ИБРР-2021): материалы XII Санкт-Петербургской межрегиональной конференции. 2021. С. 330–332.
29. Вяткин А.А., Абрамов М.В., Харитонов Н.А., Тулупьев А.Л. Применение третичной структуры алгебраической байесовской сети в задаче апостериорного вывода // Вестник Южно-Уральского государственного университета. Серия: Вычислительная математика и информатика. 2023. Т. 12. № 1. С. 61–88. https://doi.org/10.14529/cmse230104
30. Вяткин А.А., Харитонов Н.А., Тулупьев А.Л. Применение алгебраических байесовских сетей в задаче распознавания рукописных символов // Региональная информатика и информационная безопасность: сборник трудов Юбилейной XVIII Санкт-Петербургской международной конференции. 2022. С. 538–542.

14. Khan A., Sohail A., Zahoora U., Qureshi A.S. A survey of the recent architectures of deep convolutional neural networks. *Artificial Intelligence Review*, 2020, vol. 53, no. 8, pp. 5455–5516. https://doi.org/10.1007/s10462-020-09825-6
15. Oliseenko V., Abramov M. Identification of user profiles in online social networks: a combined approach with face recognition. *Journal of Physics: Conference Series*, 2021, vol. 1864, pp. 012119. https://doi.org/10.1088/1742-6596/1864/1/012119
16. Bushmelev F., Khlobystova A., Abramov M., Livshits L. Deep machine learning techniques in the problem of estimating the expression of psychological characteristics of a social media user. *Studies in Systems, Decision and Control*, 2023, vol. 457, pp. 315–324. https://doi.org/10.1007/978-3-031-22938-1_22
17. Shafiq M., Gu Z. Deep residual learning for image recognition: a survey. *Applied Sciences*, 2022, vol. 12, no. 18, pp. 8972. https://doi.org/10.3390/app12188972
18. Hossen M.I., Hei X. A low-cost attack against the hcaptcha system. *Proc. of the 2021 IEEE Security and Privacy Workshops (SPW)*, 2021, pp. 422–431. https://doi.org/10.1109/SPW53761.2021.00061
19. Kapoor A., Shah R., Bhuva R., Pandit T. *Understanding inception network architecture for image classification: Technical Report*, 2020. https://doi.org/10.13140/RG.2.2.16212.35204
20. Mittal S., Kaushik P., Hashmi S., Kumar K. Robust real time breaking of image CAPTCHAs using inception v3 model. *Proc. of the 2018 Eleventh International Conference on Contemporary Computing (IC3)*, 2018, pp. 1–5. https://doi.org/10.1109/IC3.2018.8530607
21. Goodfellow I., Pouget-Abadie J., Mirza M., Xu B., Warde-Farley D., Ozair Sh., Courville A., Bengio Y. Generative Adversarial Networks. *Communications of the ACM*, 2020, vol. 63, no. 11, pp. 139–144. https://doi.org/10.1145/3422622
22. Mirza M., Osindero S. Conditional generative Adversarial Nets. *arXiv*, 2014, arXiv:1411.1784. https://doi.org/10.48550/arXiv.1411.1784
23. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 2017, vol. 60, no. 6, pp. 84–90. https://doi.org/10.1145/3065386
24. Huang G., Liu Z., Van Der Maaten L., Weinberger K.Q. Densely connected convolutional networks. *Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 2261–2269. https://doi.org/10.1109/CVPR.2017.243
25. Ronneberger O., Fischer P., Brox T. U-Net: Convolutional networks for biomedical image segmentation. *Lecture Notes in Computer Science*, 2015, vol. 9351, pp. 234–241. https://doi.org/10.1007/978-3-319-24574-4_28
26. Chollet F. Xception: Deep learning with depthwise separable convolutions. *Proc. of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 1800–1807. https://doi.org/10.1109/CVPR.2017.195
27. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proc. of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90
28. Vyatkin A., Tulupyev A. Automation of consistency checking of ideals of conjuncts with truth probability estimates. *Information Security of Russian Regions (ISRR-2021). Proc. of the XII St. Petersburg Interregional Conference*, 2021, pp. 330–332. (in Russian).
29. Vyatkin A., Abramov M., Kharitonov N., Tulupyev A. Application of tertiary structure of algebraic bayesian network in the problem of a posteriori inference. *Bulletin of the South Ural State University. Series "Computational Mathematics and Computer Science"*, 2023, vol. 12, no. 1, pp. 61–88. (in Russian). https://doi.org/10.14529/cmse230104
30. Vyatkin A., Kharitonov N., Tulupyev A. Application of algebraic bayesian networks in handwritten character recognition. *Regional Informatics and Information Security. Proc. of the Anniversary XVIII St. Petersburg International Conference*, 2022, pp. 538–542. (in Russian).

Авторы

Задорожный Александр Сергеевич — студент, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, https://orcid.org/0009-0004-8903-7244, alexander.zadorozhnyy@yandex.ru

Корепанова Анастасия Андреевна — младший научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, sc 57218191916, https://orcid.org/0000-0003-2962-8670, aak@mail.ru

Абрамов Максим Викторович — кандидат технических наук, старший научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, sc 56938320500, https://orcid.org/0000-0002-5476-3025, mva@dscs.pro

Сабреков Артем Азатович — младший научный сотрудник, Санкт-Петербургский Федеральный исследовательский центр Российской академии наук, Санкт-Петербург, 199178, Российская Федерация, sc 56938320500, https://orcid.org/0009-0000-8378-563X, aas@dscs.pro

Authors

Alexander S. Zadorozhnyy — Student, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, https://orcid.org/0009-0004-8903-7244, alexander.zadorozhnyy@yandex.ru

Anastasia A. Korepanova — Junior Researcher, Saint Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, sc 57218191916, https://orcid.org/0000-0003-2962-8670, aak@mail.ru

Maxim V. Abramov — PhD, Senior Researcher, Saint Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, sc 56938320500, https://orcid.org/0000-0002-5476-3025, mva@dscs.pro

Artem A. Sabrekov — Junior Researcher, Saint Petersburg Federal Research Center of the Russian Academy of Sciences, Saint Petersburg, 199178, Russian Federation, sc 56938320500, https://orcid.org/0009-0000-8378-563X, aas@dscs.pro

Статья поступила в редакцию 12.10.2023
Одобрена после рецензирования 14.12.2023
Принята к печати 24.01.2024

Received 12.10.2023
Approved after reviewing 14.12.2023
Accepted 24.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-101-111

Deep attention based Proto-oncogene prediction and Oncogene transition possibility detection using moments and position based amino acid features

Manickam Vijayalakshmi[1]✉, Mahesh Vallinayagi[2]

[1,2] Sri Sarada College for Women, Tirunelveli, 627011, India

[1] vijimarthresearch@gmail.com✉, https://orcid.org/0009-0001-2012-3169
[2] vallinayagimahesh@gmail.com, https://orcid.org/0009-0006-0552-0138

Abstract

The loss of the regulatory function of tumor suppression genes and mutations in Proto-oncogene are the common underlying mechanisms for uncontrolled tumor growth in the varied complex of disorders known as cancer. Oncogene can be curable by means of diagnosing and treating the possibilities of Proto-oncogene at earlier stages. Recently, machine learning approaches helps to focus and provide information about the possibilities of Proto-oncogene that may change into oncogene in different cancer types. This study helps to diagnose the possibilities of Proto-oncogene which are possible to change oncogenes at earlier stage. Thus, this present study proposed an efficient unique predictor of Proto-oncogene with the help of Bi-Directional Long Short Term Memory added with attention concept. This approach also find the probability of Proto-oncogene to oncogene using statistical moments, position based amino-acid composition representation and deep features extracted from the sequence. Consequently, this study suggests that using a K-Nearest Neighbor classifier it is possible to find probability of changing from Proto-oncogene to cancerous oncogene.

Keywords

Proto-oncogene, PseAAC, prediction, tumour suppression genes, TSG, machine learning, Bi-directional Long Short Term Memory (BiLSTM)

Acknowledgements

Special thanks to Dr. L. Rajagopala Marthandam, HOD of Medicine, TVMCH, India for his encouragement and support.

For citation: Vijayalakshmi M., Vallinayagi M. Deep attention based Proto-oncogene prediction and Oncogene transition possibility detection using moments and position based amino acid features. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 101–111. doi: 10.17586/2226-1494-2024-24-1-101-111

Основанное на особом интересе прогнозирование протоонкогена и обнаружение возможностей его мутации в онкоген на основе первоначального анализа последовательности аминокислот

Маникам Виджаялакшми[1]✉, Махеш Валлинаяги[2]

[1,2] Женский колледж Шри Сарада, Тирунелвели, 627011, Индия

[1] vijimarthresearch@gmail.com✉, https://orcid.org/0009-0001-2012-3169
[2] vallinayagimahesh@gmail.com, https://orcid.org/0009-0006-0552-0138

Аннотация

Утрата регуляторной функции генов, подавляющих опухоль, и мутации в протоонкогенах являются общими механизмами, лежащими в основе неконтролируемого роста опухолей при разнообразном комплексе заболеваний, известных как рак. Онкоген можно излечить путем диагностики и лечения возможностей протоонкогена на ранних стадиях. В последнее время подходы машинного обучения помогают сосредоточить внимание и предоставить информацию о возможностях протоонкогена, который может превращаться в онкоген при различных типах рака или изменять его на ранних стадиях. Предложен эффективный и уникальный предиктор протоонкогена с помощью нейронной сети Bi-Directional Long Short Term Memory (BiLSTM), дополненный концепцией ухода за больными. Этот подход также позволяет определить вероятность перехода от протоонкогена

© Vijayalakshmi M., Vallinayagi M., 2024

к онкогену с использованием статистических моментов, представления аминокислотного состава на основе положения и глубоких особенностей, извлеченных из последовательности. В работе применен классификатор K-Nearest Neighbor с помощью, которого можно определить вероятность перехода от протоонкогена к раковому онкогену.

Ключевые слова
протоонкогены, PseAAC, прогнозирование, гены опухолевой супрессии, TSG, машинное обучение, двунаправленная долговременная краткосрочная память, BiLSTM

Благодарности
Особая благодарность доктору Л. Раджагопале Мартандаму, руководителю медицины, TMCH, Индия, за его поощрение и поддержку.

Ссылка для цитирования: Виджаялакшми М., Валлинаяги М. Основанное на особом интересе прогнозирование протоонкогена и обнаружение возможностей его мутации в онкоген на основе первоначального анализа последовательности аминокислот // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 101–111 (на англ. яз.). doi: 10.17586/2226-1494-2024-24-1-101-111

Introduction

A series of nucleotide bases that make up each gene are responsible for carrying information about the development and operation of cells. This essentially happens when the cells transform the genetic code into proteins. In the human body, every protein has a particular purpose. Proto-oncogenes are common cellular genes that control cell division and development in humans [1]. It has been known for a long time that cancer is a result of loss of cell cycle control. The loss of control is a result of series of genetic mutations involving activation of Proto-oncogene to oncogenes and inactivation of tumour-suppressing genes.

The process of activation, which includes insertion mutations, point mutations, protein-protein interactions, retroviral transduction, gene amplification, chromosomal translocation, and transposon integration, can turn Proto-oncogene into oncogenes. Proto-oncogenes are frequently classed according to how closely their sequences resemble those of known proteins or according to how they typically behave inside cells [2]. Oncogenomics is the study of the genes linked to the development of cancer. By point mutations or gene amplification, Proto-oncogenes are frequently activated in transformed cells [3]. These genes may have a role in the genesis of cancer, and their identification may offer fresh perspectives on cancer treatment and diagnosis [4]. Oncogenes are thought to be distinguishable from other genes by identifying their unique mutation profile since the effects of mutations on genes activities are related to those effects [5]. Due to the significant heterogeneity of mutations across individuals and various cancer types, it is challenging to identify novel oncogenes aside from those that are often mutated [6]. Consequently, it is essential to create computational techniques for the finding.

Automatic protein functional observations have gained more interest lately because they narrow the search space for effective experimental annotation [7]. Various techniques, including prediction by sequence [8], protein-protein interactions [9], evolutionary relationships [10], protein structures and structure prediction algorithms [11], microarrays [12], and integration of data kinds [13], have enhanced tools for finding protein functional annotations. Additionally, a number of algorithms have been created to identify functional proteins from an amino acid sequence [14]. In general, protein functional detection research focuses on all types of functions, whether cancer-related or not. However, the subcategory of cancer-related functional detection is particularly helpful in cancer treatment. Hence, the present study has focused on which type of cancer is most possible along with protein functional detection for the given Proto-oncogene.

Related Reviews

The personalized therapy of cancer is a current research focus. This comprises a wide range of research projects centered on Proto-oncogene, oncogenes, DNA repair genes, DNA methylation, and tumour suppressor genes. For the purpose of identifying tumor suppressor genes in silico, several computational methods have been developed. Computer-based tools are capable of classifying complete protein functional detection as well as classifying various cancer types using clinical data, SNPs, and gene expressions in combination with conventional machine learning algorithms [15–20].

A given original protein sequence was utilized by Khan et al. [21] to extract location relative features for the identification of S-nitrosocystiene sites, which is the most common posttranslational modifications of proteins. In order to forecast Proto-oncogene, Malebary et al. [22] suggested statistical moments and position-based characteristics that were merged into Pseudo Amino-Acid Composition (PseAAC) based on Chou's 5-step rules and Random Forest (RF) classifier. A strategy for locating hydroxylysine sites was put out by Mahmood et al. [23] and it is based on a potent statistical and mathematical methodology that takes into account the sequence-order impact and the makeup of each item inside protein sequences. To ascertain if an amino acid substitution (AAS) affects protein function, the "Sorting Tolerant from Intolerant" (SIFT) method was employed in [24, 25]. Yang et al. [26] employed the word segmentation strategy to extract characteristics from the protein sequence. The characteristics were then classified using the Support Vector Machine (SVM).

Ali et al. [27] created a promising classification model with good membrane protein type discrimination. PseAAC is used to extract the silent characteristics of protein sequences. SVM, Nave Bayes, K-Nearest Neighbor, Voting Feature Interval, and Probabilistic Neural Network were used as classification techniques. A categorization system

for angiogenesis and cancer angiogenesis was put up by Allehaibi et al. [28]. Using a position- and composition-based method, variable-length proteome sequences were converted into fixed-length feature vectors. Utilizing statistical moments, position related information was further transformed into a condensed form. The best outcomes were determined using the three classifiers RF, Artificial Neural Network (ANN), and SVM. In order to detect TSGs and OGs by fusing extensive genetic and epigenetic data, Lyu et al. [29] created the algorithm Discovery of Oncogenes and Tumour SupressoR genes using Genetic and Epigenetic characteristics (DORGE). By incorporating nucleotide physicochemical characteristics into pseudo K-tuple nucleotide composition (PseKNC), Feng et al. [30] created a brand-new predictor known as iDNA6mA-PseKNC.

Huang et al. [31] established a technique to predict cancer proteins and used domain information to initially annotate protein interaction. Rahman et al. [32] introduced a system that directly extracts significant characteristics from protein sequences without relying on functional domain or structural information. They used the RF approach to rank the features after feature extraction and did the prediction scheme with SVM.

In order to determine a protein DNA-binding activity, Chowdhury et al. [33] created iDNAProt-ES, which makes use of both the evolutionary profile and structural data of proteins. They derived characteristics, such as amino acid composition, bigram, Dubchak features, auto-covariance, and segmentation distribution, from the Position-Specific Scoring Matrix (PSSM) profile. Ideal set of features are extracted using recursive feature elimination with the help of SPIDER2, the model was learned using SVM with a linear kernel.

Kumar et al. [34] exploit patient bias to find oncogenes far more effectively than current techniques by identifying it as a unique signal for cancer gene identification using RF classifier with relative/absolute position-based characteristics on Chou's PseAAC. Akmal et al. [35] suggested a unique predictor called iGlycoS-PseAAC.

Proto-oncogene to Oncogene Probability Score Detection Methodology ($PSD_{(p\text{-}o \rightarrow o)}$)

The proposed framework designed to find the probability of oncogene transformation from the given Proto-oncogene amino acid sequence; progress is clearly described in the following Fig. 1. This model provides relevant score for the chance of transforming Proto-oncogene into oncogene sequence in the type of breast, lung, kidney and collateral cancers or when the Proto-oncogene is stable as normal sequence. This approach extracts statistical moments and frequency and position based features [22] along with deep recurrent neural network of Bi-directional Long Short Term Memory (BiLSTM) features to find the chance of particular type of cancerous sequence formation or not with the help of traditional machine learning K-Nearest Neighbor classifier algorithm. This approach initially predicts whether a given sequence is Proto-oncogene or not with the help of BiLSTM network. Once it is identified as Proto-oncogene, it will check is there any possibility to be changed into oncogene.

Feature Extraction

This study utilized a variety of feature extraction strategies, such as Statistical Moments Calculation (raw, central, and Hahn), Position Relative Incidence Matrix (PRIM), Frequency Vector Determination, Absolute Position Incidence Vector (AAPIV) and Deep learning based feature with the help of BiLSTM.

Let peptide sample within the dataset be expressed as

$$\mathbf{PS}_l(Z) = S_0, S_1, \ldots, S_l,$$

where **PS** is Peptide Sample, Z contains the positive and negative samples, S_0, S_1… are the individual samples and l is non-uniform index indicating that the length of a sequence may vary. In other words, l represents the arbitrary length of the primary sequence which in this case is variable for each sample.

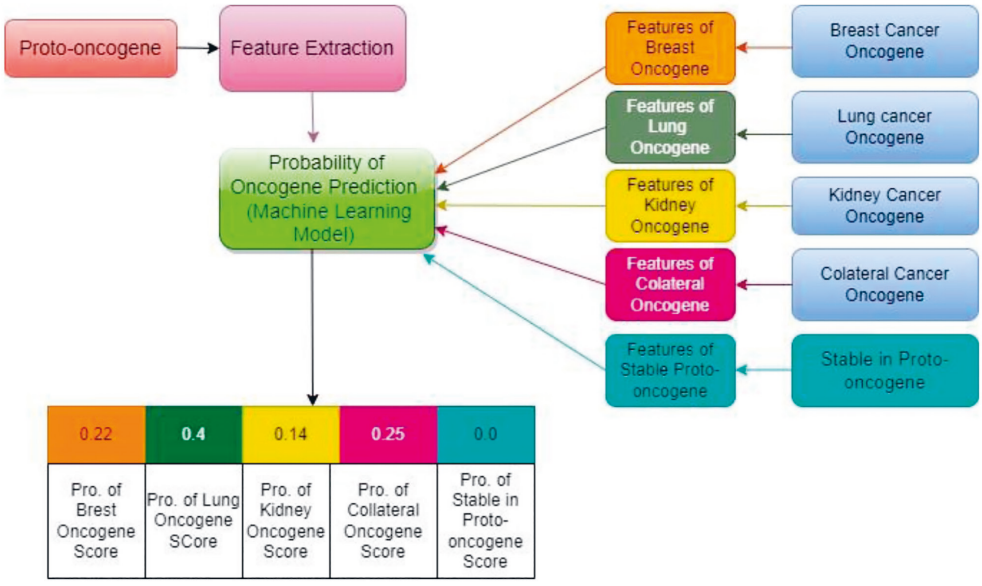


Fig. 1. The Architecture of $PSD_{(p\text{-}o \rightarrow o)}$ Model

Statistical Moments Calculation

The statistical moments are typically employed to extract particular qualities from data. Various moment sequences were employed to depict distinct attributes within the data. While some of these moments are useful for evaluating the direction and eccentricity of the data, others are useful for evaluating the magnitude of the data. There are several moment-defining polynomials depending on certain distributions have been proven by statisticians and mathematicians [36–40]. The suggested predictor raw moments, central moments, and Hahn moments were estimated up to order three. Ordering up to 3 generates enough information about the nature of data in numeric form [41].

Raw moments have location and scale variation features. Consequently, central moments are scale variant and location or position invariant. Orders up to three provide enough details on the type of data in numerical form. Additionally, the Hahn coefficient was determined using the Hahn polynomial which produces yet a different set of moments representing the initial data. The orthogonal features of these statistical moments led to their selection. The fact that orthogonal moments display a variety of features and may be utilized to recreate the original data means that they inherently include important properties that allow for exact categorization [42]. In general, these moments sufficiently transform information regarding the positioning and composition of residues in the primary structure.

The two-dimensional matrix **PS′** of size $n \times n$, which is a sequential transformation of all the amino acid residues of protein covered by **PS**, is the source of these moments computation (n is the order of the moment).

$$\mathbf{PS'} = \begin{bmatrix} ac_{11} & ac_{12} & \cdots & ac_{1n} \\ ac_{21} & ac_{22} & \cdots & ac_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ac_{n1} & ac_{n2} & \cdots & ac_{nn} \end{bmatrix}.$$

Using the function ω [43], **PS** is converted to **PS′**, and each of its arbitrary components, ac_{ij}, is an amino acid residue that has been placed in a two-dimensional setting. For all the moments up to degree 3, specific ordinal values of **PS′** elements were used. The raw moments are calculated by using equation:

$$MT_{ij} = \sum_{u=1}^{n} \sum_{v=1}^{n} u^i v^j ac_{uv},$$

where ac_{uv} is an arbitrary component of matrix **PS′**, and $i + j$ is the degree of the moments. Moreover, raw moments were denoted as MT_{00}, MT_{01}, MT_{02}, MT_{03}, MT_{10}, MT_{11}, MT_{12}, MT_{20}, MT_{21}, and MT_{30} for degree up to 3. The following equation is then used to determine central moments:

$$\eta_{ij} = \sum_{u=1}^{n} \sum_{v=1}^{n} (u - \bar{x})^i (v - \bar{y})^j ac_{uv},$$

where $\bar{x} = \dfrac{MT_{10}}{MT_{00}}$, and $\bar{y} = \dfrac{MT_{01}}{MT_{00}}$ which denotes the centroid of data.

PS was transformed into a square matrix **PS′** as it offers a substantial advantage for enumeration of Hahn moments. Discrete orthogonal moments in two dimensions require a square matrix as input. This orthogonal feature of Hahn moments suggests that they may be reversed using an inverse function. This reversible quality makes it easier to rebuild the data, which essentially means that it maintains the original data relative location, and sequence structure contained these moments.

For a one-dimensional matrix of size N, the Hahn polynomials of order n are calculated using the equation below.

$$h_n^{p,q}(r, N) = (N + V - 1)_n (N - 1)_n \times \sum_{k=0}^{n} (-1)^k \times$$

$$\times \frac{(-n)_k (-r)_k (2N + p + q - n - 1)_k}{(N + q - 1)_k (N - 1)_k} \frac{1}{k!},$$

where N is the size of the data array, V represents feature vector length n is the order of the moment; p and q are predefined constants. Additionally, the Pochhammer symbol q, which in turn employs the gamma operator as indicated in [41], is used in the equation. The two-dimensional Hahn moments are calculated using this Hahn coefficient as follows:

$$H_{ij} = \sum_{v=0}^{N-1} \sum_{u=0}^{N-1} ac_{uv} h_i^{p,q}(u, N).$$

The order of the moment is pointed out by the addition of i and j, that is, $i + j$; p, q are predefined constants; and ac_{uv} refers to any member in the square matrix **PS′**.

PRIM

To quantify the relative locations of amino acids and learn more about the relative positions of amino acid residues in the protein, a PRIM in the form of a 20×20 matrix was created. It is given as

$$\mathbf{PS_{PRIM}} = \begin{bmatrix} Seq_{1\to1} & Seq_{1\to2\ldots} & Seq_{1\to j\ldots} & Seq_{1\to20} \\ Seq_{2\to1} & Seq_{2\to2\ldots} & Seq_{2\to j\ldots} & Seq_{2\to20} \\ \vdots & \vdots & \vdots & \vdots \\ Seq_{i\to1} & Seq_{i\to2\ldots} & Seq_{i\to j\ldots} & Seq_{i\to20} \\ \vdots & \vdots & \vdots & \vdots \\ Seq_{20\to1} & Seq_{20\to2\ldots} & Seq_{20\to j\ldots} & Seq_{20\to20} \end{bmatrix}.$$

The sum of the positions of the j^{th} theresidue and the i^{th} residue initial occurrence for each element of this matrix is represented by the symbol $Seq_{i\to j}$. As a result, this matrix has 400 coefficients, which is a very large number. The opportunity to condense this information into a concise form is made possible by statistical moments. There are 30 coefficients total for degrees up to 3 after computing the PRIM raw, central, and Hahn moments. In the same way, the Reverse Position Relative Incidence Matrix (RPRIM) was created using a basic protein sequence.

The PRIM can be denoted as:

$$\mathbf{PS_{RPRIM}} = \begin{bmatrix} Seq_{1\to1} & Seq_{1\to2\ldots} & Seq_{1\to j\ldots} & Seq_{1\to20} \\ Seq_{2\to1} & Seq_{2\to2\ldots} & Seq_{2\to j\ldots} & Seq_{2\to20} \\ \vdots & \vdots & \vdots & \vdots \\ Seq_{i\to1} & Seq_{i\to2\ldots} & Seq_{i\to j\ldots} & Seq_{i\to20} \\ \vdots & \vdots & \vdots & \vdots \\ Seq_{20\to1} & Seq_{20\to2\ldots} & Seq_{20\to j\ldots} & Seq_{20\to20} \end{bmatrix}. \quad (1)$$

Statistical moments were used to decrease the dimensionality of RPRIM, resulting in the construction of a set of 30 elements. The PRIM describes the relative positions of amino acid residues in a polypeptide chain. This information is augmented by the RPRIM in (1), which reveals even more hidden information by repeating the operation on the opposite side of the primary sequence.

Frequency Vector Determination

Simple counting of each amino acid residue inside the main sequence yields the frequency vector. The frequency of each amino acid residue in the supplied sequence is represented by an element in the frequency vector. As a result, the frequency vector has 20 coefficients.

$$\mathbf{fv} = \{\propto_1, \propto_2 \ldots \ldots \propto_{20}\}.$$

Absolute Position Incidence Vector

The frequency matrix significantly provides information on how amino acid residues are composed. The AAPIV gives a summary of the residue location. It had a length of 20 elements and was composed of a single coefficient for each amino acid residue. The sum of the positions of every natural amino acid in the fundamental structure is included in elements of the AAPIV, which is given as equation:

$$\mathbf{PO} = \{pos_1, pos_2, pos_3, \ldots, pos_{20}\}.$$

Equation below made it possible to calculate any i^{th} element of the AAPIV.

$$pos_i = \sum_{PO=1}^{n} pos_k,$$

where pos_k represents the location of the i^{th} amino acid residue. Consequently, the Reverse Accumulative Absolute Position Vector (RAAPIV) assessed additional specific data based on the absolute positions of amino acids in peptide samples. RAAPIV was produced by reversing the basic sequence and calculating AAPIV. That can be denoted as

$$\mathbf{RAAPIV} = \{\lambda_1, \lambda_2, \lambda_3, \ldots, \lambda_{20}\},$$

where λ_i is the total number of positions in the main structure; then the i^{th} amino acid residue can be found.

BiLSTM

The architecture of BiLSTM is depicted in Fig. 2. The input sequences will be routed into the BiLSTM with size 128 filter. The outcome of BiLSTM is sent into maxpooling layer MP1 with size 2. Then, the Relu Layer will be applied to the output. The output of the Relu layer is then fed into the attention layer. The dropout layer will be given the results of the attention layer. The output of the BiLSTM layers is fed into the maxpooling layer MP2 with size 2. The outcome will then be applied to the Relu layer and then into the dense layer with size 128. The SoftMax classification will be used in the dense output to predict whether a given sequence is Proto-oncogene or not at the end of the architecture shown in the Fig. 2. The same layer model is used for extracting deep features as well as for the prediction of Proto-oncogene.

Feature Vector Description

The final step in processing primary sequences (**PS′**) through all of the aforementioned phases is combining them to create an accumulative feature vector. Two dimensional representation of the major sequence matrices **PRIM**, **PS′**, and **RPRIM** are changed into a concise form through calculating their statistical moments (raw, central, and Hahn). Thus, it produces 90 coefficients. Another 60 coefficients are included to the vector by pooling the frequency vector (**fv**), AAPIV (**PO**), and RAAPIV (**RAAPIV**). The deep features extracted by the final dense layer of the model with size of 128. Hence the final feature vector of the size 278 is used to find whether the given Proto-oncogene will change into oncogene or not.

Probability Score Detection using ML Algorithm

After features were extracted using the feature extraction approaches, a fixed-sized feature vector with 278 coefficients was created to be used for further processing in the machine learning algorithm for computing the possibilities of oncogene prediction. The features of Proto-oncogene sequences were extracted by utilizing certain

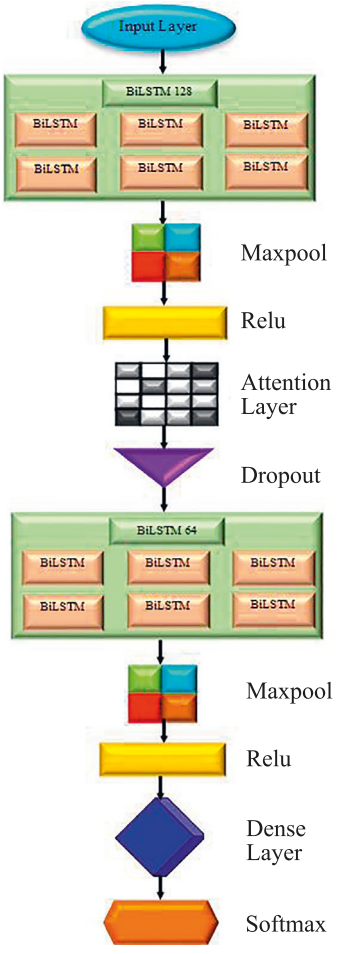


Fig. 2. BiLSTM Architecture

feature extraction strategies like Statistical Moments Calculation (raw, central, and Hahn), PRIM, Frequency Vector Determination, AAPIV, and they are fed into the K-Nearest Neighbor classifier Machine learning model which predicts the probability of the formation of Oncogene that causes cancer. Furthermore, the different categories of oncogene sequences, like breast cancer oncogene, lung cancer oncogene, kidney cancer oncogene, collateral oncogene and stable Proto-oncogene, were extracted by the same feature extraction strategies. All the above feature vectors are fed into the proposed oncogene probability prediction machine learning model. This proposed model compares the available features to identify the probability score of the Proto-oncogene to mutate into different types of oncogene. It also predicts the probability score of Proto-oncogene sample tested is a stable Proto-oncogene, which means it does not mutates into an oncogene.

Experimental Results

Dataset Description

The UniProt website consists of different kind resources, such as UniProt Knowledge Base (UniProtKB), UniProt Reference (UniRef), UniProt Archives (UniParc), and Protein Sets (Proteomes) of Fully Sequenced Genomes. Supporting datasets include protein information, such as transcript, distribution, sub cellular location, keywords, cross reference datasets, and disease, currently available in the UniProtKB protein entry. With the help of UniProt's "Search/ID Mapping" tool, different described in UniProt [43]. The following Fig. 3 shows the UniProt search and downloading section for genomic data. In this work, the 630 negative samples and 252 positive samples from the dataset provided by the ProtoPred [22] are used to train the BiLSTM with attention model for Proto-oncogene prediction as well as those 252 positive samples are used to find the probability of its status from Proto-oncogene to oncogene which causes different types of cancer. The following Table 1 shows the sample Proto-oncogene [22] and oncogenes in different types of cancer from UniProtKB [43].

The benchmark data set is split into k (10) disjoint fold partitions for cross-validation. Table 2 displays the findings of the KFold cross validations of the proposed model and the following Table 3 shows the performance of the Proto-oncogene prediction using the designed BiLSTM with attention model compared with state art of works, as an independent test, 30 % samples used for testing and remaining 70 % of samples used for training.

From the Table 3 it is found that the proposed BiLSTM_ATT model archives 97 % F1-Score, which is significantly better compared to the existing approaches. In the second phase of Proto-oncogene to oncogene transformation probability finding process, we trained the KNN classifier model with all the oncogene from different types of disease features along with stable Proto-oncogene features.

In the evaluation progress, for all of the 252 Proto-oncogene the same set of statistical moments, frequency based features along with BiLSTM features are extracted and score for each five classes, such as breast cancer, lung cancer, kidney cancer, colorectal cancer and the Proto-oncogene remains in the same state as denoted by stable, is estimated. Among the five scores, the class belonging to the maximum score is considered as the probability of changing Proto-oncogene to that particular type of oncogene or shows that there is no transition. The following Pie chart in the Fig. 4 clearly describes that the percentage of Proto-oncogene in the dataset has the probability of changing into particular types of cancer disease or it won't affect anything. A Pie chart is drawn based on the probability score attained by each Proto-oncogene sequence in the benchmark dataset [22], and the highest score identifies the type of oncogene that is most probable.

From the execution of the design, it is found that the up to 43.3 % of sequence data has the probability of changing into breast cancer oncogene. Similarly, 30 % related to kidney cancer, 13.3 % possibility of colorectal cancer, 10.8 % of lung cancer and 3.6 % of no transition. In this mode the probability of changes is estimated based on the highest score among all 5 classes. The highest score may be in any range from zero to one. In order to estimate the

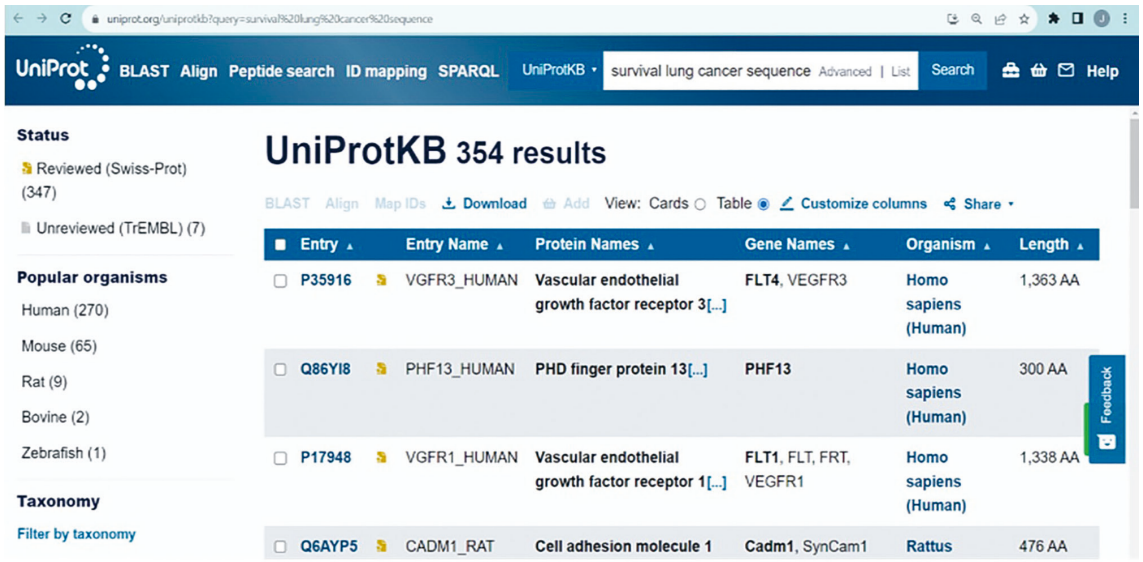


Fig. 3. UniProtKBSearch and its Result Page

Table 1. Samples studied

Sample of Proto-oncogene sequence [22]
MECPSCQHVSKEETPKFCSQCGERLPPAAPIADSENNNSTMASA…
MAHSCRWRFPARPGTTGGGGGGGRRGLGGAPRQRVPALLLPPGP…
MKLNPQQAPLYGDCVVTVLLAEEDKAEDDVVFYLVFLGSTLRHC…
MDQTCELPRRNCLLPFSNPVNLDAPEDKDSPFGNGQSNFSEPLN…
MTSGGSASRSGHRGVPMTSRGFDGSRRGSLRRAGARETASEAAD…
Sample of the Kidney Cancer Sequence
MFASCHCVPRGRRTMKMIHFRSSSVKSLSQEMRCTIRLLDDSEISCHI…
MGQDAFMEPFGDTLGVFQCKIYLLLFGACSGLKVTVPSHTVHGVRG…
MRLEELKRLQNPLEQVNDGKYSFENHQLAMDAENNIEKYPLNLQPL…
MTVKTEAAKGTLTYSRMRGMVAILIAFMKQRRMGLNDFIQKIANN…
MAEQDVENDLLDYDEEEEPQAPQESTPAPPKKDIKGSYVSIHSSGFR…
Sample of the Lung Cancer Sequence
MENEKENLFCEPHKRGLMKTPLKESTTANIVLAEIQPDFGPLTTP…
MENFTALFGAQADPPPPPTALGFGPGKPPPPPPPAGGGPGTA…
MKIIILLGFLGATLSAPLIPQRLMSASNSNELLLNLNNGQLLPLQL…
MPVSTSLHQDGSQERPVSLTSTTSSSGSSCDSRSAMEEPSSSEA…
MAFSDLTSRTVHLYDNWIKDADPRVEDWLLMSSPLPQTILLGF…
Sample of the Breast Cancer Sequence
MGQDAFMEPFGDTLGVFQCKIYLLLFGACSGLKVTVPSHTVHG…
MVQYELWAALPGASGVALACCFVAAAVALRWSGRRTARGAV…
MNYSLHLAFVCLSLFTERMCIQGSQFNVEVGRSDKLSLPGFENL…
MAGFGAMEKFLVEYKSAVEKKLAEYKCNTNTAIELKLVRFPEDL…
MDRSKENCISGPVKATAPVGGPKRVLVTQQFPCQNPLPVNSG…
Sample of the Colorectal Cancer Sequence
MKIIILLGFLGATLSAPLIPQRLMSASNSNELLLNLNNGQLLPLQ…
MSEKPKVYQGVRVKITVKELLQQRRAHQAASGGTRSGGSSVH…
MELSGATMARGLAVLLVLFLHIKNLPAQAADTCPEVKVVGLEG…
MIPPADSLLKYDTPVLVSRNTEKRSPKARLLKVSPQQPGPSGSA…
MEGAALLRVSVLCIWMSALFLGVGVRAEEAGARVQQNVPSGT…

Table 2. Findings of the KFold testing for Accuracy and F1-Score, %

Fold #	Accuracy	Precision	Recall	F1-Score
1	94	96	90	92
2	97	98	96	97
3	94	96	90	92
4	96	96	95	95
5	94	96	90	92
6	97	97	97	97
7	95	97	92	94
8	97	98	96	97
9	98	99	98	98
10	100	100	100	100
Average	96	96	94	95

chance of changing from Proto-oncogene to oncogene in reliable form, a marginal threshold is fixed for the maximum score checking. In this work the threshold is fixed as 0.5. Here the chance of oncogene cancer type is determined not only by its maximum score, it also checks if that maximum score is greater than the pre-defined threshold. The probability of changes from Proto-oncogene to particular four types of cancerous gene or stable in the current state is shown in the following Fig. 5 with the threshold as 0.5. The predicted data whose highest score below 0.5 means it is in unpredictable class. Whereas in Fig. 4, the data is assigned to the class which has highest predicted score among all trained five classes.

From the above Fig. 5, it is clear that the based on the threshold the probability of changing from Proto-oncogene to breast cancer oncogene is reduced from 43.3 % without threshold (Fig. 4) to 35.4 %. The following Fig. 6 shows the probability score for changing from Proto-oncogene to oncogene for the given testing unknown Proto-oncogene sequence in the Table 4.

Table 3. Results of Proto-oncogene prediction using BiLSTM with Spatial Attention, %

Methods	Accuracy	F1-Score	Precision	Recall	Specificity	Mcc	AUC
PSSM [44]	81	78	77	79	79	56	83
PseAAC [45]	85	82	81	83	83	64	89
ProtoPred_RF [22]	97	96	94	98	98	92	97
BiLSTM_ATT Model	97	97	96	98	98	94	98

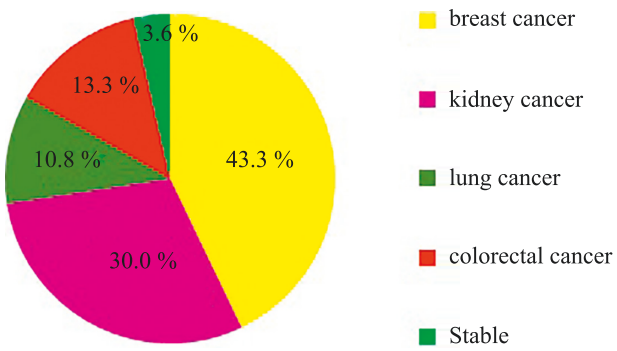


Fig. 4. Percentage of Proto-oncogene has the possibility to change into particular type of oncogene or remains as it as Proto-oncogene

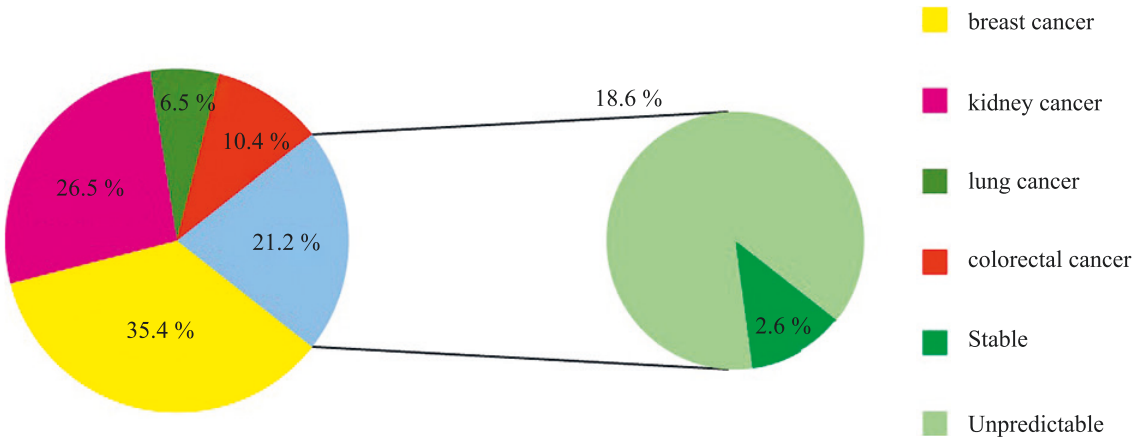


Fig. 5. Probability of Transition from Proto-oncogene to Oncogenes with Score Threshold

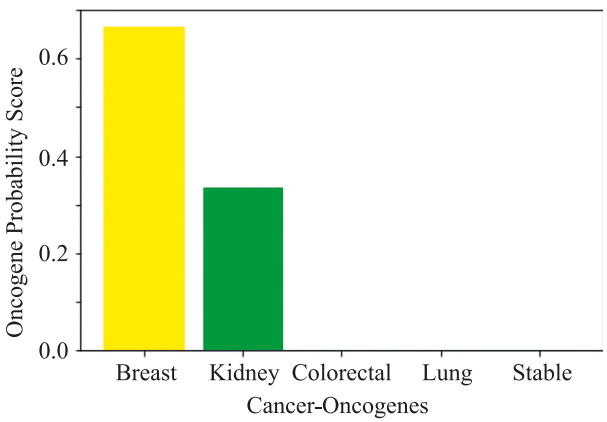


Fig. 6. Predicted Probability Score for the Cancer Oncogene

Table 4. Unknown Proto-oncogene Sequence for Prediction

MEYMSTGSDEKEEIDLLIKHLNVSEVIDIMENLYASEEPGVYEPSLMTMYPDS
NQNEERSESLLRSGQEVPWLSSVRYGTVEDLLAFANHVSNMTKHFYGRRPQ
ECGILLNMVISPQNGRYQIDSDVLLVPWKLTYRNIGSGFVPRGAFGKVYLA
QDMKTKKRMACKLIPIDQFKPSDVEIQACFRHENIAELYGAVLWGDTVHLFM
EAGEGGSVLEKLESCGPMREFEIIWVTKHILKGLDFLHSKKVIHHDIKPSNIV
FMSTKAVLVDFGLSVKMTEDVYLPKDLRGTEIYMSPEVILCRGHSTKADIY
SLGATLIHMQTGTPPWVKRYPRSAYPSYLYIIHKQAPPLEDIAGDCSPGMRELI
EAALERNPNHRPKAADLLKHEALNPPREDQPRCQSLDSALFERKRLLSRKEL
QLPENIADSSCTGSTEESEVLRRQRSLYIDLGALAGYFNIVRGPPTLEYG

Conclusion

Mutations in Proto-oncogene are the leading causes of cancer because of exposure to a mutagen. Proto-oncogene proteins are formed when Proto-oncogene are translated. These proteins function as a biomarker for cancer susceptibility. The proposed approach offers a reliable in-silico method for detecting such proteins. The suggested method incorporates all of the suggestions from the state of the art to create a computationally intelligent predictor. The features of a two-dimensional representation of the key structure of proteins, such as Statistical Moment Calculation (raw, central, and Hahn), Position Relative Incidence Matrix, Frequency Vector Determination, Absolute Position Incidence Vector and Deep RNN features, are gathered to form feature vectors. Following the extraction of feature vectors from both positive and negative sequences, the data is used to train the K-Nearest Neighbor machine learning classifier algorithm that is employed to find the probability score for each classes, such as stable, unpredictable or breast, lung, kidney and collateral cancers. The obtained results are evaluated using the benchmark ProroPred dataset. They show that the designed BiLSTM_Attention model achieves 97 % accuracy for the prediction of Proto-oncogene. The deep feature extraction along with statistical and moments features support the model to find the probability of transformation from Proto-oncogene to oncogene. This approach helps to find transformation from of Proto-oncogene to oncogene at earlier stage, which saves human life.

References

1. Williams D.E., Eisenman J., Baird A., Rauch C., Van Ness K., March C.J., Park L.S., Martin U., Mochizukl D.Y., Boswell H.S., Burgess G.S., Cosman D., Lyman S.D. Identification of a ligand for the c-kit Proto-oncogene. *Cell*, 1990, vol. 63, no. 1, pp. 167–174. https://doi.org/10.1016/0092-8674(90)90297-r
2. Cooper G.M. *Oncogenes*. 2nd ed. Jones and Bartlett Publishers Inc. Boston, 1995, 384 p.
3. Mulligan L.M., Kwok J.B., Healey C.S., Elsdon M.J., Eng C., Gardner E., Love D.R., Mole S.E., Moore J.K., Papi L., Ponder M.A., Telenius H., Tunnacliffe A., Ponder B.A. Germ-line mutations of the *RET* Proto-oncogene in multiple endocrine neoplasia type 2A. *Nature,* 1993, vol. 363, no. 6428, pp. 458–460. https://doi.org/10.1038/363458a0
4. Croce C.M. Oncogenes and cancer. *New England journal of medicine*, 2008, vol. 358, no. 5, pp. 502–511. https://doi.org/10.1056/NEJMra072367
5. Vogelstein B., Papadopoulos N., Velculescu V.E., Diaz L.A., Kinzler K.W. Cancer genome landscapes. *Science*, 2013, vol. 339, no. 6127, pp. 1546–1558. https://doi.org/10.1126/science.1235122
6. Pon J.R., Marra M.A. Driver and passenger mutations in cancer. *Annual Review of Pathology: Mechanisms of Disease*, 2015, vol. 10, pp. 25–50. https://doi.org/10.1146/annurev-pathol-012414-040312
7. Kulmanov M., Khan M.A., Hoehndorf R. DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier. *Bioinformatics*, 2018, vol. 34, no. 4, pp. 660–668. https://doi.org/10.1093/bioinformatics/btx624
8. Wass M.N., Sternberg M.J. ConFunc–functional annotation in the twilight zone. *Bioinformatics*, 2008, vol. 24, no. 6, pp. 798–806. https://doi.org/10.1093/bioinformatics/btn037
9. Deng M., Zhang K., Mehta S., Chen T., Sun F. Prediction of protein function using protein-protein interaction data. *Journal of Computational Biology*, 2003, vol. 10, no. 6, pp. 947–960. https://doi.org/10.1089/106652703322756168
10. Marcotte E.M., Pellegrini M., Ng H.L., Rice D.W., Yeates T.O., Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. *Science*, 1999, vol. 285, no. 5428, pp. 751–753. https://doi.org/10.1126/science.285.5428.751

Литература

1. Williams D.E., Eisenman J., Baird A., Rauch C., Van Ness K., March C.J., Park L.S., Martin U., Mochizukl D.Y., Boswell H.S., Burgess G.S., Cosman D., Lyman S.D. Identification of a ligand for the c-kit Proto-oncogene // Cell. 1990. V. 63. N 1. P. 167–174. https://doi.org/10.1016/0092-8674(90)90297-r
2. Cooper G.M. Oncogenes / 2nd ed. Jones and Bartlett Publishers Inc. Boston, 1995. 384 p.
3. Mulligan L.M., Kwok J.B., Healey C.S., Elsdon M.J., Eng C., Gardner E., Love D.R., Mole S.E., Moore J.K., Papi L., Ponder M.A., Telenius H., Tunnacliffe A., Ponder B.A. Germ-line mutations of the RET Proto-oncogene in multiple endocrine neoplasia type 2A // Nature. 1993. V. 363. N 6428. P. 458–460. https://doi.org/10.1038/363458a0
4. Croce C.M. Oncogenes and cancer // New England journal of medicine. 2008. V. 358. N 5. P. 502–511. https://doi.org/10.1056/NEJMra072367
5. Vogelstein B., Papadopoulos N., Velculescu V.E., Diaz L.A., Kinzler K.W. Cancer genome landscapes // Science. 2013. V. 339. N 6127. P. 1546–1558. https://doi.org/10.1126/science.1235122
6. Pon J.R., Marra M.A. Driver and passenger mutations in cancer // Annual Review of Pathology: Mechanisms of Disease. 2015. V. 10. P. 25–50. https://doi.org/10.1146/annurev-pathol-012414-040312
7. Kulmanov M., Khan M.A., Hoehndorf R. DeepGO: predicting protein functions from sequence and interactions using a deep ontology-aware classifier // Bioinformatics. 2018. V. 34. N 4. P. 660–668. https://doi.org/10.1093/bioinformatics/btx624
8. Wass M.N., Sternberg M.J. ConFunc–functional annotation in the twilight zone // Bioinformatics. 2008. V. 24. N 6. P. 798–806. https://doi.org/10.1093/bioinformatics/btn037
9. Deng M., Zhang K., Mehta S., Chen T., Sun F. Prediction of protein function using protein-protein interaction data // Journal of Computational Biology. 2003. V. 10. N 6. P. 947–960. https://doi.org/10.1089/106652703322756168
10. Marcotte E.M., Pellegrini M., Ng H.L., Rice D.W., Yeates T.O., Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences // Science. 1999. V. 285. N 5428. P. 751–753. https://doi.org/10.1126/science.285.5428.751

11. Pal D., Eisenberg D. Inference of protein function from protein structure. *Structure*, 2005, vol. 13, no. 1, pp. 121–130. https://doi.org/10.1016/j.str.2004.10.015
12. Huttenhower C., Hibbs M., Myers C., Troyanskaya O.G. A scalable method for integration and functional analysis of multiple microarray datasets. *Bioinformatics*, 2006, vol. 22, no. 23, pp. 2890–2897. https://doi.org/10.1093/bioinformatics/btl492
13. Kourmpetis Y.A.I., van Dijk A.D.J., Bink M.C.A., van Ham M.R.C.H.J., terBraak C.J.F. Bayesian markov random field analysis for protein function prediction based on network data. *PLoS One*, 2010, vol. 5, no. 2. https://doi.org/10.1371/journal.pone.0009293
14. Radivojac P., Clark W.T., Oron T.R. et al. A large-scale evaluation of computational protein function prediction. *Nature Methods*, 2013, vol. 10, no. 3, pp. 221–227. https://doi.org/10.1038/nmeth.2340
15. Mihaylov I., Nisheva M., Vassilev D. Application of machine learning models for survival prognosis in breast cancer studies. *Information*, 2019, vol. 10, no. 3, pp. 93. https://doi.org/10.3390/info10030093
16. Cruz J.A., Wishart D.S. Applications of machine learning in cancer prediction and prognosis. *Cancer Informatics*, 2006, vol. 2, pp. 59–77. https://doi.org/10.1177/117693510600200030
17. Sotiriou C., Neo S.-Y., McShane L.M., Korn E.L., Long P.M., Jazaeri A., Martiat P., Fox S.B., Harris A.L., Liu E.T. Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proceedings of the National Academy of Sciences of the United States of America*, 2003, vol. 100, no. 18, pp. 10393–10398. https://doi.org/10.1073/pnas.1732912100
18. Vural S., Wang X., Guda C. Classification of breast cancer patients using somatic mutation profiles and machine learning approaches. *BMC Systems Biology*, 2016, vol. 10, no. 3, pp. 62. https://doi.org/10.1186/s12918-016-0306-z
19. Cai Z., Xu D., Zhang Q., Zhang J., Ngai S.-M., Shao J. Classification of lung cancer using ensemble-based feature selection and machine learning methods. *Molecular BioSystems*, 2015, vol. 11, no. 3, pp. 791–800. https://doi.org/10.1039/c4mb00659c
20. Kourou K., Exarchos T.P., Exarchos K.P., Karamouzis M.V. Fotiadis D.I. Machine learning applications in cancer prognosis and prediction. *Computational and Structural Biotechnology Journal*, 2015, vol. 13, pp. 8–17. https://doi.org/10.1016/j.csbj.2014.11.005
21. Khan Y.D., Batool A., Rasool N., Khan S.A., Chou K.-C.J. Prediction of nitrosocysteine sites using position and composition variant features. *Letters in Organic Chemistry*, 2019, vol. 16, no. 4, pp. 283–293. https://doi.org/10.2174/1570178615666180802122953
22. Malebary S.J., Khan R., Khan Y.D. ProtoPred: Advancing oncological research through identification of proto-oncogene proteins. *IEEE Access*, 2021, vol. 9, pp. 68788–68797. https://doi.org/10.1109/ACCESS.2021.3076448
23. Mahmood M.K., Ehsan A., Khan Y.D., Chou K.-C. iHyd-LysSite (EPSV): identifying hydroxylysine sites in protein using statistical formulation by extracting enhanced position and sequence variant feature technique. *Current Genomic*, 2020, vol. 21, no. 7, pp. 536–545. https://doi.org/10.2174/1389202921999200831142629
24. Kumar P., Henikoff S., Ng P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols*, 2009, vol. 4, no. 7, pp. 1073–1081. https://doi.org/10.1038/nprot.2009.86
25. Vaser R., Adusumalli S., Leng S., Sikic M., Ng P.C. SIFT missense predictions for genomes. *Nature Protocols*, 2016, vol. 11, no. 1, pp. 1–9. https://doi.org/10.1038/nprot.2015.123
26. Yang Y., Lu B.L., Yang W.Y. Classification of protein sequences based on word segmentation methods. *Proc. of the 6th Asia-Pacific Bioinformatics Conference (APBC '08)*, 2008, pp. 177–186. https://doi.org/10.1142/9781848161092_0020
27. Ali F., Hayat M. Classification of membrane protein types using Voting Feature Interval in combination with Chou's Pseudo Amino Acid Composition. *Journal of Theoretical Biology*, 2015, vol. 384, pp. 78–83. https://doi.org/10.1016/j.jtbi.2015.07.034
28. Allehaibi K., Daanial Khan Y., Khan S.A. iTAGPred: A two-level prediction model for identification of angiogenesis and tumor angiogenesis biomarkers. *Applied Bionics and Biomechanics*, 2021, vol. 2021, pp. 2803147. https://doi.org/10.1155/2021/2803147
29. Lyu J., Li J.J., Su J., Peng F., Chen Y.E., Ge X., Li W. DORGE: Discovery of Oncogenes and tumoR suppressor genes using Genetic and Epigenetic features. *Science Advances*, 2020, vol. 6, no. 46, pp. 1–17. https://doi.org/10.1126/sciadv.aba6784
30. Feng P., Yang H., Ding H., Lin H., Chen W., Chou K.C. iDNA6mA-PseKNC: Identifying DNA N6-methyladenosine sites by

11. Pal D., Eisenberg D. Inference of protein function from protein structure // Structure. 2005. V. 13. N 1. P. 121–130. https://doi.org/10.1016/j.str.2004.10.015
12. Huttenhower C., Hibbs M., Myers C., Troyanskaya O.G. A scalable method for integration and functional analysis of multiple microarray datasets // Bioinformatics. 2006. V. 22. N 23. P. 2890–2897. https://doi.org/10.1093/bioinformatics/btl492
13. Kourmpetis Y.A.I., van Dijk A.D.J., Bink M.C.A., van Ham M.R.C.H.J., terBraak C.J.F. Bayesian markov random field analysis for protein function prediction based on network data // PLoS One. 2010. V. 5. N 2. https://doi.org/10.1371/journal.pone.0009293
14. Radivojac P., Clark W.T., Oron T.R. et al. A large-scale evaluation of computational protein function prediction // Nature Methods. 2013. V. 10. N 3. P. 221–227. https://doi.org/10.1038/nmeth.2340
15. Mihaylov I., Nisheva M., Vassilev D. Application of machine learning models for survival prognosis in breast cancer studies // Information. 2019. V. 10. N 3. P. 93. https://doi.org/10.3390/info10030093
16. Cruz J.A., Wishart D.S. Applications of machine learning in cancer prediction and prognosis // Cancer Informatics. 2006. V. 2. P. 59–77. https://doi.org/10.1177/117693510600200030
17. Sotiriou C., Neo S.-Y., McShane L.M., Korn E.L., Long P.M., Jazaeri A., Martiat P., Fox S.B., Harris A.L., Liu E.T. Breast cancer classification and prognosis based on gene expression profiles from a population-based study // Proceedings of the National Academy of Sciences of the United States of America. 2003. V. 100. N 18. P. 10393–10398. https://doi.org/10.1073/pnas.1732912100
18. Vural S., Wang X., Guda C. Classification of breast cancer patients using somatic mutation profiles and machine learning approaches // BMC Systems Biology. 2016. V. 10. N 3. P. 62. https://doi.org/10.1186/s12918-016-0306-z
19. Cai Z., Xu D., Zhang Q., Zhang J., Ngai S.-M., Shao J. Classification of lung cancer using ensemble-based feature selection and machine learning methods // Molecular BioSystems. 2015. V. 11. N 3. P. 791–800. https://doi.org/10.1039/c4mb00659c
20. Kourou K., Exarchos T.P., Exarchos K.P., Karamouzis M.V. Fotiadis D.I. Machine learning applications in cancer prognosis and prediction // Computational and Structural Biotechnology Journal. 2015. V. 13. P. 8–17. https://doi.org/10.1016/j.csbj.2014.11.005
21. Khan Y.D., Batool A., Rasool N., Khan S.A., Chou K.-C.J. Prediction of nitrosocysteine sites using position and composition variant features // Letters in Organic Chemistry. 2019. V. 16. N 4. P. 283–293. https://doi.org/10.2174/1570178615666180802122953
22. Malebary S.J., Khan R., Khan Y.D. ProtoPred: Advancing oncological research through identification of proto-oncogene proteins // IEEE Access. 2021. V. 9. P. 68788–68797. https://doi.org/10.1109/ACCESS.2021.3076448
23. Mahmood M.K., Ehsan A., Khan Y.D., Chou K.-C. iHyd-LysSite (EPSV): identifying hydroxylysine sites in protein using statistical formulation by extracting enhanced position and sequence variant feature technique // Current Genomic. 2020. V. 21. N 7. P. 536–545. https://doi.org/10.2174/1389202921999200831142629
24. Kumar P., Henikoff S., Ng P.C. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm // Nature Protocols. 2009. V. 4. N 7. P. 1073–1081. https://doi.org/10.1038/nprot.2009.86
25. Vaser R., Adusumalli S., Leng S., Sikic M., Ng P.C. SIFT missense predictions for genomes // Nature Protocols. 2016. V. 11. N 1. P. 1–9. https://doi.org/10.1038/nprot.2015.123
26. Yang Y., Lu B.L., Yang W.Y. Classification of protein sequences based on word segmentation methods // Proc. of the 6th Asia-Pacific Bioinformatics Conference (APBC '08). 2008. P. 177–186. https://doi.org/10.1142/9781848161092_0020
27. Ali F., Hayat M. Classification of membrane protein types using Voting Feature Interval in combination with Chou's Pseudo Amino Acid Composition // Journal of Theoretical Biology. 2015. V. 384. P. 78–83. https://doi.org/10.1016/j.jtbi.2015.07.034
28. Allehaibi K., Daanial Khan Y., Khan S.A. iTAGPred: A two-level prediction model for identification of angiogenesis and tumor angiogenesis biomarkers // Applied Bionics and Biomechanics. 2021. V. 2021.P. 2803147. https://doi.org/10.1155/2021/2803147
29. Lyu J., Li J.J., Su J., Peng F., Chen Y.E., Ge X., Li W. DORGE: Discovery of Oncogenes and tumoR suppressor genes using Genetic and Epigenetic features // Science Advances. 2020. V. 6. N 46. P. 1–17. https://doi.org/10.1126/sciadv.aba6784
30. Feng P., Yang H., Ding H., Lin H., Chen W., Chou K.C. iDNA6mA-PseKNC: Identifying DNA N6-methyladenosine sites by

incorporating nucleotide physicochemical properties into PseKNC. *Genomics*, 2018, vol. 111, no. 1, pp. 96–102. https://doi.org/10.1016/j.ygeno.2018.01.005
31. Huang C.H., Peng H.S., Ng K.L. Prediction of cancer proteins by integrating protein interaction, domain frequency, and domain interaction data using machine learning algorithms. *BioMed Research International*, 2015, vol. 2015, pp. 312047. https://doi.org/10.1155/2015/312047
32. Rahman M.S., Shatabda S., Saha S., Kaykobad M., Rahman M.S. DPP-PseAAC: a DNA-binding protein prediction model using Chou's general PseAAC. *Journal of Theoretical Biology*, 2018, vol. 452, pp. 22–34. https://doi.org/10.1016/j.jtbi.2018.05.006
33. Chowdhury S.Y., Shatabda S., Dehzangi A. iDNAProt-ES: Identification of DNA-binding proteins using evolutionary and structural features. *Scientific Reports*, 2017, vol. 7, pp. 14938. https://doi.org/10.1038/s41598-017-14945-1
34. Kumar R.D., Searleman A.C., Swamidass S.J., Griffith O.L., Bose R. Statistically identifying tumor suppressors and oncogenes from pan-cancer genome-sequencing data. *Bioinformatics*, 2015, vol. 31, no. 22, pp. 3561–3568. https://doi.org/10.1093/bioinformatics/btv430
35. Akmal M.A., Hussain W., Rasool N., Khan Y.D., Khan S.A., Chou K.-C. Using CHOU'S 5-steps rule to predict O-linked serine glycosylation sites by blending position relative features and statistical moment. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2021, vol. 18, no. 5, pp. 2045–2056. https://doi.org/10.1109/TCBB.2020.2968441
36. Khan Y.D., Ahmad F., Anwar M.W. Aneuro-cognitive approach for iris recognition using back propagation. *World Applied Sciences Journal*, 2012, vol. 16, no. 5, pp. 678–685.
37. Khan Y.D., Ahmed F., Khan S.A. Situation recognition using image moments and recurrent neural networks. *Neural Computing and Applications*, 2014, vol. 24, no. 7–8, pp. 1519–1529. https://doi.org/10.1007/s00521-013-1372-4
38. Khan Y.D., Khan N.S., Farooq S., Abid A., Khan S.A., Ahmad F., Mahmood M.K. An efficient algorithm for recognition of human actions. *Scientific World Journal*, 2014, vol. 2014, pp. 875879. https://doi.org/10.1155/2014/875879
39. Khan Y.D., Khan S.A., Ahmad F., Islam S. Iris recognition using image moments and K-means algorithm. *Scientific World Journal*, 2014, vol. 2014, pp. 723595. https://doi.org/10.1155/2014/723595
40. Mahmood S., Khan Y.D., Mahmood M.K. A treatise to vision enhancement and color fusion techniques in night vision devices. *Multimedia Tools and Applications*, 2018, vol. 77, no. 2, pp. 2689–2737. https://doi.org/10.1007/s11042-017-4365-y
41. Butt H., Rasool N., Khan Y.D. A treatise to computational approaches towards prediction of membrane protein and its subtypes. *The Journal of Membrane Biology*, 2017, vol. 250, no. 1, pp. 55–76. https://doi.org/10.1007/s00232-016-9937-7
42. Akmal M.A., Rasool N., Khan Y.D. Prediction of N-linked glycosylation sites using position relative features and statistical moments. *PLoS ONE*, 2017, vol. 12, no. 8, pp. 1–21. https://doi.org/10.1371/journal.pone.0181966
43. Pundir S., Magrane M., Martin M.J., O'Donovan C. Searching and navigating UniProt databases. *Current Protocols in Bioinformatics*, 2015. pp. 1.27.1–1.27.10 https://doi.org/10.1002/0471250953.bi0127s50
44. Delorenzi M., Speed T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions. *Bioinformatics*, 2002, vol. 18, no. 4, pp. 617–625. https://doi.org/10.1093/bioinformatics/18.4.617
45. Jia J., Liu Z., Xiao X., Liu B., Chou K.-C. iSuc-PseOpt: identifying lysine succinylation sites in proteins by incorporating sequence-coupling effects into pseudo components and optimizing imbalanced training dataset. *Analytical Biochemistry*, 2016, vol. 497, pp. 48–56. https://doi.org/10.1016/j.ab.2015.12.009

incorporating nucleotide physicochemical properties into PseKNC // Genomics. 2018. V. 111. N 1. P. 96–102. https://doi.org/10.1016/j.ygeno.2018.01.005
31. Huang C.H., Peng H.S., Ng K.L. Prediction of cancer proteins by integrating protein interaction, domain frequency, and domain interaction data using machine learning algorithms // BioMed Research International. 2015. V. 2015. P. 312047. https://doi.org/10.1155/2015/312047
32. Rahman M.S., Shatabda S., Saha S., Kaykobad M., Rahman M.S. DPP-PseAAC: a DNA-binding protein prediction model using Chou's general PseAAC // Journal of Theoretical Biology. 2018. V. 452. P. 22–34. https://doi.org/10.1016/j.jtbi.2018.05.006
33. Chowdhury S.Y., Shatabda S., Dehzangi A. iDNAProt-ES: Identification of DNA-binding proteins using evolutionary and structural features // Scientific Reports. 2017. V. 7. P. 14938. https://doi.org/10.1038/s41598-017-14945-1
34. Kumar R.D., Searleman A.C., Swamidass S.J., Griffith O.L., Bose R. Statistically identifying tumor suppressors and oncogenes from pan-cancer genome-sequencing data // Bioinformatics. 2015. V. 31. N 22. P. 3561–3568. https://doi.org/10.1093/bioinformatics/btv430
35. Akmal M.A., Hussain W., Rasool N., Khan Y.D., Khan S.A., Chou K.-C. Using CHOU'S 5-steps rule to predict O-linked serine glycosylation sites by blending position relative features and statistical moment // IEEE/ACM Transactions on Computational Biology and Bioinformatics. 2021. V. 18. N 5. P. 2045–2056. https://doi.org/10.1109/TCBB.2020.2968441
36. Khan Y.D., Ahmad F., Anwar M.W. Aneuro-cognitive approach for iris recognition using back propagation // World Applied Sciences Journal. 2012. V. 16. N 5. P. 678–685.
37. Khan Y.D., Ahmed F., Khan S.A. Situation recognition using image moments and recurrent neural networks // Neural Computing and Applications. 2014. V. 24. N 7–8. P. 1519–1529. https://doi.org/10.1007/s00521-013-1372-4
38. Khan Y.D., Khan N.S., Farooq S., Abid A., Khan S.A., Ahmad F., Mahmood M.K. An efficient algorithm for recognition of human actions // Scientific World Journal. 2014. V. 2014. P. 875879. https://doi.org/10.1155/2014/875879
39. Khan Y.D., Khan S.A., Ahmad F., Islam S. Iris recognition using image moments and K-means algorithm // Scientific World Journal. 2014. V. 2014. P. 723595. https://doi.org/10.1155/2014/723595
40. Mahmood S., Khan Y.D., Mahmood M.K. A treatise to vision enhancement and color fusion techniques in night vision devices // Multimedia Tools and Applications. 2018. V. 77. N 2. P. 2689–2737. https://doi.org/10.1007/s11042-017-4365-y
41. Butt H., Rasool N., Khan Y.D. A treatise to computational approaches towards prediction of membrane protein and its subtypes // The Journal of Membrane Biology. 2017. V. 250. N 1. P. 55–76. https://doi.org/10.1007/s00232-016-9937-7
42. Akmal M.A., Rasool N., Khan Y.D. Prediction of N-linked glycosylation sites using position relative features and statistical moments // PLoS ONE. 2017. V. 12. N 8. P. 1–21. https://doi.org/10.1371/journal.pone.0181966
43. Pundir S., Magrane M., Martin M.J., O'Donovan C. Searching and navigating UniProt databases // Current Protocols in Bioinformatics. 2015. P. 1.27.1–1.27.10 https://doi.org/10.1002/0471250953.bi0127s50
44. Delorenzi M., Speed T. An HMM model for coiled-coil domains and a comparison with PSSM-based predictions // Bioinformatics. 2002. V. 18. N 4. P. 617–625. https://doi.org/10.1093/bioinformatics/18.4.617
45. Jia J., Liu Z., Xiao X., Liu B., Chou K.-C. iSuc-PseOpt: identifying lysine succinylation sites in proteins by incorporating sequence-coupling effects into pseudo components and optimizing imbalanced training dataset // Analytical Biochemistry. 2016. V. 497. P. 48–56. https://doi.org/10.1016/j.ab.2015.12.009

Authors

Manickam Vijayalakshmi — Research Scholar, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Assistant Professor, Sri Sarada College for Women, Tirunelveli, 627011, India, https://orcid.org/0009-0001-2012-3169, vijimarthresearch@gmail.com
Mahesh Vallinayagi — PhD, Head, Associate Professor, Sri Sarada College for Women, Tirunelveli, 627011, India, https://orcid.org/0009-0006-0552-0138, vallinayagimahesh@gmail.com

Received 27.09.2023
Approved after reviewing 04.01.2024
Accepted 30.01.2024

Авторы

Виджаялакшми Маникам — научный сотрудник, Университет Манонманиам Сундаранар, Абишекапати, Тируневелли-627012; доцент, Женский колледж Шри Сарада, Тирунелвели, 627011, Индия, https://orcid.org/0009-0001-2012-3169, vijimarthresearch@gmail.com
Валлинаяги Махеш — PhD, руководитель, доцент, Женский колледж Шри Сарада, Тирунелвели, 627011, Индия, https://orcid.org/0009-0006-0552-0138, vallinayagimahesh@gmail.com

Статья поступила в редакцию 27.09.2023
Одобрена после рецензирования 04.01.2024
Принята к печати 30.01.2024

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2024-24-1-112-117
УДК 004.021

Метод хранения векторных представлений в сжатом виде с применением кластеризации

Никита Андреевич Томилов[1], Владимир Павлович Туров[2], Александр Амаякович Бабаянц[3]✉, Алексей Владимирович Платонов[4]

[1,2,3,4] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] programmer174@icloud.com, https://orcid.org/0000-0001-9325-0356
[2] firemoon@icloud.com, https://orcid.org/0009-0009-1470-7633
[3] babayants.alexander@gmail.com✉, https://orcid.org/0009-0004-6367-6398
[4] avplatonov@itmo.ru, https://orcid.org/0000-0002-8485-1296

Аннотация

Введение. Алгоритмы машинного обучения для информационного поиска позволяют представить текстовые и мультимодальные документы в виде векторов. Такие векторные представления (embeddings) сохраняют семантическое содержание документов и сводят задачу поиска к задаче определения расстояния между векторами. Сжатие векторных представлений позволяет уменьшить объем памяти, занимаемый ими, и повысить эффективность вычислений. В работе рассмотрены существующие способы сжатия векторных представлений без потери и с потерей точности. Предложен метод уменьшения ошибки путем кластеризации векторных представлений при использовании сжатия с потерей точности. **Метод.** Сущность метода состоит в предварительной кластеризации векторных представлений, сохранении центров каждого кластера и значений координат каждого векторного представления относительно центра его кластера. Центры каждого кластера сжимаются без потери точности, а получившиеся смещенные векторные представления с потерей точности. **Основные результаты.** Предложенный метод протестирован на наборах данных fashion-mnist-784-euclidean и NYT-256-angular. Проведено сравнение векторных представлений, сжатых с потерей точности при помощи уменьшения разрядности, с векторными представлениями, сжатыми по предложенному методу. При незначительном, около 10 %, увеличении размера сжатых данных средняя абсолютная величина ошибки от потери точности для наборов fashion-mnist-784-euclidean и NYT-256-angular снизилась в четыре и примерно в два раза соответственно. **Обсуждение.** Разработанный метод может быть применен для решения задач хранения и обработки векторных представлений мультимодальных документов, например, при разработке поисковых систем.

Ключевые слова

векторные представления, эмбеддинг, кластеризация, k-means, сжатие векторных представлений

Ссылка для цитирования: Томилов Н.А., Туров В.П., Бабаянц А.А., Платонов А.В. Метод хранения векторных представлений в сжатом виде с применением кластеризации // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 112–117. doi: 10.17586/2226-1494-2024-24-1-112-117

A method of storing vector data in compressed form using clustering

Nikita A. Tomilov[1], Vladimir P. Turov[2], Alexander A. Babayants[3]✉, Alexey V. Platonov[4]

[1,2,3,4] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] programmer174@icloud.com, https://orcid.org/0000-0001-9325-0356
[2] firemoon@icloud.com, https://orcid.org/0009-0009-1470-7633
[3] babayants.alexander@gmail.com✉, https://orcid.org/0009-0004-6367-6398
[4] avplatonov@itmo.ru, https://orcid.org/0000-0002-8485-1296

Abstract

The development of the machine learning algorithms for information search in recent years made it possible to represent text and multimodal documents in the form of vectors. These vector representations (embeddings) preserve the semantic

© Томилов Н.А., Туров В.П., Бабаянц А.А., Платонов А.В., 2024

content of documents and allow the search to be performed as the calculation of distance between vectors. Compressing embeddings can reduce the amount of memory they occupy and improve computational efficiency. The article discusses existing methods for compressing vector representations without loss of accuracy and with loss of accuracy. A method is proposed to reduce error by clustering vector representations using lossy compression. The essence of the method is in performing the preliminary clustering of vector representations, saving the centers of each cluster, and saving the coordinate value of each vector representation relative to the center of its cluster. Then, the centers of each cluster are compressed without loss of accuracy, and the resulting shifted vector representations are compressed with loss of accuracy. To restore the original vector representations, the coordinates of the center of the corresponding cluster are added to the coordinates of the displaced representation. The proposed method was tested on the fashion-mnist-784-euclidean and NYT-256-angular datasets. A comparison has been made of compressed vector representations with loss of accuracy by reducing the bit depth with vector representations compressed using the proposed method. With a slight (around 10 %) increase in the size of the compressed data, the absolute value of the error from loss of accuracy decreased by four and two times, respectively, for the tested sets. The developed method can be applied in tasks where it is necessary to store and process vector representations of multimodal documents, for example, in the development of search engines.

Keywords

vector representations, embeddings, clustering, k-means, compression of vector representations

For citation: Tomilov N.A., Turov V.P., Babayants A.A., Platonov A.V. A method of storing vector data in compressed form using clustering. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 112–117 (in Russian). doi: 10.17586/2226-1494-2024-24-1-112-117

Введение

В современном мире стремительно растет количество генерируемых данных, как в текстовом формате (статьи, публикации), так и в прочих форматах (изображения, звук). Соответственно, возникает необходимость быстрого поиска информации по этим данным. В последнее время наибольшую популярность получают алгоритмы поиска с применением машинного обучения, способные формировать векторные представления, позволяющие отображать семантическую составляющую текстовых [1] или мультимодальных документов в виде векторов (эмбеддингов). Далее процесс информационного поиска сводится к проведению операций над полученными векторами [2]. Для хранения векторных представлений в неизмененном виде часто требуется большой объем дискового пространства и памяти. С целью уменьшения размера векторных представлений используются различные алгоритмы сжатия и кодирования чисел. Наиболее популярным из таких алгоритмов является квантизация — пересчет каждого из значений в представление с меньшей точностью. Например, из FP32 (число с плавающей точкой 32-битной разрядности) в представление FP16 или FP8 (число с плавающей точкой 16-битной или 8-битной разрядности), и дальнейшее хранение и использование FP16 или FP8 [3]. Такой алгоритм является достаточно эффективным и позволяет кратно уменьшить размер векторного представления. Недостаток алгоритма — очевидная потеря точности, которая может влиять при дальнейшем использовании данных эмбеддингов, например, для их использования в моделях машинного обучения [4].

В настоящей работе предложен алгоритм сжатия векторных представлений, при котором вместо хранения исходных векторных представлений производится их кластеризация и вычисление центроидов кластеров. В результате каждое векторное представление может быть сохранено как пара из индекса кластера и компоненты разницы до центроида кластера (дельты), и затем записано на диск в сжатом виде. Данный метод позволяет сохранять центроиды и дельты с разной потерей точности, например: сжатие центроидов без потерь и сжатие дельт с пересчетом из FP32 в FP16 или FP8. Выдвигаемая гипотеза заключается в том, что это позволит снизить потерю точности относительно преобразования исходных векторных представлений из FP32 в FP16 или FP8.

Описание алгоритма

В работе предложен метод хранения и сжатия векторных представлений с дополнительной информацией. Каждое векторное представление, являющееся массивом чисел с плавающей точкой 32-битной разрядности, характеризуется, помимо непосредственно значений массива чисел, первичным и вторичным ключами. Такой подход расширяет и уточняет индексацию данных и позволяет, например, при хранении векторных представлений текстовых документов, указывать в качестве первичного ключа — идентификатор текстового документа, а в качестве вторичного ключа — идентификатор отрывка, например, номер предложения, при условии, что необходимо хранить и использовать именно эмбеддинги предложений [5].

Для хранения формируются два файла, которые далее помещаются в хранилище. Первый файл хранит непосредственно векторное представление в двоичном формате (N — количество кортежей, с одинаковым первичным ключом; кортежи, список длиной N). Каждый кортеж хранит информацию отдельного векторного представления, т. е. (вторичный ключ; длину массива; массив байт). Второй файл реализует индекс и хранит соответствие «первичный ключ — смещение в первом файле, начиная с которого находятся все векторные представления, относящиеся к этому первичному ключу».

Перед сохранением в файл данные каждого векторного представления кодируются и сжимаются. Рассмотрим три вида кодирования.

— FP32 (float32) — кодирование без потерь точности. Каждый компонент векторного представления занимает 4 Б.

— FP16 (float16) — кодирование с потерей точности. Каждый компонент векторного представления преобразуется в формат с плавающей запятой 16-битной разрядности. Занимает в два раза меньше памяти.
— FP8 (float8/minifloat) — кодирование с потерей точности. Каждый компонент векторного представления преобразуется в формат с плавающей запятой 8-битной разрядности. Занимает в четыре раза меньше памяти.

После кодирования каждый из массивов байт сжимается одним из трех алгоритмов и форматов сжатия — ZLib[1], LZMA, XZ [6]. Все алгоритмы сжатия сконфигурированы такими параметрами, при которых предполагается достижение наибольшего коэффициента сжатия. Все целочисленные значения, а именно, первичные и вторичные ключи и смещения кодируются с применением кодирования переменной длины, что позволяет, несмотря на использование 4- и 8-байтных типов данных, использовать только необходимое число байт (например, числа до 127 кодируются одним байтом).

Такой способ хранения предоставляет возможность быстрого доступа к любому векторному представлению без необходимости распаковывать избыточные данные. Данная организация хранения позволяет создать потокобезопасную и быструю реализацию алгоритма перебора (аналогичную операции «full table scan» [7] для реляционных систем управления базами данных), необходимую для выполнения векторного поиска по всем векторным представлениям или для их экспорта в другие хранилища. Помимо этого, доступен выбор любой комбинации алгоритмов сжатия и кодирования векторных представлений.

Для проверки гипотезы о применимости кластеризации и хранении векторных представлений как дельт относительно центроидов кластеров, рассматриваемый алгоритм был дополнен следующими условиями. Все векторные представления с одинаковым первичным ключом кластеризуются с помощью алгоритма кластеризации k-means [8] до указанного (параметром N) числа кластеров, и каждому векторному представлению ставится в соответствие индекс кластера, к которому он принадлежит.

На основе всех доступных данных создано два хранилища. В первое хранилище помещены центроиды кластеров, у которых первичным ключом является собственный индекс кластера, а вторичным ключом — ноль. Во второе хранилище записаны дельты — расстояние между векторным представлением и центроидом кластера, к которому он относится, с сохранением первичного и вторичного ключей. Связь между хранилищами обеспечивает дополнительный служебный файл, который хранит кортежи (первичный и вторичный ключи, индекс кластера). Для извлечения оригинального векторного представления из описанной структуры необходимо произвести следующие операции.

— Из хранилища дельт извлечь дельту, соответствующую первичному и вторичному ключам.
— Из файла соответствий получить индекс кластера, к которому относится необходимое векторное представление.
— Из хранилища центроидов извлечь необходимый центроид кластера, к которому относится требуемое векторное представление.
— Рассчитать сумму центроида кластера и дельты и получить искомое векторное представление.

Такой подход позволяет выполнить сохранение центроидов и дельт кластеров с разными настройками сжатия. В результате возможно увеличить точность сохранения векторных представлений при сжатии с потерями при одинаковых настройках сжатия исходных векторных представлений и их дельт ценой небольшого дискового пространства, необходимого для хранения самих центроидов кластеров.

Описание эксперимента

Для проверки предложенного алгоритма было выбрано два тестовых набора данных, fashion-mnist-784-euclidean [9] и NYT-256-angular[2]. Данный выбор наборов данных обусловлен их общедоступностью, частотой использования в тестировании алгоритмов машинного обучения, а также тем, что наборы сильно отличаются друг от друга по семантике закодированных данных, диапазону значений и метрике расстояния. Это условия позволяют оценить работу алгоритма на разных входных данных.

Оба набора данных были преобразованы в форматы, соответствующие описанным хранилищам в разделе «Описание алгоритма». Для каждого полученного хранилища выполним замеры занимаемого места на диске и метрики ошибки. Метрика ошибки — расстояние между векторным представлением, сохраненным и затем извлеченным из хранилища, и исходным. При этом расстояние может быть евклидовым или угловым и выбирается в зависимости от используемого набора данных. Обозначив векторное представление из хранилища, заданное набором компонентов $(vc_0, \ldots, vc_n) - vc$; а исходное векторное представление $(vs_0, \ldots, vs_n) - vs$, получим следующие формулы для расчета метрик ошибок при использовании:

— евклидова расстояния

$$e = \sqrt{\sum_{k=0}^{n} (vc_k - vs_k)^2};$$

— углового расстояния

$$e = \frac{\sum_{k=0}^{n} vc_k vs_k}{\sqrt{\sum_{k=0}^{n} vc_n^2} \sqrt{\sum_{k=0}^{n} vs_n^2}}.$$

[1] Gailly J., Adler M. Zlib compression library [Электронный ресурс]. Режим доступа: https://www.repository.cam.ac.uk/items/248f35af-3d9c-40a4-8a70-52520d274a47, свободный. Яз. англ. (дата обращения: 28.05.2023).

[2] Dua D. and Graff C. UCI Machine Learning Repository. 2019. [Электронный ресурс]. Режим доступа: https://archive.ics.uci.edu/ml/datasets/bag+of+words, свободный. Яз. англ. (дата обращения: 30.05.2023).

Вычислим метрики для каждого из векторных представлений и запишем их минимальное, максимальное и медианное значение. Чем меньше полученное значение ошибки, тем ближе векторные представления друг к другу и тем меньше потеря точности при сохранении в хранилище.

Для сравнения с эталонным примером хранения данных, векторные представления из обоих наборов преобразуем в базы данных SQLite3 [10] с сохранением в таблицу в виде массивов байт после применения алгоритмов сжатия ZLib, LZMA, XZ. Отметим, что для замера места на диске был включен оригинальный файл набора данных в формате HDF5, который может дать «порядок» размера оригинальных данных несмотря на то что, кроме использованных векторных представлений (взятых из «обучающей» выборки наборов данных), содержит еще и «проверочную» выборку наборов данных.

Тестовый стенд имеет следующие характеристики: AMD Ryzen 7 7700X 8C16T; 64GB RAM; NVMe WD SN850X 2TB; операционная система Ubuntu 22.04; OpenJDK 17; разработанный в рамках предыдущих исследований инструмент сравнения алгоритмов векторного поиска [11].

Заметим, что измерений времени извлечения единичного векторного представления и времени выполнения полного перебора не производилось. Это связано с тем, что использованные реализации алгоритмов сжатия LZMA и XZ из пакета Apache Commons Compress задействовали промежуточные преобразования массивов байт, ведущие к большим накладным расходам.

Результаты эксперимента

Первый набор данных, fashion-mnist-784-euclidean, содержит 60 000 векторных представлений изображений различных предметов одежды. Каждое векторное представление состоит из 784 компонентов [9]. В силу общедоступности, простоты анализа и не очень большого размера, этот набор данных получил широкое распространение как один из эталонов для тестирования алгоритмов машинного обучения [12]. Для работы с первым набором данных использована метрика евклидова расстояния. Каждый компонент векторного представления кодирует яркость соответствующего пиксела числом в диапазоне от 0 до 255. Данный набор векторных представлений содержит неплотные (разреженные) вектора, содержащие большое количество нулевых компонентов.

Размеры файлов хранилищ представлены в табл. 1. При использовании формата FP32 ошибки нет, так как исходные данные были представлены в этом же формате — потери точности не происходит. При использовании формата FP8 наблюдается очень большая потеря точности. При этом, если сравнивать кодирование в FP8 без кластеризации и с кластеризацией, то можно отметить, что хранение центроидов кластеров значительно снижает ошибку, хотя и не устраняет ее полностью. Однако, в связи с тем, что набор данных содержит значения в ограниченном диапазоне от 0 до 255, представляется целесообразным хранить их как 8-битные целые числа. При условии, что все значения векторных представлений оригинального набора данных фактически целочисленные (хотя и хранятся как FP32), при преобразовании в INT8 и сжатия с помощью кодирования переменной длины [13] потерь точности не происходит, и метрика потери точности для всех полученных хранилищ равна нулю. Тем не менее, эти результаты показывают, насколько метод кодирования с кластеризацией может улучшать точность при использовании форматов с малым диапазоном значений (FP8), а также что дельты сжимаются хуже, чем оригинальные векторные представления той же размерности. Также из этих результатов можно сделать вывод о том, что применение предложенного алгоритма с хранением

Таблица 1. Размер на диске и ошибка (по метрике евклидова расстояния) для набора данных fashion-mnist-784-euclidean, сохраненного с использованием разных алгоритмов сжатия

Table 1. Disk size (in megabytes) and error (as euclidean distance) for the fashion-mnist-784-euclidean dataset, saved with different compression algorithms

Хранилище	Тип данных	Размер на диске при алгоритме сжатия, МБ				Метрика ошибки (евклидово расстояние)		
		без сжатия	ZLib	XZ	LZMA	среднее	медиана	максимум
Оригинальный файл HDF5	FP32	227,5	—	—	—	нет		
База данных SQLite3	FP32	235,7	46,1	40,6	37,5	нет		
Бинарное без кластеризации	FP32	180,1	40,8	36,2	33,6	нет		
	FP16	89,7	33,6	40,6	38,0	0,001	0,001	0,013
	FP8	45,4	25,5	27,7	24,6	4750,7	4568,5	16 362,0
Бинарное, 1000 кластеров	FP32	183,0	111,2	110,0	107,3	нет		
	FP16	93,2	67,5	70,2	67,5	0,045	0,039	0,327
	FP8	51,0	24,6	24,0	25,1	656,9	569,8	4318,4
Бинарное, 2000 кластеров	FP32	186,2	108,6	106,4	103,7	нет		
	FP16	96,3	67,5	70,6	67,9	0,037	0,031	0,250
	FP8	50,9	33,1	34,1	31,0	598,5	509,0	4293,3

Таблица 2. Размер на диске и ошибка (по метрике углового расстояния) для набора данных NYT-256-angular, сохраненного с использованием разных алгоритмов сжатия

Table 2. Disk size (in megabytes) and error (as angular distance) for the NYT-256-angular dataset, saved with different compression algorithms

Хранилище	Тип данных	Размер на диске при алгоритме сжатия, МБ				Метрика ошибки (угловое расстояние)		
		без сжатия	ZLib	XZ	LZMA	среднее	медиана	максимум
Оригинальный файл HDF5	FP32	300,6	—	—	—	нет		
База данных SQLite3	FP32	382,6	288,3	382,3	382,3	нет		
Бинарное без кластеризации	FP32	286,1	280,2	300,1	287,5	нет		
	FP16	144,1	145,9	161,1	150,4	0,0001	0,0001	0,0001
	FP8	73,0	53,9	80,7	68,0	0,0011	0,0007	0,0221
Бинарное, 1000 кластеров	FP32	289,4	282,8	302,9	290,2	нет		
	FP16	146,7	148,7	163,9	153,2	0,0001	0,0001	0,0001
	FP8	77,4	71,0	77,0	72,1	0,0007	0,0004	0,0199
Бинарное, 2000 кластеров	FP32	290,2	282,9	303,0	290,3	нет		
	FP16	147,9	149,2	164,6	153,8	0,0001	0,0001	0,0001
	FP8	77,4	72,0	73,4	71,1	0,0007	0,0004	0,0199

векторов в двоичных файлах занимает на несколько мегабайт меньше памяти, чем хранение векторов в SQLite при использовании одинаковых алгоритмов сжатия.

Второй набор данных, NYT-256-angular [10], содержит 290 тысяч векторных представлений текстов статей газеты New York Times. Каждое векторное представление состоит из 256 компонентов, для работы с этим набором данных использована метрика углового (angular) расстояния. Второй набор данных был выбран для тестирования, так как он сильно отличается по содержанию и диапазону значений от набора fashion-mnist-784-euclidean, а также использует другую метрику расстояния. В значениях компонентов каждого из векторных представлений практически нет нулей, а все значения лежат в диапазоне от [–1; 1].

Объемы файлов хранилищ, а также вычисленные расстояния (показатель ошибки) представлены в табл. 2. Аналогично первому набору данных, при использовании формата FP32 ошибки нет, исходные данные были представлены в этом же формате и потери точности не происходит. Для второго набора данных практически отсутствует разница в эффективности сжатия дельт по сравнению с эффективностью сжатия исходных векторных представлений, что не наблюдалось для набора данных fashion-mnist-784-euclidean с повторяющимися нулевыми компонентами в векторных представлениях. При сжатии через кластеризацию для одинаковых настроек хранилище, использующее кластеризацию, занимает на фиксированное число мегабайт больше дискового пространства, чем хранилище без кластеризации, а разница обусловлена именно хранением центроидов кластеров; однако эта разница небольшая и в худшем случае центроиды занимают около 4 МБ. Из данных результатов можно сделать вывод что использование кластеризации для тысячи кластеров (и хранении центроидов кластеров в FP32 без потерь сжатия) позволяет уменьшить значения метрики ошибки, что подтверждает заявленную гипотезу.

Заключение

При хранении векторных представлений применение универсальных алгоритмов сжатия без потерь, а именно ZLIB, LZMA и XZ, оправданно и позволяет существенно снизить объем занимаемого дискового пространства для хранения. При работе с разреженными векторными представлениями предложенный и описанный алгоритм хранения векторных представлений с использованием кодирования FP8 позволяет на порядок уменьшить занимаемое дисковое пространство для хранения. Также происходит уменьшение пространства при сравнении хранения векторных представлений в SQLite и с применением универсальных алгоритмов сжатия без потерь. В частности, для набора данных fashion-mnist-784-euclidean векторные представления, закодированные FP8 и сжатые с помощью ZLIB, используют 25,5 МБ дискового пространства, что почти в два раза меньше, чем сжатие оригинальных векторных представлений с помощью ZLIB (40,8 МБ), в 60 раз меньше, чем использование баз данных SQLite3 без сжатия (236 МБ) и на несколько мегабайт меньше, чем — SQLite3 со сжатием. При работе с плотными векторными представлениями подход с использованием кластеризации позволил увеличить точность сохранения векторных представлений при использовании сжатия с потерями, ценой незначительного увеличения дискового пространства за счет дополнительного хранения центроидов кластеров. На протестированных наборах NYT-256-angular и fashion-mnist-784-euclidean при использовании дополнительного хранения центроидов кластеров метрика ошибки уменьшилась в два и четыре раза соответственно.

Литература

1. Grbovic M., Cheng H. Real-time personalization using embeddings for search ranking at airbnb // Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. 2018. P. 311–320. https://doi.org/10.1145/3219819.3219885
2. Berry M.W., Drmac Z., Jessup E.R. Matrices, vector spaces, and information retrieval // SIAM Review. 1999. V. 41. N 2. P. 335–362. https://doi.org/10.1137/S0036144598347035
3. Micikevicius P., Narang S., Alben J., Diamos G., Elsen E., Garcia D., Ginsburg B., Houston M., Kuchaiev O., Venkatesh G., Wu H. Mixed precision training // arXiv. 2017. arXiv:1710.03740. https://doi.org/10.48550/arXiv.1710.03740
4. Zhang J., Yang J., Yuen H. Training with low-precision embedding tables // Systems for Machine Learning Workshop at NeurIPS. 2018. V. 2018.
5. Moghadasi M.N., Zhuang Y. Sent2Vec: A new sentence embedding representation with sentimental semantic // Proc. of the 2020 IEEE International Conference on Big Data (Big Data). 2020. P. 4672–4680. https://doi.org/10.1109/BigData50022.2020.9378337
6. Parekar P.M., Thakare S.S. Lossless data compression algorithm–a review // International Journal of Computer Science & Information Technologies. 2014. V. 5. N 1. P. 276–278.
7. Manro A., Sinha S., Chaturvedi B., Mohan J. Index seek versus table scan performance and implementation of RDBMS // Lecture Notes in Electrical Engineering. 2019. V. 526. P. 411–420. https://doi.org/10.1007/978-981-13-2553-3_40
8. Kanungo T., Mount D., Netanyahu N., Piatko Ch., Silverman R., Wu A. The analysis of a simple k-means clustering algorithm // Proc. of the sixteenth annual symposium on Computational geometry. 2000. P. 100–109. https://doi.org/10.1145/336154.336189
9. Xiao H., Rasul K., Vollgraf R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms // arXiv. 2017. arXiv:1708.07747. https://doi.org/10.48550/arXiv.1708.07747
10. Bi C., China P.R. Research and application of SQLite embedded database technology // WSEAS Transactions on Computers. 2009. V. 8. N 1. P. 83–92.
11. Туров В.П., Томилов Н.А., Бабаянц А.А. Разработка инструмента сравнения алгоритмов векторного поиска // XI Конгресс молодых учёных: сборник научных трудов. Т. 1. 2022. С. 446–450.
12. Aumüller M., Bernhardsson E., Faithfull A. ANN-benchmarks: A benchmarking tool for approximate nearest neighbor algorithms // Lecture Notes in Computer Science. 2017. V. 10609. P. 34–49. https://doi.org/10.1007/978-3-319-68474-1_3
13. Golomb S. Run-length encodings (Corresp.) // IEEE Transactions on Information Theory. 1966. V. 12. N 3. P. 399–401. https://doi.org/10.1109/TIT.1966.1053907

References

1. Grbovic M., Cheng H. Real-time personalization using embeddings for search ranking at airbnb. *Proc. of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 2018, pp. 311–320. https://doi.org/10.1145/3219819.3219885
2. Berry M.W., Drmac Z., Jessup E.R. Matrices, vector spaces, and information retrieval. *SIAM Review*, 1999, vol. 41, no. 2, pp. 335–362. https://doi.org/10.1137/S0036144598347035
3. Micikevicius P., Narang S., Alben J., Diamos G., Elsen E., Garcia D., Ginsburg B., Houston M., Kuchaiev O., Venkatesh G., Wu H. Mixed precision training. *arXiv*, 2017, arXiv:1710.03740. https://doi.org/10.48550/arXiv.1710.03740
4. Zhang J., Yang J., Yuen H. Training with low-precision embedding tables. *Systems for Machine Learning Workshop at NeurIPS*, 2018, vol. 2018.
5. Moghadasi M.N., Zhuang Y. Sent2Vec: A new sentence embedding representation with sentimental semantic. *Proc. of the 2020 IEEE International Conference on Big Data (Big Data)*, 2020, pp. 4672–4680. https://doi.org/10.1109/BigData50022.2020.9378337
6. Parekar P.M., Thakare S.S. Lossless data compression algorithm–a review. *International Journal of Computer Science & Information Technologies*, 2014, vol. 5, no. 1, pp. 276–278.
7. Manro A., Sinha S., Chaturvedi B., Mohan J. Index seek versus table scan performance and implementation of RDBMS. *Lecture Notes in Electrical Engineering*, 2019, vol. 526, pp. 411–420. https://doi.org/10.1007/978-981-13-2553-3_40
8. Kanungo T., Mount D., Netanyahu N., Piatko Ch., Silverman R., Wu A. The analysis of a simple k-means clustering algorithm. *Proc. of the sixteenth annual symposium on Computational geometry*, 2000, pp. 100–109. https://doi.org/10.1145/336154.336189
9. Xiao H., Rasul K., Vollgraf R. Fashion-MNIST: a novel image dataset for benchmarking machine learning algorithms. *arXiv*, 2017, arXiv:1708.07747. https://doi.org/10.48550/arXiv.1708.07747
10. Bi C., China P.R. Research and application of SQLite embedded database technology. *WSEAS Transactions on Computers*, 2009, vol. 8, no. 1, pp. 83–92.
11. Turov V.P., Tomilov N.A., Babaiantc A.A. Development of a tool for comparing vector search algorithms. *Proc. of the XI Congress of Young Scientists. V. 1*. 2022, pp. 446–450. (in Russian)
12. Aumüller M., Bernhardsson E., Faithfull A. ANN-benchmarks: A benchmarking tool for approximate nearest neighbor algorithms. *Lecture Notes in Computer Science*, 2017, vol. 10609, pp. 34–49. https://doi.org/10.1007/978-3-319-68474-1_3
13. Golomb S. Run-length encodings (Corresp.). *IEEE Transactions on Information Theory*, 1966, vol. 12, no. 3, pp. 399–401. https://doi.org/10.1109/TIT.1966.1053907

Авторы

Томилов Никита Андреевич — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, SC 57225127284, https://orcid.org/0000-0001-9325-0356, programmer174@icloud.com

Туров Владимир Павлович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0009-1470-7633, firemoon@icloud.com

Бабаянц Александр Амаякович — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0004-6367-6398, babayants.alexander@gmail.com

Платонов Алексей Владимирович — кандидат технических наук, доцент, доцент, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0002-8485-1296, avplatonov@itmo.ru

Authors

Nikita A. Tomilov — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, SC 57225127284, https://orcid.org/0000-0001-9325-0356, programmer174@icloud.com

Vladimir P. Turov — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0009-1470-7633, firemoon@icloud.com

Alexander A. Babayants — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0004-6367-6398, babayants.alexander@gmail.com

Alexey V. Platonov — PhD, Associate Professor, Associate Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0002-8485-1296, avplatonov@itmo.ru

Статья поступила в редакцию 08.06.2023
Одобрена после рецензирования 10.01.2024
Принята к печати 29.01.2024

Received 08.06.2023
Approved after reviewing 10.01.2024
Accepted 29.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-118-123

Monocular depth estimation for 2D mapping of simulated environments

Majd Barhoum[1], Anton A. Pyrkin[2]✉

[1,2] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] barhoum.majd213@gmail.com, https://orcid.org/0009-0007-8022-7932
[2] a.pyrkin@gmail.com✉, https://orcid.org/0000-0001-8806-4057

Abstract

This article addresses the problem of constructing maps for 2D simulated environments. An algorithm based on monocular depth estimation is proposed achieving comparable accuracy to methods utilizing expensive sensors such as RGBD cameras and LIDARs. To solve the problem, we employ a multi-stage approach. First, a neural network predicts a relative disparity map from an RGB flow provided by RGBD camera. Using depth measurements from the same camera, two parameters are estimated that connect the relative and absolute displacement maps in the form of a linear regression relation. Based on a simpler RGB camera, by applying a neural network and estimates of scaling parameters, an estimate of the absolute displacement map is formed, which allows to obtain an estimate of the depth map. Thus, a virtual scanner has been designed providing Cartographer SLAM with depth information for environment mapping. The proposed algorithm was evaluated on a ROS 2.0 simulation of a simple mobile robot. It achieves faster depth prediction compared to other depth estimation algorithms. Furthermore, maps generated by our approach demonstrated a high overlap ratio with those obtained using an ideal RGBD camera. The proposed algorithm can find applicability in crucial tasks for mobile robots, like obstacle avoidance, and path planning. Moreover, it can be used to generate accurate cost maps, enhancing safety and adaptability in mobile robot navigation.

Keywords

monocular depth estimation, mapping, linear regression, disparity maps, neural network

Acknowledgements

This paper was supported by the Ministry of Science and Higher Education of the Russian Federation (State Assignement No. 2019-0898).

For citation: Barhoum M., Pyrkin A.A. Monocular depth estimation for 2D mapping of simulated environments. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 118–123. doi: 10.17586/2226-1494-2024-24-1-118-123

УДК 004.896

Использование монокулярной оптики при оценке глубины объектов для двумерного картирования моделируемой среды

Мажд Бархум[1], Антон Александрович Пыркин[2]✉

[1,2] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] barhoum.majd213@gmail.com, https://orcid.org/0009-0007-8022-7932
[2] a.pyrkin@gmail.com✉, https://orcid.org/0000-0001-8806-4057

Аннотация

Введение. Рассмотрена задача построения карты двумерной среды. Предложен алгоритм оценки на основе монокулярной оптики и RGB-изображений. Алгоритм позволяет получать результаты, сопоставимые с подходами на основе дорогостоящих датчиков, таких как RGBD-камеры и лидары. **Метод.** Решение задачи включает нескольких этапов. На начальном этапе выполняется обучение нейронной сети, которая формирует относительную карту несоответствия (смещений) на основе входного потока RGB-изображений от RGBD-камеры. С использованием измерений глубин от той же камеры выполняется оценка двух параметров, связывающих относительную и абсолютную карты смещений в виде линейного регрессионного соотношения. На основе более простой RGB-камеры, путем применения нейронной сети и оценок масштабирующих параметров

© Barhoum M., Pyrkin A.A., 2024

формируется оценка абсолютной карты смещений, позволяющей получить оценку карты глубин. Таким образом, синтезирован виртуальный сканер, который формирует данные о глубине для построения карты окружающей среды. **Основные результаты.** Представленный алгоритм апробирован при моделировании движения мобильного робота в среде ROS 2.0. Удалось достичь более быстрого прогнозирования глубины объектов по сравнению с другими алгоритмами оценки глубины. Карты, сгенерированные согласно разработанному алгоритму, продемонстрировали высокую степень совпадения с картами, полученными с помощью идеальной RGBD-камеры. **Обсуждение.** Предложенный алгоритм может найти применение в ключевых задачах управления мобильными роботами, такими как избегание препятствий и планирование пути. Алгоритм может быть использован при разметке карт по областям с различной степенью сложности прохождения, повышая безопасность и адаптивность навигации мобильных роботов.

Ключевые слова
монокулярная оценка глубины, картографирование, линейная регрессия, карты несоответствия (смещений), нейронные сети

Благодарности
Работа выполнена при поддержке Министерства науки и высшего образования Российской Федерации (госзадание 2019-0898).

Ссылка для цитирования: Бархум М., Пыркин А.А. Использование монокулярной оптики при оценке глубины объектов для двумерного картирования моделируемой среды // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 118–123 (на англ. яз.). doi: 10.17586/2226-1494-2024-24-1-118-123

Introduction

In the realm of robotics, an accurate map plays a pivotal role in the robot's interaction with the surrounding world, granting the robot the ability to navigate seamlessly through obstacle filled environments. To that end, Simultaneous Localization and Mapping (SLAM) algorithms are used to enable robots to keep track of their own position and the position of all obstacles around them utilizing the robot's on-board sensors. Visual SLAM algorithms build 3D maps of the environment using multiple cameras or specialized depth sensors. However, considering compact robotic systems used in cost-constrained applications such algorithms may not be practical. On the other hand, Monocular visual SLAM algorithms rely on a single camera to build 3D maps of the environment. Nevertheless, they suffer from scale ambiguity since only a single image with little to no depth information is used for both mapping and localization. 2D SLAM algorithms require less computational demand making them a convenient alternative to visual SLAM, provided a 2D map is sufficient for the problem at hand. But they still require depth information in the form of laser scans to build the surrounding map. Monocular Depth Estimation (MDE) offers the means to retrieve depth information without the dependence on sensors like LIDARS, or RGBD cameras. In the heart of it, the problem of MDE arises from the lack of depth cues in single images rendering conventional estimation approaches useless, but due to the importance of the problem recent advances showed the formidability of neural networks in predicting near perfect depth maps. The problem of MDE can be broken down into two separate sub-problems, namely relative depth estimation and metric depth estimation, the most common segment is the metric one, where the neural network trains to predict a depth map in metric space for situations similar to the datasets it trained on [1–5]. The other segment addresses the problem from a different angle, where instead of direct depth maps, disparity being the inverse depth is predicted, and the disparity of each pixel is only consistent relative to each other. This is achieved by training on a mixture of datasets with different camera models to increase the generalization of the model [6–8].

Integrating MDE with SLAM is an attempt to meet the surging demand for autonomous systems. That combines the power of artificial intelligence and conventional bundle adjustment techniques to improve the visual mapping process. The authors of [9] implemented a pseudo-RGBD SLAM system which used a Convolutional neural network to predict depth maps, and those depth maps are used to improve the accuracy of monocular SLAM. While the authors of [10] implement an unsupervised MDE pipeline that improves the accuracy of monocular SLAM and provides a fail-proof method for when ORB-SLAM does not match enough features, and significantly decreases the initialization time of the SLAM system. In this work, we tackle the integration of MDE as a sensor for 2D mapping using Cartographer SLAM, aiming at the balance between real-time mapping and accuracy we propose a simple yet effective algorithm to map simulated environments.

Monocular Depth Estimation

When choosing a neural network to perform the task of MDE, one needs to take into account the delicate balance between speed and accuracy by testing multiple networks and evaluating the trade-off between frames per second and performance, we reached the conclusion of working with a network with a low parameter count and good performance, even if the accuracy is lower than its peers. In our algorithm we chose one of the networks proposed by the authors of [6] where they trained multiple networks on a mixture of datasets with different input-output resolutions and showed how the improvement varies depending on the model trained and input image resolution. The **Swin2-T** network they trained can promise near real-time operation with acceptable accuracy provided a GPU with adequate capabilities is used when running the network. In their work, the architecture of the network has a smaller image size input 256 × 256 making the network more efficient to deploy. The architecture uses the pre-trained **Swin2-T** as an encoder where the transformer progresses through four

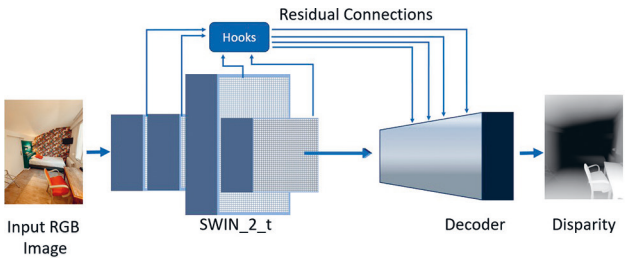


Fig. 1. Architecture of the network with Swin2-T transformer encoder and a decoder that uses the residual connection coming from the encoder stages to decode depth

stages each of which has a number **Swin2** blocks, the total number of these blocks is 12, divided as 2 in each stage except in the third one where there are 6 of them. The output of each stage is passed to the next stage and also hooked to the decoder which consists of 4 convolutional layers to map the encoded data into a depth map similar to input in size except that it has only one channel. The architecture is depicted in Fig. 1.

Relative to Metric Depth Maps

In their work, the authors of [8] propose a loss function to facilitate training on a mix of multiple datasets in disparity space, instead of the usual depth ground truth in online datasets. The loss function used to produce a prediction of relative disparity which is scale and shift-invariant, where, if $\mathbf{d}$ is the predicted disparity and $\mathbf{d}^*$ is the ground truth disparity, they define:

$$\hat{\mathbf{d}} = s\mathbf{d} + t,\ \hat{\mathbf{d}}^* = \mathbf{d}^*, \tag{1}$$

where, s, t from (1) can be determined as a least-squares problem for each pixel i from M as:

$$(s, t) = \operatorname{argmin}_{s,t} \sum_{i=1}^{M} (s\mathbf{d}_i + t - \mathbf{d}_i^*)^2. \tag{2}$$

And the loss for one image using $\hat{\mathbf{d}}$, $\mathbf{d}^*$ from (1) becomes:

$$\mathcal{L}(\hat{\mathbf{d}}, \hat{\mathbf{d}}^*) = \frac{1}{2M} \sum_{i=1}^{M} \operatorname{mae}(\hat{\mathbf{d}}_i - \hat{\mathbf{d}}_i^*). \tag{3}$$

They add to it the gradient loss to capture edges of objects in the image with the term proposed by [11] as:

$$\mathcal{L}_{reg}(\hat{\mathbf{d}}, \hat{\mathbf{d}}^*) = \frac{1}{M} \sum_{k=1}^{K} \sum_{i=1}^{M} (|\nabla_x R_i^k| + |\nabla_y R_i^k|, \tag{4}$$

where k refers to the index of scale level (they use 4), ∇_y, ∇_x are the gradients on x-axis and y-axis respectively, and $R_i = \hat{\mathbf{d}}_i - \mathbf{d}_i^*$. To adapt these losses to multiple datasets, they combine (3) and (4) to get the final loss function applied to the experiments.

$$\mathcal{L}_l = \frac{1}{N_l} \sum_{n=1}^{N_l} \mathcal{L}(\hat{\mathbf{d}}^n, \hat{\mathbf{d}}^{*n}) + \alpha \mathcal{L}_{reg}(\hat{\mathbf{d}}^n, \hat{\mathbf{d}}^{*n}),$$

where N_l is the size of the mixture of datasets in the training experiment, and α is 0.5.

Notation. Let vectors $\hat{\mathbf{D}} = col(\widehat{d_1}, \widehat{d_2}, \ldots, \widehat{d_M})$ be the predicted disparity map by the network, and $\mathbf{D} = col(d_1, d_2, \ldots, d_M)$ be the inverse of the ground truth depth obtained from the RGBD camera (images reshaped as vectors of the same length), $b \in \mathbb{R}$, $W \in \mathbb{R}$ is the shift between maps, and the scale respectively as an argument of (2), then we can define $W = \bar{s}$, $b = \bar{t}$ as average values of the scale and shift in our simulation, $\mathbf{E}$ is a vector column of ones of length M.

Retrieving the scale and shift of the disparity map can align it with the metric one as:

$$\mathbf{D} = [\hat{\mathbf{D}}\ \mathbf{E}] \begin{bmatrix} W \\ b \end{bmatrix}. \tag{5}$$

Equation (5) can be treated as a least squares problem then, there exist optimal values for W, b as:

$$[\hat{W}, \hat{b}]^T = \left([\hat{\mathbf{D}}\ \mathbf{E}]^T [\hat{\mathbf{D}}\ \mathbf{E}]\right)^{-1} \left([\hat{\mathbf{D}}\ \mathbf{E}]^T \mathbf{D}\right). \tag{6}$$

Then we reason that we can solve this problem using a linear regression model minimizing the square errors as:

$$L = \sum_{i=1}^{M} \left(\mathbf{D_i} - [\hat{\mathbf{D}}_\mathbf{i}\ \mathbf{E_i}] \begin{bmatrix} \hat{\mathbf{W}} \\ \hat{\mathbf{b}} \end{bmatrix} \right)^2.$$

Leaving the metric predicted depth to be denoted using the optimal values from (6) as:

$$\mathbf{D}^* = \frac{1}{(\hat{W}\hat{\mathbf{D}} + \hat{b})}.$$

Pseudo Laser Scanner

From the pinhole camera model with no distortion, the invertible projection equation can be written as follows:

$$\begin{bmatrix} u \\ v \\ 1 \\ 1/z \end{bmatrix} = \frac{1}{z} \mathbf{K}[\mathbf{R}|\mathbf{t}] \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \tag{7}$$

where u, v are the pixel indices in the image, x, y, z are the 3D coordinates of the point, and $\mathbf{K}[\mathbf{R}|\mathbf{t}]$ are the intrinsic and extrinsic parameters of the camera. Using equation (7) and the estimated depth, we can re-project the pixels in the image onto a 3D point cloud, but for 2D mapping only one row of the image is sufficient, thus we re-project all the pixels of the middle row of the image onto a 3D line; that mimics the work of a laser scanner as was done by the authors of [12]. We also introduce a limit on the range of the point cloud re-projected, since we compromised accuracy for speed when choosing the network. Not to mention that a smaller range corresponding to a smaller search window can prove to be easier to handle when mapping.

$$z = \begin{cases} \mathbf{D}^*, \textit{if } 0 < D^* \leq 1.5 \\ \Phi, \textit{otherwise} \end{cases}.$$

Cartographer SLAM

Cartographer is a popular SLAM system designed to create highly accurate 2D and 3D maps of environments

while localizing a robot within those maps. Cartographer SLAM uses the concept of drawing multiple sub-maps and then stitching them together to form the entire map of the environment, it draws the map as a grid and estimates the pose of the robot using a graph optimization algorithm called the CERES scan matching. The main Steps in this Graph-based SLAM system are the following:

1. Transform points from the scan frame which is the sensor frame into the sub-map frame.

$$T_{\xi P} = \begin{Bmatrix} \cos(\xi_\theta) & -\sin(\xi_\theta) \\ \sin(\xi_\theta) & \cos(\xi_\theta) \end{Bmatrix} P + \begin{Bmatrix} \xi_x \\ \xi_y \end{Bmatrix},$$

where the pose is described by the position x, y and the orientation θ.

2. Then a hit or miss VOXEL probabilistic filter is used to find the probability of a cell in the grid of the sub-map being occupied or not, based on the observed points using the laser scan. If a point was already observed its odds $odds(p) = \frac{p}{1-p}$ are updated as:

$$M = odds^{-1}(odds(M_{old})odds(p_{hit})).$$

3. CERES scan matching: before adding a scan to a sub-map, the pose of the scan, denoted as ξ, is optimized in relation to the current local sub-map using a scan matcher based on the Ceres library. The scan matcher finds the scan pose that maximizes the probabilities associated with the scan points within the sub-map.

$$\underset{\xi}{\operatorname{argmin}} \sum_{k=1}^{K} (1 - M_{smooth}(T_\xi h_k))^2.$$

4. Loop closure: it is formulated as a nonlinear least squares problem that allows adding residuals to take additional data into account. Every few seconds the CERES scan matching is applied to compute the nonlinear optimization problem.

$$\underset{\Xi^m, \Xi^s}{\arg\min}\ 1/2 \sum_{i,j} \rho(E^2(\xi_i^m, \xi_j^s; \sum, \xi_{ji})),$$

where the sub-map poses Ξ^m and the scan poses Ξ^s are optimized according to constraints [13].

Results and Discussion

Experiment. The simulation features a simple differential drive robot. The simulation process passes two stages as depicted in Fig. 2. First, the linear regression model is trained on relative depth estimated as an input, and the target will be the ground truth depth from an RGBD camera. When triggered (after linear regression finishes training), the second stage starts, the robot is loaded into a new environment where the trained linear regression model rescales and shifts the relative disparity transforming it to a metric one which is then inversed to achieve metric depth. From this metric depth only the middle row is projected as a 3D point cloud forming a pseudo laser scanner that is used as an input to Cartographer SLAM.

Metric Depth Results. We first present the performance of the aligned metric depth of our algorithm. In the process of evaluating the MDE performance, the testing prediction and ground truth are masked according to cropping criteria corresponding to a specific region of interest in the images of the testing dataset, like the Eigen-crop proposed by [14], or the Garg-crop proposed by [15]. These criteria are typically used to ensure that the evaluation metrics reflect the performance of the model in areas that are most relevant to the task. In our work, the alignment procedure aims to render pre-trained work applicable in an environment of choice; hence, we evaluate the entire image (no cropping) pursuing a better comprehensive assessment of the performance across the entire field of view as shown in Fig. 3. MDE is treated as a regression problem and is evaluated using the following metrics, where M is the number of pixels d_i, d_i^*, are ground truth and predicted depth:

— Root Mean Squared Error (RMSE):

$$RMSE = \sqrt{\frac{1}{M}\sum_{i}^{M} |d_i - d_i^*|^2}.$$

— Mean Relative Error (REL):

$$REL = \frac{1}{M}\sum_{i}^{M} \frac{|d_i - d_i^*|}{d_i}.$$

— Threshold accuracy (δ): ratio of predicted pixels having relative errors within a threshold.

$$\max\left(\frac{d^*}{d}, \frac{d}{d^*}\right) < \delta_i,\ \delta_i = 1.25^i.$$

We compare the metric depth performance with the ground truth by evaluating multiple metrics, and the results are shown in Table.

Better performance is highlighted and made bold. Note that the maximum depth in our testing environment is 22 m.

It is worth mentioning that these results are lacking in terms of accuracy, but when projecting to 3D point clouds, we can see that the nearest object seen by the camera is aligned almost perfectly with the ground truth. We figure that limiting the range of the projected depth on the

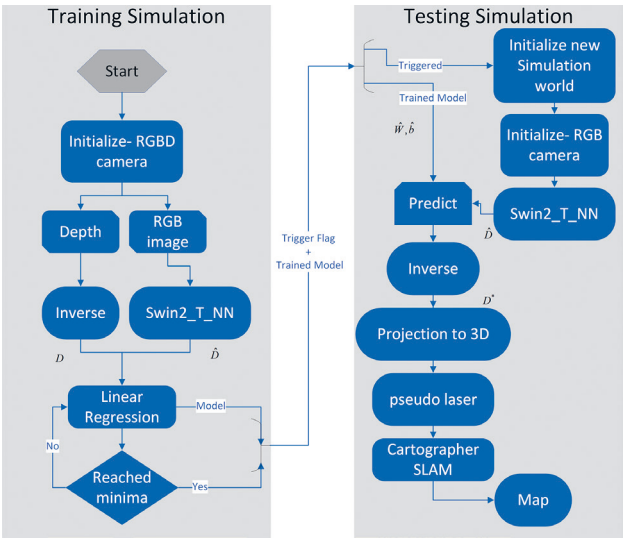


Fig. 2. Simulation algorithm

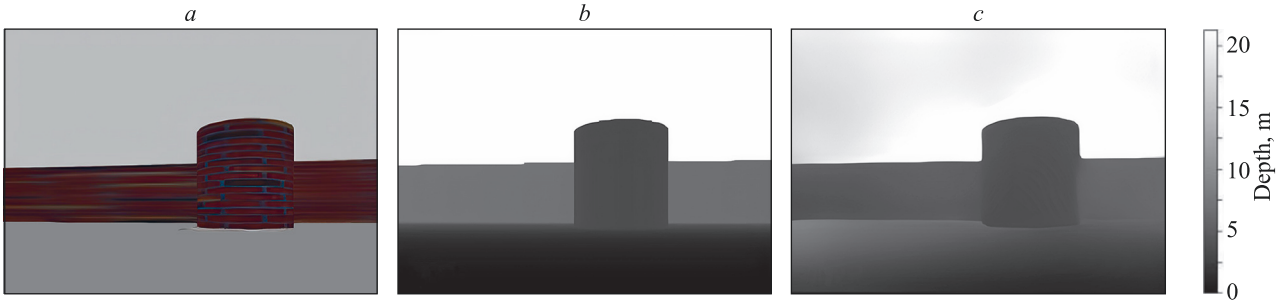


Fig. 3. Input RGB image (*a*), ground truth depth (*b*), aligned predicted depth (*c*), maximum width and height seen in the scene is 12 × 6.5 m

Table. Metric Depth results. Downward arrows ↓ mean the lower the value the better, upward ↑ mean higher is better

MDE algorithm	δ_1↑	δ_2↑	δ_3↑	*REL*↓, m	*RMSE*↓, m	Image dimension, pixel	Frames per second, Hz
Ada_bins [1]	0.438	0.588	0.760	1.158	3.886	640 × 480	8.5
Proposed algorithm	0.454	0.499	0.587	1.759	4.452	256 × 256	28.2

synthetic 2D laser scanner proves to be useful in solving depth accuracy problems.

Mapping. In this section, we present the results of mapping using our algorithm in comparison to the map generated by using the RGBD camera. The objective is to assess the accuracy of our approach in simulating the performance of real-world RGBD sensors. To that end, we run two consecutive simulations in the same world where in the first one we drive the robot equipped with the RGBD camera to map the environment, saving the trajectory the robot traversed while mapping as a file containing the action commands (linear, and angular velocities). When done, the map is saved as an image. In the second simulation we let the robot equipped with only an RGB camera use the same saved action commands to traverse the environment again following the same trajectory and mapping using our approach, since Cartographer SLAM is a pose graph optimization algorithm; making the robot pass through the same poses will produce similar grid maps that can be compared. This is done to ensure that acquired maps are aligned and variations between sub-maps are minimal. It is worth mentioning that the robot starts from the same position in the two simulations (near the bottom right wall), the trajectory of choice followed by the robot is a non-complete clockwise lap along the wall. When it reaches the center cylinders, it makes full rotation around its *z*-axis.

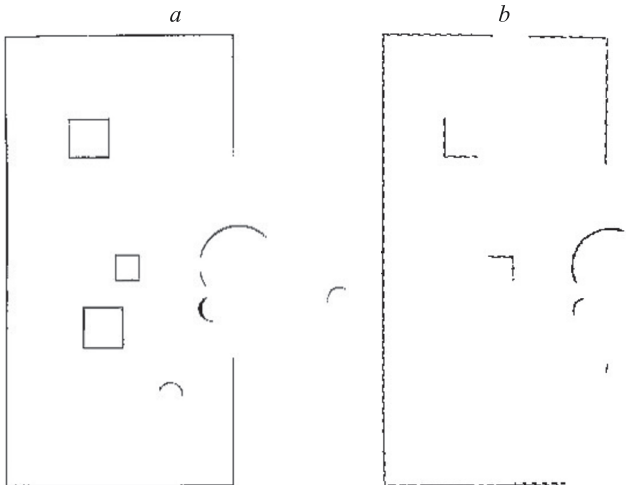


Fig. 4. Mapping results using the ideal RGBD camera (*a*), mapping using our algorithm (*b*)

To evaluate, we first address the accuracy of the mapping in terms of spatial fidelity and spatial consistency where we examine the alignment of key landmarks. To that end, we calculate the overlap between the map generated using our approach and the map saved using the RGBD which amounts to 88.17 %, indicating a high degree of correspondence between the two maps. We address obstacle recognition where it can be seen from Fig. 4 that in certain areas our algorithm fails to recognize the entirety of the object. This is due to the uncertainty in mapping generated by Frame-to-Frame variability since the network is designed to predict the depth in each frame. In some cases, the prediction may exhibit small variations due to the network sensitivity to visual changes and the rabid change between consecutive frames.

Conclusion

In conclusion, our approach showcases significant promise in mapping simulated environments. In terms of metric depth estimation, it falls short when compared to other algorithms trained to directly produce metric depth on certain datasets, which is reasonable since we traded the accuracy of the depth map with a faster inference rate. However, this low-accuracy depth map does not hinder the algorithm where in terms of mapping simulated environments, the high map overlap presented in the experimental results reinforces its effectiveness and potential for various robotics and simulation applications, emphasizing the balance between accuracy and real-time performance. It is paramount moving forward to boost the algorithm ability to handle dynamic scenes as well

as enhance depth estimation and improve its accuracy. Conducting real-world testing and further optimizing, the algorithm are also venues that should be addressed in the future in an effort to pave the way for robust and accurate mapping in resource-constrained platforms.

References

1. Bhat S.F., Alhashim I., Wonka P. AdaBins: Depth estimation using adaptive bins. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2021, pp. 4009–4018. https://doi.org/10.1109/cvpr46437.2021.00400
2. Li Z., Wang X., Liu X., Jiang J. BinsFormer: Revisiting adaptive bins for monocular depth estimation. *arXiv*, 2022, arXiv:2204.00987. https://doi.org/10.48550/arXiv.2204.00987
3. Zhang S., Yang L., Mi M.B., Zheng X., Yao A. Improving deep regression with ordinal entropy. *arXiv*, 2023, arXiv:2301.08915. https://doi.org/10.48550/arXiv.2301.08915
4. Ranftl R., Bochkovskiy A., Koltun V. Vision transformers for dense prediction. *Proc. of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 12179–12188. https://doi.org/10.1109/iccv48922.2021.0119
5. Xie Z., Geng Z., Hu J., Zhang Z., Hu H., Cao Y. Revealing the dark secrets of masked image modeling. *Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 14475–14485. https://doi.org/10.1109/cvpr52729.2023.01391
6. Birkl R., Wofk D., Müller M. MiDaS v3.1 – A Model zoo for robust monocular relative depth estimation. *arXiv*, 2023, arXiv:2307.14460. https://doi.org/10.48550/arXiv.2307.14460
7. Bhat S.F., Birkl R., Wofk D., Wonka P., Müller M. ZoeDepth: Zero-shot transfer by combining relative and metric depth. *arXiv*, 2023, arXiv:2302.12288. https://doi.org/10.48550/arXiv.2302.12288
8. Ranftl R., Lasinger K., Hafner D., Schindler K., Koltun V. Towards robust monocular depth estimation: mixing datasets for zero-shot cross-dataset transfer. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2022, vol. 44, no. 3, pp. 1623–1637. https://doi.org/10.1109/tpami.2020.3019967
9. Xing X., Cai Y., Lu T., Yang Y., Wen D. Joint self-supervised monocular depth estimation and SLAM. *Proc. of the 26th International Conference on Pattern Recognition (ICPR)*, 2022, pp. 4030–4036. https://doi.org/10.1109/icpr56361.2022.9956576
10. Geng M., Shang S., Ding B., Wang H., Zhang P. Unsupervised learning-based depth estimation-aided visual slam approach. *Circuits, Systems, and Signal Processing*, 2020, vol. 39, pp. 543–570. https://doi.org/10.1007/s00034-019-01173-3
11. Li Z., Snavely N. MegaDepth: Learning single-view depth prediction from internet photos. *Proc. of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2041–2050. https://doi.org/10.1109/cvpr.2018.00218
12. Tran M., Ly N. Mobile robot planner with low-cost cameras using deep reinforcement learning. *Proc. of the 7th NAFOSTED Conference on Information and Computer Science (NICS)*, 2020, pp. 54–59. https://doi.org/10.1109/nics51282.2020.9335852
13. Hess W., Kohler D., Rapp H., Andor D. Real-time loop closure in 2D LIDAR SLAM. *Proc. of the 2016 IEEE International Conference on Robotics and Automation (ICRA)*, 2016, pp. 1271–1278. https://doi.org/10.1109/icra.2016.7487258
14. Eigen D., Puhrsch C., Fergus R. Depth map prediction from a single image using a multi-scale deep network. *Advances in Neural Information Processing Systems*, 2014, vol. 27, pp. 2, 5, 6.
15. Garg R., Kumar B.G.V., Carneiro G., Reid I. Unsupervised CNN for single view depth estimation: Geometry to the rescue. *Lecture Notes in Computer Science*, 2016, vol. 9912, pp. 740–756. https://doi.org/10.1007/978-3-319-46484-8_45

Литература

1. Bhat S.F., Alhashim I., Wonka P. AdaBins: Depth estimation using adaptive bins // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021. P. 4009–4018. https://doi.org/10.1109/cvpr46437.2021.00400
2. Li Z., Wang X., Liu X., Jiang J. BinsFormer: Revisiting adaptive bins for monocular depth estimation // arXiv. 2022. arXiv:2204.00987. https://doi.org/10.48550/arXiv.2204.00987
3. Zhang S., Yang L., Mi M.B., Zheng X., Yao A. Improving deep regression with ordinal entropy // arXiv. 2023. arXiv:2301.08915. https://doi.org/10.48550/arXiv.2301.08915
4. Ranftl R., Bochkovskiy A., Koltun V. Vision transformers for dense prediction // Proc. of the IEEE/CVF International Conference on Computer Vision. 2021. P. 12179–12188. https://doi.org/10.1109/iccv48922.2021.01196
5. Xie Z., Geng Z., Hu J., Zhang Z., Hu H., Cao Y. Revealing the dark secrets of masked image modeling // Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023. P. 14475–14485. https://doi.org/10.1109/cvpr52729.2023.01391
6. Birkl R., Wofk D., Müller M. MiDaS v3.1 – A Model zoo for robust monocular relative depth estimation // arXiv. 2023. arXiv:2307.14460. https://doi.org/10.48550/arXiv.2307.14460
7. Bhat S.F., Birkl R., Wofk D., Wonka P., Müller M. ZoeDepth: Zero-shot transfer by combining relative and metric depth // arXiv. 2023. arXiv:2302.12288. https://doi.org/10.48550/arXiv.2302.12288
8. Ranftl R., Lasinger K., Hafner D., Schindler K., Koltun V. Towards robust monocular depth estimation: mixing datasets for zero-shot cross-dataset transfer // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2022. V. 44. N 3. P. 1623–1637. https://doi.org/10.1109/tpami.2020.3019967
9. Xing X., Cai Y., Lu T., Yang Y., Wen D. Joint self-supervised monocular depth estimation and SLAM // Proc. of the 26th International Conference on Pattern Recognition (ICPR). 2022. P. 4030–4036. https://doi.org/10.1109/icpr56361.2022.9956576
10. Geng M., Shang S., Ding B., Wang H., Zhang P. Unsupervised learning-based depth estimation-aided visual slam approach // Circuits, Systems, and Signal Processing. 2020. V. 39. P. 543–570. https://doi.org/10.1007/s00034-019-01173-3
11. Li Z., Snavely N. MegaDepth: Learning single-view depth prediction from internet photos // Proc. of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2018. P. 2041–2050. https://doi.org/10.1109/cvpr.2018.00218
12. Tran M., Ly N. Mobile robot planner with low-cost cameras using deep reinforcement learning // Proc. of the 7th NAFOSTED Conference on Information and Computer Science (NICS). 2020. P. 54–59. https://doi.org/10.1109/nics51282.2020.9335852
13. Hess W., Kohler D., Rapp H., Andor D. Real-time loop closure in 2D LIDAR SLAM // Proc. of the 2016 IEEE International Conference on Robotics and Automation (ICRA). 2016. P. 1271–1278. https://doi.org/10.1109/icra.2016.7487258
14. Eigen D., Puhrsch C., Fergus R. Depth map prediction from a single image using a multi-scale deep network // Advances in Neural Information Processing Systems. 2014. V. 27. P. 2, 5, 6.
15. Garg R., Kumar B.G.V., Carneiro G., Reid I. Unsupervised CNN for single view depth estimation: Geometry to the rescue // Lecture Notes in Computer Science. 2016. V. 9912. P. 740–756. https://doi.org/10.1007/978-3-319-46484-8_45

Authors

Majd Barhoum — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0007-8022-7932, barhoum.majd213@gmail.com

Anton A. Pyrkin — D.Sc., Full Professor, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 26656070700, https://orcid.org/0000-0001-8806-4057, a.pyrkin@gmail.com

Авторы

Бархум Мажд — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0007-8022-7932, barhoum.majd213@gmail.com

Пыркин Антон Александрович — доктор технических наук, профессор, профессор, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 26656070700, https://orcid.org/0000-0001-8806-4057, a.pyrkin@gmail.com

Received 09.11.2023
Approved after reviewing 18.12.2023
Accepted 20.01.2024

Статья поступила в редакцию 09.11.2023
Одобрена после рецензирования 18.12.2023
Принята к печати 20.01.2024.

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-124-132
УДК 004.891

Сегментация мышечной ткани на снимках компьютерной томографии на уровне позвонка L3

Анастасия Романовна Теплякова[1]✉, Роман Владимирович Шершнев[2], Сергей Олегович Старков[3], Татев Артаковна Агабабян[4], Валерия Андреевна Кукарская[5]

[1,2,3] Обнинский институт атомной энергетики — филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский ядерный университет «МИФИ» (ИАТЭ НИЯУ МИФИ), Обнинск, 249039, Российская Федерация
[4,5] Медицинский радиологический научный центр им. А.Ф. Цыба — филиал ФГБУ «Национальный медицинский исследовательский центр радиологии» Минздрава России, Обнинск, 249036, Российская Федерация

[1] anastasija-t23@mail.ru✉, https://orcid.org/0000-0001-8612-3850
[2] rvshershnev@yandex.ru, https://orcid.org/0009-0000-2324-5893
[3] sergeystarkov56@mail.ru, https://orcid.org/0000-0002-0420-7856
[4] tatevik.05@mail.ru, https://orcid.org/0000-0002-9971-3451
[5] kukarely@mail.ru, https://orcid.org/0009-0001-4872-3182

Аннотация

Введение. С возрастающей рутинной нагрузкой на врачей-рентгенологов, связанной с необходимостью анализировать большое количество снимков, возникает потребность в автоматизации части процесса анализа. Саркопения представляет собой состояние, при котором происходит потеря мышечной массы. Для диагностики саркопении наиболее часто применяется компьютерная томография, по снимкам которой может быть оценен объем мышечной ткани. Первым этапом анализа является ее оконтуривание, которое выполняется вручную, занимает продолжительное время и не всегда производится достаточно качественно, что оказывает влияние на точность оценок и, как следствие, на план лечения пациента. Предметом исследования является применение подходов компьютерного зрения для точной сегментации мышечной ткани по снимкам компьютерной томографии с целью саркометрии. Цель исследования — разработка подхода к решению задачи сегментации собранных и размеченных снимков. **Метод.** Представленный подход включает в себя этапы предварительной обработки снимков, сегментации при помощи нейронных сетей семейства U-Net и постобработки. Всего рассмотрено 63 различных конфигурации подхода, которые отличаются с точки зрения данных, подаваемых на вход моделей, и архитектур моделей. Оценено влияние предложенного способа постобработки получаемых бинарных масок на точность сегментации. **Основные результаты.** Для конфигурации подхода, включающей предварительную обработку с маскированием стола томографа и применением анизотропной диффузионной фильтрации, сегментацию моделью, имеющей архитектуру Inception U-Net и постобработку на базе анализа контуров, получен коэффициент схожести Дайса (Dice similarity coefficient, DSC) 0,9379 и пересечение над объединением (Intersection over Union, IoU) — 0,8824. Девять из исследованных в работе конфигураций также продемонстрировали высокие значения метрик DSC (в диапазоне 0,9356–0,9374) и IoU (0,8794–0,8822). Предлагаемый подход на базе предобработанных трехканальных изображений позволяет достигать значений 0,9364 для DSC и 0,8802 для IoU с применением легковесной модели сегментации U-Net. В соответствии с описанным подходом реализован программный модуль на языке Python. **Обсуждение.** Результаты исследования подтверждают целесообразность применения компьютерного зрения для оценки показателей мышечной ткани. Разработанный модуль может применяться для снижения рутинной нагрузки на рентгенологов.

Ключевые слова

компьютерное зрение, сегментация, компьютерная томография, мышечная ткань, индекс скелетной мускулатуры, саркопения, диагностика

Ссылка для цитирования: Теплякова А.Р., Шершнев Р.В., Старков С.О., Агабабян Т.А., Кукарская В.А. Сегментация мышечной ткани на снимках компьютерной томографии на уровне позвонка L3 // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 124–132. doi: 10.17586/2226-1494-2024-24-1-124-132

© Теплякова А.Р., Шершнев Р.В., Старков С.О., Агабабян Т.А., Кукарская В.А., 2024

Segmentation of muscle tissue in computed tomography images at the level of the L3 vertebra

Anastasia R. Teplyakova[1]✉, Roman V. Shershnev[2], Sergey O. Starkov[3], Tatev A. Agababian[4], Valeria A. Kukarskaya[5]

[1,2,3] Obninsk Institute for Nuclear Power Engineering, Obninsk, 249039, Russian Federation
[4,5] A. Tsyb Medical Radiological Research Centre – branch of the National Medical Research Radiological Center of the Ministry of Health of the Russia, Obninsk, 249036, Russian Federation

[1] anastasija-t23@mail.ru✉, https://orcid.org/0000-0001-8612-3850
[2] rvshershnev@yandex.ru, https://orcid.org/0009-0000-2324-5893
[3] sergeystarkov56@mail.ru, https://orcid.org/0000-0002-0420-7856
[4] tatevik.05@mail.ru, https://orcid.org/0000-0002-9971-3451
[5] kukarely@mail.ru, https://orcid.org/0009-0001-4872-3182

Abstract

With the increasing routine workload on radiologists associated with the need to analyze large numbers of images, there is a need to automate part of the analysis process. Sarcopenia is a condition in which there is a loss of muscle mass. To diagnose sarcopenia, computed tomography is most often used, from the images of which the volume of muscle tissue can be assessed. The first stage of the analysis is its contouring, which is performed manually, takes a long time and is not always performed with sufficient quality affecting the accuracy of estimates and, as a result, the patient's treatment plan. The subject of the study is the use of computer vision approaches for accurate segmentation of muscle tissue from computed tomography images for the purpose of sarcometry. The purpose of the study is to develop an approach to solving the problem of segmentation of collected and annotated images. An approach is presented that includes the stages of image pre-processing, segmentation using neural networks of the U-Net family, and post-processing. In total, 63 different configurations of the approach are considered, which differ in terms of data supplied to the input models and model architectures. The influence of the proposed method of post-processing the resulting binary masks on the segmentation accuracy is also evaluated. The approach, which includes pre-processing with table masking and anisotropic diffusion filtering, segmentation with an Inception U-Net architecture model, and post-processing based on contour analysis, achieves a Dice similarity coefficient of 0.9379 and Intersection over Union of 0.8824. Nine other configurations, the experimental results for which are reflected in the article, also demonstrated high values of these metrics (in the ranges of 0.9356–0.9374 and 0.8794–0.8822, respectively). The approach proposed in the article based on preprocessed three-channel images allows us to achieve metrics of 0.9364 and 0.8802, respectively, using the lightweight U-Net segmentation model. In accordance with the described approach, a software module was implemented in Python. The results of the study confirm the feasibility of using computer vision to assess muscle tissue parameters. The developed module can be used to reduce the routine workload on radiologists.

Keywords

computer vision, segmentation, computed tomography, muscle tissue, skeletal muscle index, sarcopenia, diagnostics

For citation: Teplyakova A.R., Shershnev R.V., Starkov S.O., Agababian T.A., Kukarskaya V.A. Segmentation of muscle tissue in computed tomography images at the level of the L3 vertebra. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 124–132 (in Russian). doi: 10.17586/2226-1494-2024-24-1-124-132

Введение

Саркопения имеет большую значимость в прогнозировании онкологических заболеваний [1]. В последние годы растет количество источников, подтверждающих тот факт, что оценка мышечной массы и определение вероятности развития саркопении у пациента при помощи рентгенологических методов позволяют получить эффективные биомаркеры. На сегодняшний день компьютерная томография (КТ) — не только точный и удобный с практической точки зрения метод оценки структуры тела, но и обязательная составная часть плановых исследований для диагностики, стадирования, наблюдения и оценки ответа на терапию у больных с опухолями [2, 3]. Однако в силу своего главного недостатка, заключающегося в дозовой нагрузке на пациента, КТ не может служить инструментом для оценки мышечной массы по результатам сканирования тела полностью. В связи с этим возникает необходимость применения подходов, позволяющих оценивать скелетно-мышечный индекс по аксиальному срезу КТ на определенном уровне. В [4–6] показано, что измеренная при КТ площадь мышечной и жировой ткани на уровне третьего поясничного позвонка (L3) надежно коррелирует с объемом скелетной мускулатуры и жировой клетчатки всего тела у здоровых людей.

Обзор существующих подходов

В рамках исследования выполнен анализ ряда научных работ, посвященных подходам к диагностике саркопении по снимкам КТ. В работе [7], опубликованной Научно-практическим клиническим центром диагностики и телемедицинских технологий Департамента здравоохранения города Москвы, представлен обзор современных подходов к диагностике саркопении. Авторы подчеркивают перспективность исследований, направленных на создание программного обеспечения для оценки мышечной ткани. Рассмотрены две группы существующих методов сегментации (на базе поиска мышечной ткани по плотности в диапазонах, задаваемых в единицах Хаунсфилда (Hounsfield units, HU), и на базе применения машинного обучения). Предложен метод решения задачи через применение сети U-Net++

на обучающей выборке 15 пациентов. В результате получено значение коэффициента схожести Дайса (Dice similarity coefficient, DSC) — 0,9312. Применение сети с более сложной архитектурой помогло решить проблему, связанную с малым объемом выборки, однако привело к значительному увеличению вычислительных затрат на этапах обучения сети и предсказания в рамках клинического использования.

В работе [8] предложено применение архитектуры U-Net. Для обучения использован набор данных, включающий размеченные вручную снимки 102 уникальных пациентов. Усредненное значение DSC для всех сегментируемых в исследовании групп мышц составило 0,930 ± 0,049. На предлагаемых в работе [9] (в ней применен описанный в [8] алгоритм) снимках с демонстрацией работы алгоритма заметны артефакты сегментации, лежащие вне зоны интереса, что обусловлено отсутствием постобработки получаемых масок. В [10] реализована сегментация мышечной ткани при помощи ансамбля из 15 моделей глубокого обучения с архитектурой U-Net с 4-уровневой глубиной и 18 начальными фильтрами. Использовано 1025 отдельных срезов КТ на уровне позвонка L3 (для обучения и валидации) и 500 дополнительных срезов — для тестирования. Полученное среднее значение DSC равно 0,97. В работе [11] предложен подход к сегментации мышц и жировой ткани по снимкам КТ на уровне позвонка L3 у пациентов с политравмой, в качестве архитектуры выбрана U-Net, модель обучена на снимках 3413 пациентов. Получено среднее значение DSC для мышечной ткани — 0,926. Недостаток данного подхода — отсутствие предварительной обработки, из-за которого в масках сегментации присутствуют руки пациентов (авторы исследования выделили это для будущей работы). В [12] при помощи U-Net сегментированы отдельные группы мышц. Для обучения использованы снимки 130 пациентов, удаление артефактов реализовано за счет исключения пикселей со значениями, выходящими за рамки интервала от –20 до 200 HU. Для отдельных групп мышц получены значения DSC в диапазоне 0,86–0,95, для общей мышечной массы — 0,95. В работе [13] акцент сделан на сегментации отдельных групп мышц в случаях с цирротической саркопенией. Для решения задачи применена архитектура 2,5D U-Net в сочетании с ветвью 3D-кодирования. Набор данных содержит снимки 317 пациентов (из них 216 с циррозом печени). Достигнуто среднее значение DSC равное 0,937. Метод, описанный в работе [14], основан на сегментации при помощи полностью сверточной нейронной сети (Fully Convolutional Network, FCN). Для обучения использовано три набора данных, включающих в себя 6651 снимок КТ на уровне позвонка L3, достигнуто значение DSC — 0,9811. В [15] описано создание системы, состоящей из модулей сегментации мышечной и жировой тканей, их количественной оценки и анализа саркопении. В рамках обучения и тестирования моделей сегментации собраны и размечены снимки КТ 67 пациентов, проведены эксперименты с архитектурами Attention U-Net, U-Net и SegNet. Attention U-Net продемонстрировала самое высокое значение DSC для мышечной ткани равное 0,96. Преимуществом работы является нацеленность на применение системы в клинической практике, учет времени обработки модулями исследования отдельного пациента (на сегментацию снимка затрачивается порядка 6 с). В [16] предложен полностью автоматизированный подход, который не только реализует сегментацию снимка КТ на уровне позвонка L3, но и обнаруживает его в массиве снимков исследования КТ пациента. Сегментация реализована обученной на 1070 снимках КТ FCN, основанной на U-Net-подобной архитектуре. DSC для отдельных групп мышц достигают значений в диапазоне 0,94–0,97, для объединенной мышечной массы — 0,96, заявленное время обработки одного исследования — 1 с.

Главные достоинства рассмотренных работ заключаются в том, что данные для исследований собраны в медицинских центрах и размечены квалифицированными рентгенологами, а также в том, что результаты востребованы и протестированы в условиях клинической практики. Однако в большинстве исследований мало внимания уделено технической реализации решений: предварительной обработке снимков, сравнению различных архитектур, методам постобработки масок сегментации. Программные части и наборы данных, полученные в рассмотренных зарубежных исследованиях, являются закрытыми, что влечет за собой невозможность применения результатов в клинической практике и доработки функционала под запросы специалистов-рентгенологов. Стоит также отметить отсутствие готового программного обеспечения для точной сегментации мышечной ткани по снимкам КТ на уровне позвонка L3 на российском рынке. В связи с этим возникла потребность в реализации собственного программного обеспечения, один из модулей которого может решать описанную задачу.

Преимущество предлагаемого в настоящей работе подхода заключается в реализации большого количества способов решения задачи: представлены результаты 63 различных модификаций подхода, для каждого из которых обучена своя модель. Варьируются 9 типов входных данных модели (в зависимости от применяемого алгоритма предварительной обработки снимков) и 7 архитектур сегментации (U-Net, Attention U-Net, U-Net++, ResU-Net, Inception U-Net, Dense U-Net, U^2-Net). Для каждой из конфигураций выполнено сравнение метрик до и после применения алгоритма постобработки масок сегментации. Отметим, что предложен новый подход к обработке входных снимков, позволяющий повысить производительность самой легковесной из перечисленных архитектур (U-Net) до уровня производительности более сложных из них.

Постановка задачи

При проведении анализа состава тела пациента врач-рентгенолог выполняет количественную оценку мышц по снимкам КТ брюшной полости. Для этого он вручную осуществляет разметку мышечной ткани на снимке, соответствующем уровню позвонка L3. Разметка снимка КТ представляет собой формирование маски сегментации, основанной на точках контуров (каждая наносится нажатием кнопки компьютерной

мыши) мышечной ткани, и занимает порядка 5,5 мин без использования предварительной разметки, выдаваемой полуавтоматическими программными средствами (с нуля), и порядка 2,8 мин — с ее использованием [17]. Задача заключается в разработке программного модуля, получающего на вход DICOM файл со снимком КТ на уровне позвонка L3 и возвращающего его в обработанном виде (с оверлеем, содержащим контуры мышечной ткани). Задача включает в себя подзадачи:

— создание набора данных (сбор снимков КТ брюшной полости на уровне позвонка L3, разметка мышечной ткани);
— разработка алгоритмов предварительной обработки снимков КТ;
— построение, обучение и тестирование моделей сегментации;
— разработка алгоритмов постобработки масок сегментации;
— разработка алгоритмов, отвечающих за работу с файлами (загрузка данных из входных файлов, передача их алгоритмам обработки и сегментации, запись результатов в выходные файлы).

Описание набора данных

Набор данных, примененный для обучения нейронных сетей и тестирования подходов к обработке снимков, основанных на классическом компьютерном зрении, собран из данных КТ онкологических больных. Пациенты получили комбинированное лечение в Медицинском радиологическом научном центре им. А.Ф. Цыба и для них были доступны изображения КТ, полученные до операции. Для формирования набора рентгенологами были отобраны снимки КТ, соответствующие уровню позвонка L3.

Исследования выполнялись на мультиспиральном компьютерном томографе Optima CT660 (GE). Параметры КТ-исследования: напряжение 120 кВ; сила тока в среднем 95 мА·с (в зависимости от конституции пациента); матрица изображения 512 × 512 пикселей; толщина томографического среза — 5 мм, коллимация — 6 ×2 мм. Полученные КТ-данные реконструированы с толщиной среза 2,5 мм и инкрементом 1–1,5 мм.

КТ-исследования проведены в положении больного лежа на спине с вытянутыми над головой руками, при необходимости применялось полипозиционное исследование с целью улучшения визуализации. Топограмма выполнена при задержке дыхания в фазу вдоха в краниокаудальном направлении. По топограмме произведено планирование области исследования. Сделано нативное (бесконтрастное) исследование, затем сканирование с контрастным усилением по стандартной методике.

Разметка мышечной ткани, заключающаяся в формировании бинарных масок для всех снимков набора, произведена опытным рентгенологом, имеющим более 10 лет медицинского стажа. Набор данных содержит 207 снимков КТ на уровне позвонка L3 119 уникальных пациентов (67 мужчин, 52 женщины). Медианные возраста — 63 и 59,5 лет для снимков пациентов мужского и женского полов. Самому молодому пациенту 28 лет, самому возрастному — 81. Половозрастное распределение снимков показано на рис. 1. Набор разделен на обучающую, валидационную и тестовую выборки, содержащие 161, 23 и 23 снимков.

Описание разработанного модуля сегментации снимков

В основе логики функционирования разработанного модуля лежат три этапа.

Этап 1. Получение из файлов DICOM входных данных для нейронных сетей. Вне зависимости от того, какой метод обработки применялся к снимкам КТ, для всех DICOM-файлов выполнялись следующие шаги.

Шаг 1. Чтение массивов пикселей (ширина и высота каждого составляют 512 пикселей);

Шаг 2. Переход от значений пикселей к HU.

Для дальнейшей обработки снимков реализованы три подхода, с использованием которых осуществлено формирование различных вариантов входных данных для моделей сегментации. Первый подход основан на анализе изображения в HU (уровень и ширина окна принимаются равными рекомендованным, считываемым из DICOM, для большинства исследований 40 HU и 400 HU соответственно) при помощи применения порога и выделения контура тела (через анализ размеров и положений). Данный подход необходим для маскирования всего, что лежит за пределами тела пациента: пиксели, относящиеся к областям стола томографа и рук пациента, преобразуются в черные.

Второй подход базируется на одном из методов классической обработки снимков — адаптивном выравнивании гистограмм с ограниченным усилением контраста (Contrast Limited Adaptive Histogram Equalization,

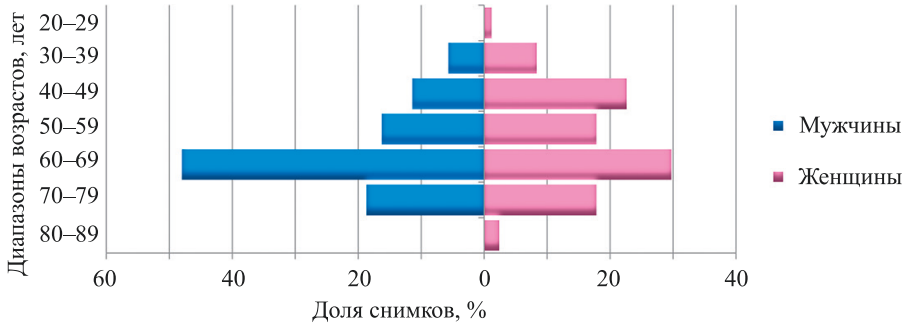


Рис. 1. Половозрастное распределение снимков
Fig. 1. Age and gender distribution of images

CLAHE). Цель применения подхода заключается в повышении контрастности изображения, в рамках исследования применяются пределы отсечения (Clip Limit, CL), равные 1, 2 и 3.

Для снимков КТ характерно нежелательное изменение значений пикселей — шум. Он приводит к снижению качества изображения и ухудшает контрастное разрешение. Увеличение отношения сигнал/шум (Signal/Noise Rate, SNR) представляет собой важную задачу. При уменьшении уровня шума должны сохраняться детали исходного изображения, чтобы оно не теряло клинически значимого содержания. Метод шумоподавления на базе применения анизотропного диффузионного фильтра (Anisotropic Diffusion Filter, ADF) лишен недостатков, связанных с размытием краев (для пространственных фильтров), сложностью правильного выбора частотных характеристик (для временных фильтров), порогов и масштабов (для вейвлетов), возникновением артефактов, отбрасыванием деталей мелких структур [18]. Исходя из этого, третий подход служит для предварительной обработки, отвечает за уменьшение уровня шума и основан на ADF.

Для обучения моделей сегментации применены 9 вариантов формирования входных данных:
— одноканальные базовые снимки с переходом к HU (*base*);
— одноканальные снимки *base*, на которых замаскировано все, кроме тела пациента (*masked*);
— одноканальные снимки masked с применением CLAHE с CL, равными 1, 2 и 3 (*CLAHE 1*, *CLAHE 2* и *CLAHE 3*);
— одноканальные снимки masked с примененным ADF (*ADF*);
— одноканальные снимки masked с последовательно примененными CLAHE с CL, равным 2, и ADF (*CLAHE ADF*);
— трехканальные снимки, у которых первый канал представляет *masked*, второй — *CLAHE 2*, третий — *ADF* (*3 channels*);
— двухканальные снимки, у которых первый канал представляет *masked*, второй — *CLAHE 2* (*2 channels*).

На рис. 2 показаны разные варианты одного снимка КТ.

Этап 2. Сегментация снимков и формирование для них бинарных масок, в которых белые пиксели соответствуют мышечной ткани, а черные — областям, лежащим за ее пределами. Для этого построены, обучены и протестированы 63 нейросетевых модели, имеющие 7 различных архитектур (для каждого из 9 вариантов входных данных — отдельная сеть). Все реализованные архитектуры объединяет то, что в основе лежат две подсети (кодер и декодер), однако каждая из них имеет свои особенности:
— U-Net представляет собой базовую архитектуру, самую легковесную из рассмотренных;
— Attention U-Net имеет ворота внимания, которые позволяют сети подавлять ненужные области и сосредотачивать внимание на существенных деталях различных форм и размеров;
— U-Net++ основана на вложенных плотных пропускных соединениях, введение которых призвано устранить семантический разрыв между картами признаков кодера и декодера;
— ResU-Net основана на использовании остаточных соединений, изучающих остаточные признаки со ссылкой на входные данные;
— Inception U-Net имеет блоки, которые позволяют использовать несколько типов размера фильтра вместо того, чтобы ограничиваться одним размером;
— Dense U-Net основана на замене базовых блоков U-Net на плотные блоки, позволяющей создавать более глубокие модели для более точной сегментации;
— U^2-Net представляет собой двухуровневую вложенную архитектуру, способную захватывать больше контекстной информации из разных масштабов и увеличивающую глубину анализа.

Этап 3. Постобработка бинарных масок сегментации, предсказываемых моделями, для устранения «артефактов» сегментации — белых пикселей, лежащих вне области интереса. Этот этап важен, так как позволяет предотвратить не только некорректную оценку площади мышечной ткани, но и ошибки в определении ее показателей в HU (минимум, максимум, среднее, стандартное отклонение). Предложенный алгоритм постобработки базируется на анализе контуров:
— белые пиксели, лежащие внутри эллипса, вписанного в область интереса (ограничена внутренней стороной мышечной ткани), преобразуются в черные;
— белые пиксели, лежащие за пределами эллипса, описанного вокруг области интереса (ограничивает внешнюю стороной мышечной ткани), преобразуются в черные.

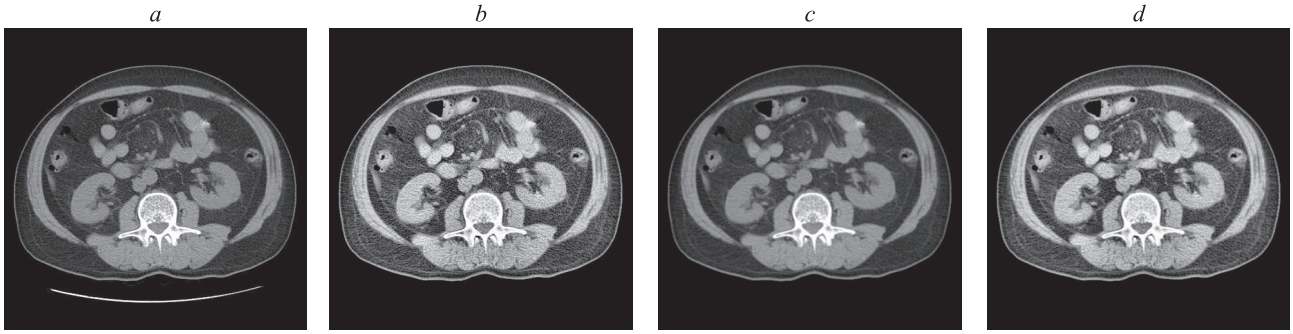


Рис. 2. Варианты входных снимков: *base* (*a*); *CLAHE 2* (*b*); *ADF* (*c*); *CHALE ADF* (*d*)
Fig. 2. Variations of input images: *base* (*a*); *CLAHE 2* (*b*); *ADF* (*c*); *CHALE ADF* (*d*)

Полученные результаты

Программный код разработанного модуля реализован на языке Python. Для создания, обучения и тестирования моделей сегментации использован фреймворк Keras, созданный на базе платформы TensorFlow.

Модуль получает на вход файл DICOM, формирует входные данные для модели, преобразуя содержащийся в нем снимок к необходимому виду, сегментирует снимок, обрабатывает полученную маску сегментации. На рис. 3 показан пример предварительно обработанного снимка и соответствующих ему масок сегментации мышечной ткани: истинной (сформирована рентгенологом) и предсказанной (результат работы модуля).

Для оценки моделей сегментации применены две метрики: DSC и пересечение над объединением (Intersection over Union, IoU). Для их расчета использованы следующие формулы:

$$DSC = 2 \cdot \frac{|pred \cap ground\_truth|}{|pred| + |ground\_truth|},$$

$$\mathrm{IoU} = \frac{|pred \cap ground\_truth|}{|pred \cup ground\_truth|}.$$

где *pred* — маска, предсказанная моделью; *ground_truth* — маска, полученная при разметке снимка рентгенологом. Значения DSC и IoU (табл. 1) рассчитаны для масок тестовых выборок, соответствующих всем описанным типам входных данных. Ко всем маскам, предсказанным моделями, применен алгоритм постобработки, что позволило повысить значения DSC и IoU в среднем на 0,18 % и 0,32 %. Для 5 архитектур из 7 рассмотренных входных данных с применением методов обработки дают лучшие DSC, чем базовые снимки с переходом к HU. В 6 случаях из 7 обработанных данных получены наилучшие значения IoU. Самая легкая из рассмотренных архитектур, U-Net, имеющая 1 940 817 обучаемых параметров с применением данных *3 channel*, достигла значения DSC, которое всего на 0,0015 меньше, чем максимальное из достигнутых (для Inception U-Net, имеющей 18 082 757 обучаемых параметров). При этом U-Net требует значительно меньше вычислительных и временных ресурсов для обучения, а также обеспечивает возможность использования модуля на менее мощных вычислительных устройствах. Это говорит о преимуществах применения предложенного подхода к формированию входных данных для моделей сегментации.

Время обработки модулем одного снимка в зависимости от выбранной модели сегментации (табл. 2) измерено для двух конфигураций вычислительного устройства: с графическим процессором NVIDIA GeForce RTX 3080 и без него, т. е. только с центральным процессором. Среднее время обработки снимка для конфигурации с центрального процессора не превышает 2 с, что значительно меньше времени, затрачиваемого рентгенологом при ручной разметке.

Таблица 1. Значения DSC и IoU
Table 1. DSC and IoU values

Типы входных данных	Архитектуры моделей сегментации						
	U-Net	Attention U-Net	U-Net++	ResU-Net	Inception U-Net	Dense U-Net	U²-Net
Значения DSC							
base	0,9320	0,9235	0,9333	**0,9366**	0,9346	**0,9374**	0,9358
masked	0,9302	0,9288	0,9321	0,9290	0,9353	0,9341	0,9350
CLAHE 1	0,9322	0,9062	0,9317	0,9322	0,9355	0,9325	**0,9370**
CLAHE 2	0,9304	0,9296	0,9287	0,9303	0,9329	0,9349	0,9343
CLAHE 3	0,9315	0,9202	0,9287	0,9296	0,9339	0,9333	0,9351
ADF	0,9328	0,9277	0,9338	0,9341	**0,9379**	0,9360	0,9347
CLAHE ADF	0,9295	**0,9297**	0,9290	0,9241	0,9323	0,9357	0,9353
3 channels	**0,9364**	0,9180	0,9330	0,9363	0,9333	0,9370	0,9350
2 channels	0,9310	0,9285	**0,9356**	0,9342	0,9351	0,9349	0,9366
Значения IoU							
base	0,8735	0,8583	0,8765	**0,8811**	0,8770	0,8821	0,8796
masked	0,8700	0,8679	0,8737	0,8679	0,8783	0,8771	0,8781
CLAHE 1	0,8722	0,8330	0,8728	0,8742	0,8789	0,8727	**0,8810**
CLAHE 2	0,8691	0,8688	0,8688	0,8701	0,8745	0,8772	0,8755
CLAHE 3	0,8713	0,8525	0,8682	0,8679	0,8760	0,8749	0,8772
ADF	0,8736	0,8669	0,8771	0,8771	**0,8824**	0,8796	0,8767
CLAHE ADF	0,8673	**0,8691**	0,8678	0,8602	0,8748	0,8789	0,8787
3 channels	**0,8802**	0,8506	0,8760	0,8804	0,8759	**0,8822**	0,8779
2 channels	0,8706	0,8675	**0,8794**	0,8778	0,8779	0,8770	0,8802

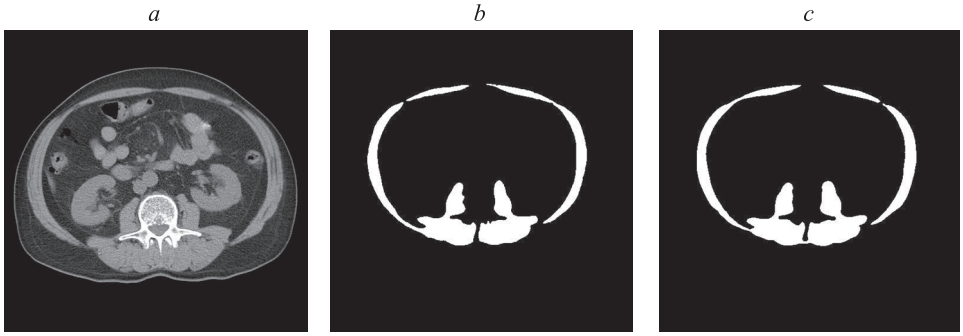


Рис. 3. Пример снимка и соответствующих ему масок сегментации: снимок модификации *masked* (*a*); истинная (*b*); предсказанная (*c*)

Fig. 3. Example of an image and corresponding segmentation masks: masked image (*a*); true mask (*b*); predicted mask (*c*)

Таблица 2. Среднее время обработки одного снимка, мс

Table 2. Average processing time per image, ms

Архитектуры моделей сегментации	Конфигурация вычислительного устройства	
	с графическим и центральным процессорами	с центральным процессором
U-Net	42,0	133
Attention U-Net	54,0	389
U-Net++	60,0	510
ResU-Net	52,2	358
Inception U-Net	90,8	1650
Dense U-Net	72,5	685
U^2-Net	102,0	925

Контуры, формируемые модулем сегментации, позволяют получать оценку площади мышц. Далее эта площадь должна быть нормализована к квадрату роста, чтобы быть репрезентативной для оценки мышечной массы человека в целом. Исходя из этого, следующим шагом при КТ-саркометрии является вычисление скелетно-мышечного индекса, равного отношению полученного показателя площади скелетной мускулатуры на уровне тела L3 позвонка к квадрату показателя роста пациента. Расчет скелетно-мышечного индекса (SMI) выполняется по формуле:

$$SMI = \frac{S}{h^2},$$

где S — площадь всех групп мышц, попавших в срез, см^2; h — рост человека, м.

Далее полученные значения скелетно-мышечного индекса сравниваются с референсными. Большинство авторов считают пороговым значением *SMI*, ниже которого состояние мышечной ткани расценивается как саркопения: 52,4 см^2/м^2 для мужчин и 38,5 см^2/м^2 для женщин [19, 20].

Заключение

Разработан программный модуль, позволяющий в автоматическом режиме сегментировать мышечную ткань по снимкам компьютерной томографии на уровне позвонка L3, тем самым упрощая процесс оценки мышечной массы при диагностике саркопении. Для обучения нейросетевых моделей собран и размечен набор данных, содержащий 207 снимков компьютерной томографии брюшной полости. Несмотря на небольшой объем обучающей выборки, достигнуты значения: коэффициента схожести Дайса равное 0,9374 и пересечение над объединением — 0,8822. Реализован ряд подходов к формированию входных данных моделей, позволяющих повысить точность сегментации. Один из подходов приближает точность легковесной U-Net к точностям более сложных архитектур, что является важным для применения в условиях ограниченных вычислительных ресурсов при применении в клинической практике. Результаты исследования подтверждают целесообразность применения компьютерного зрения для оценки показателей мышечной ткани, поэтому в рамках его развития планируется собрать и разметить больше обучающих данных, а также разработать модуль автоматического поиска снимка на уровне позвонка L3 по полной компьютерной томограмме пациента, модуль анализа гистограмм в единицах Хаунсфилда, программу для отображения медицинских данных с возможностью интеграции вышеупомянутых модулей.

Литература

1. Шеберова Е.В., Силантьева Н.К., Агабабян Т.А., Потапов А.Л., Невольских А.А., Иванов С.А., Каприн А.Д. Роль компьютерной томографии в диагностике саркопении // Сибирский онкологический журнал. 2023. Т. 22. № 3. С. 125–133. https://doi.org/10.21294/1814-4861-2023-22-3-125-133
2. Агабабян Т.А., Кукарская В.А., Силантьева Н.К., Потапов А.Л., Скоропад В.Ю., Шеберова Е.В., Дорожкин А.Д., Иванов С.А., Каприн А.Д. Роль КТ-саркометрии в прогнозировании послеоперационных осложнений у больных раком желудка // Современная онкология. 2023. Т. 25. № 3. С. 284–288. https://doi.org/10.26442/18151434.2023.3.202260
3. Tagliafico A.S., Bignotti B., Torri L., Rossi F. Sarcopenia: how to measure, when and why // La radiologia medica. 2022. V. 127. N 3. P. 228–237. https://doi.org/10.1007/s11547-022-01450-3
4. Shen W., Punyanitya M., Wang Z., Gallagher D., St-Onge M.P., Albu J., Heymsfield S.B., Heshka S. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image // Journal of Applied Physiology. 2004. V. 97. N 6. P. 2333–2338. https://doi.org/10.1152/japplphysiol.00744.2004
5. Van den Broeck J., Sealy M.J., Brussaard C., Kooijman J., Jager-Wittenaar H., Scafoglieri A. The correlation of muscle quantity and quality between all vertebra levels and level L3, measured with CT: An exploratory study // Frontiers in Nutrition. 2023. V. 10. P. 1148809. https://doi.org/10.3389/fnut.2023.1148809
6. Vedire Y., Nitsche L., Tiadjeri M., McCutcheon V., Hall J., Barbi J., Yendamuri S., Ray A.D. Skeletal muscle index is associated with long term outcomes after lobectomy for non-small cell lung cancer // BMC Cancer. 2023. V. 23. N 1. P. 778. https://doi.org/10.1186/s12885-023-11210-9
7. Сморчкова А.К., Петряйкин А.В., Семёнов Д.С., Шарова Д.Е. Саркопения: современные подходы к решению диагностических задач // Digital Diagnostics. 2022. Т. 3. № 3. С. 196–211. https://doi.org/10.17816/DD110721
8. Burns J.E., Yao J., Chalhoub D., Chen J.J., Summers R.M. A machine learning algorithm to estimate sarcopenia on abdominal CT // Academic Radiology. 2020. V. 27. N 3. P. 311–320. https://doi.org/10.1016/j.acra.2019.03.011
9. Graffy P.M., Liu J., Pickhardt P.J., Burns J.E., Yao J., Summers R.M. Deep learning-based muscle segmentation and quantification at abdominal CT: application to a longitudinal adult screening cohort for sarcopenia assessment // British Journal of Radiology. 2019. V. 92. N 1100. P. 20190327. https://doi.org/10.1259/bjr.20190327
10. Blanc-Durand P., Schiratti J.B., Schutte K., Jehanno P., Herent P., Pigneur F., Lucidarme O., Benaceur Y., Sadate A., Luciani A., Ernst O., Rouchaud A., Creze M., Dallongeville A., Banaste N., Cadi M., Bousaid I., Lassau N., Jegou S. Abdominal musculature segmentation and surface prediction from CT using deep learning for sarcopenia assessment // Diagnostic and Interventional Imaging. 2020. V. 101. N 12. P. 789–794. https://doi.org/10.1016/j.diii.2020.04.011
11. Ackermans L.L.G.C., Volmer L., Wee L., Brecheisen R., Sánchez-González P., Seiffert A.P., Gómez E.J., Dekker A., Ten Bosch J.A., Olde Damink S.M.W., Blokhuis T.J. Deep learning automated segmentation for muscle and adipose tissue from abdominal computed tomography in polytrauma patients // Sensors. 2021. V. 21. N 6. P. 2083. https://doi.org/10.3390/s21062083
12. Kreher R., Hinnerichs M., Preim B., Saalfeld S., Surov A. Deep-learning-based Segmentation of Skeletal Muscle Mass in Routine Abdominal CT Scans // In Vivo. 2022. V. 36. N 4. P. 1807–1811. https://doi.org/10.21873/invivo.12896
13. Song G., Zhou J., Wang K., Yao D., Chen S., Shi Y. Segmentation of multi-regional skeletal muscle in abdominal CT image for cirrhotic sarcopenia diagnosis // Frontiers in Neuroscience. 2023. V. 17. P. 1203823. https://doi.org/10.3389/fnins.2023.1203823
14. Dabiri S., Popuri K., Cespedes Feliciano E.M., Caan B.J., Baracos V.E., Beg M.F. Muscle segmentation in axial computed tomography (CT) images at the lumbar (L3) and thoracic (T4) levels for body composition analysis // Computerized Medical Imaging and Graphics. 2019. V. 75. P. 47–55. https://doi.org/10.1016/j.compmedimag.2019.04.007
15. Gu S., Wang L., Han R., Liu X., Wang Y., Chen T., Zheng Z. Detection of sarcopenia using deep learning-based artificial intelligence body part measure system (AIBMS) // Frontiers in Physiology. 2023. V. 14. P. 1092352. https://doi.org/10.3389/fphys.2023.1092352

References

1. Sheberova E.V., Silanteva N.K., Agababian T.A., Potapov A.L., Nevolskikh A.A., Ivanov S.A., Kaprin A.D. Role of computed tomography in sarcopenia detection. *Siberian Journal of Oncology*, 2023, vol. 22, no. 3, pp. 125–133. (in Russian). https://doi.org/10.21294/1814-4861-2023-22-3-125-133
2. Agababyan T.A., Kukarskaya V.A., Silanteva N.K., Potapov A.L., Skoropad V.YU., Sheberova E.V., Dorozhkin A.D., Ivanov S.A., Kaprin A.D. CT sarcometry in the prediction of postoperative complications in patients with gastric cancer: retrospective cohort study. *Journal of Modern Oncology*, 2023, vol. 25, no. 3, pp. 284–288. (in Russian). https://doi.org/10.26442/18151434.2023.3.202260
3. Tagliafico A.S., Bignotti B., Torri L., Rossi F. Sarcopenia: how to measure, when and why. *La radiologia medica*, 2022, vol. 127, no. 3, pp. 228–237. https://doi.org/10.1007/s11547-022-01450-3
4. Shen W., Punyanitya M., Wang Z., Gallagher D., St-Onge M.P., Albu J., Heymsfield S.B., Heshka S. Total body skeletal muscle and adipose tissue volumes: estimation from a single abdominal cross-sectional image. *Journal of Applied Physiology*, 2004, vol. 97, no. 6, pp. 2333–2338. https://doi.org/10.1152/japplphysiol.00744.2004
5. Van den Broeck J., Sealy M.J., Brussaard C., Kooijman J., Jager-Wittenaar H., Scafoglieri A. The correlation of muscle quantity and quality between all vertebra levels and level L3, measured with CT: An exploratory study. *Frontiers in Nutrition*, 2023, vol. 10, pp. 1148809. https://doi.org/10.3389/fnut.2023.1148809
6. Vedire Y., Nitsche L., Tiadjeri M., McCutcheon V., Hall J., Barbi J., Yendamuri S., Ray A.D. Skeletal muscle index is associated with long term outcomes after lobectomy for non-small cell lung cancer. *BMC Cancer*, 2023, vol. 23, no. 1, pp. 778. https://doi.org/10.1186/s12885-023-11210-9
7. Smorchkova A.K., Petraikin A.V., Semenov D.S., Sharova D.E. Sarcopenia: Modern approaches to solving diagnosis problems. *Digital Diagnostics*, 2022, vol. 3, no. 3, pp. 196–211. https://doi.org/10.17816/DD110721
8. Burns J.E., Yao J., Chalhoub D., Chen J.J., Summers R.M. A machine learning algorithm to estimate sarcopenia on abdominal CT. *Academic Radiology*, 2020, vol. 27, no. 3, pp. 311–320. https://doi.org/10.1016/j.acra.2019.03.011
9. Graffy P.M., Liu J., Pickhardt P.J., Burns J.E., Yao J., Summers R.M. Deep learning-based muscle segmentation and quantification at abdominal CT: application to a longitudinal adult screening cohort for sarcopenia assessment. *British Journal of Radiology*, 2019, vol. 92, no. 1100, pp. 20190327. https://doi.org/10.1259/bjr.20190327
10. Blanc-Durand P., Schiratti J.B., Schutte K., Jehanno P., Herent P., Pigneur F., Lucidarme O., Benaceur Y., Sadate A., Luciani A., Ernst O., Rouchaud A., Creze M., Dallongeville A., Banaste N., Cadi M., Bousaid I., Lassau N., Jegou S. Abdominal musculature segmentation and surface prediction from CT using deep learning for sarcopenia assessment. *Diagnostic and Interventional Imaging*, 2020, vol. 101, no. 12, pp. 789–794. https://doi.org/10.1016/j.diii.2020.04.011
11. Ackermans L.L.G.C., Volmer L., Wee L., Brecheisen R., Sánchez-González P., Seiffert A.P., Gómez E.J., Dekker A., Ten Bosch J.A., Olde Damink S.M.W., Blokhuis T.J. Deep learning automated segmentation for muscle and adipose tissue from abdominal computed tomography in polytrauma patients. *Sensors*, 2021, vol. 21, no. 6, pp. 2083. https://doi.org/10.3390/s21062083
12. Kreher R., Hinnerichs M., Preim B., Saalfeld S., Surov A. Deep-learning-based Segmentation of Skeletal Muscle Mass in Routine Abdominal CT Scans. *In Vivo*, 2022, vol. 36, no. 4, pp. 1807–1811. https://doi.org/10.21873/invivo.12896
13. Song G., Zhou J., Wang K., Yao D., Chen S., Shi Y. Segmentation of multi-regional skeletal muscle in abdominal CT image for cirrhotic sarcopenia diagnosis. *Frontiers in Neuroscience*, 2023, vol. 17, pp. 1203823. https://doi.org/10.3389/fnins.2023.1203823
14. Dabiri S., Popuri K., Cespedes Feliciano E.M., Caan B.J., Baracos V.E., Beg M.F. Muscle segmentation in axial computed tomography (CT) images at the lumbar (L3) and thoracic (T4) levels for body composition analysis. *Computerized Medical Imaging and Graphics*, 2019, vol. 75, pp. 47–55. https://doi.org/10.1016/j.compmedimag.2019.04.007
15. Gu S., Wang L., Han R., Liu X., Wang Y., Chen T., Zheng Z. Detection of sarcopenia using deep learning-based artificial intelligence body part measure system (AIBMS). *Frontiers in Physiology*, 2023, vol. 14, pp. 1092352. https://doi.org/10.3389/fphys.2023.1092352

16. Islam S., Kanavati F., Arain Z., Da Costa O.F., Crum W., Aboagye E.O., Rockall A.G. Fully automated deep-learning section-based muscle segmentation from CT images for sarcopenia assessment // Clinical Radiology. 2022. V. 77. N 5. P. e363–e371. https://doi.org/10.1016/j.crad.2022.01.036
17. Takahashi N., Sugimoto M., Psutka S.P., Chen B., Moynagh M.R., Carter R.E. Validation study of a new semi-automated software program for CT body composition analysis // Abdominal Radiology. 2017. V. 42. N 9. P. 2369–2375. https://doi.org/10.1007/s00261-017-1123-6
18. Kaur R., Juneja M., Mandal A.K. A comprehensive review of denoising techniques for abdominal CT images // Multimedia Tools and Applications. 2018. V. 77. N 17. P. 22735–22770. https://doi.org/10.1007/s11042-017-5500-5
19. Масенко В.Л., Коков А.Н., Григорьева И.И., Кривошапова К.Е. Лучевые методы диагностики саркопении // Исследования и практика в медицине. 2019. Т. 6. № 4. С. 127–137. https://doi.org/10.17709/2409-2231-2019-6-4-13
20. Lee J.S., Kim Y.S., Kim E.Y., Jin W. Prognostic significance of CT-determined sarcopenia in patients with advanced gastric cancer // PLoS ONE. 2018. V. 13. N 8. P. e0202700. https://doi.org/10.1371/journal.pone.0202700

16. Islam S., Kanavati F., Arain Z., Da Costa O.F., Crum W., Aboagye E.O., Rockall A.G. Fully automated deep-learning section-based muscle segmentation from CT images for sarcopenia assessment. *Clinical Radiology*, 2022, vol. 77, no. 5, pp. e363–e371. https://doi.org/10.1016/j.crad.2022.01.036
17. Takahashi N., Sugimoto M., Psutka S.P., Chen B., Moynagh M.R., Carter R.E. Validation study of a new semi-automated software program for CT body composition analysis. *Abdominal Radiology*, 2017, vol. 42, no. 9, pp. 2369–2375. https://doi.org/10.1007/s00261-017-1123-6
18. Kaur R., Juneja M., Mandal A.K. A comprehensive review of denoising techniques for abdominal CT images. *Multimedia Tools and Applications*, 2018, vol. 77, no. 17, pp. 22735–22770. https://doi.org/10.1007/s11042-017-5500-5
19. Masenko V.L., Kokov A.N., Grigoreva I.I., Krivoshapova K.E. Radiology methods of the sarcopenia diagnosis. *Research and Practical Medicine Journal*, 2019, vol. 6, no. 4, pp. 127–137. (in Russian). https://doi.org/10.17709/2409-2231-2019-6-4-13
20. Lee J.S., Kim Y.S., Kim E.Y., Jin W. Prognostic significance of CT-determined sarcopenia in patients with advanced gastric cancer. *PLoS ONE*, 2018, vol. 13, no. 8, pp. e0202700. https://doi.org/10.1371/journal.pone.0202700

Авторы

Теплякова Анастасия Романовна — преподаватель, аспирант, Обнинский институт атомной энергетики — филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский ядерный университет «МИФИ» (ИАТЭ НИЯУ МИФИ), Обнинск, 249039, Российская Федерация, sc 57220985322, https://orcid.org/0000-0001-8612-3850, anastasija-t23@mail.ru

Шершнев Роман Владимирович — старший преподаватель, Обнинский институт атомной энергетики — филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский ядерный университет «МИФИ» (ИАТЭ НИЯУ МИФИ), Обнинск, 249039, Российская Федерация, https://orcid.org/0009-0000-2324-5893, rvshershnev@yandex.ru

Старков Сергей Олегович — доктор физико-математических наук, профессор, старший научный сотрудник, Обнинский институт атомной энергетики — филиал федерального государственного автономного образовательного учреждения высшего образования «Национальный исследовательский ядерный университет «МИФИ» (ИАТЭ НИЯУ МИФИ), Обнинск, 249039, Российская Федерация, sc 6701907645, https://orcid.org/0000-0002-0420-7856, sergeystarkov56@mail.ru

Агабабян Татев Артаковна — кандидат медицинских наук, заведующая отделением, Медицинский радиологический научный центр им. А.Ф. Цыба — филиал ФГБУ «Национальный медицинский исследовательский центр радиологии» Минздрава России, Обнинск, 249036, Российская Федерация, sc 57202285176, https://orcid.org/0000-0002-9971-3451, tatevik.05@mail.ru

Кукарская Валерия Андреевна — клинический ординатор, Медицинский радиологический научный центр им. А.Ф. Цыба — филиал ФГБУ «Национальный медицинский исследовательский центр радиологии» Минздрава России, Обнинск, 249036, Российская Федерация, https://orcid.org/0009-0001-4872-3182, kukarely@mail.ru

Authors

Anastasia R. Teplyakova — PhD Student, Lecturer, Obninsk Institute for Nuclear Power Engineering, Obninsk, 249039, Russian Federation, sc 57220985322, https://orcid.org/0000-0001-8612-3850, anastasija-t23@mail.ru

Roman V. Shershnev — Senior Lecturer, Obninsk Institute for Nuclear Power Engineering, Obninsk, 249039, Russian Federation, https://orcid.org/0009-0000-2324-5893, rvshershnev@yandex.ru

Sergey O. Starkov — D.Sc. (Physics & Mathematics), Professor, Senior Researcher, Obninsk Institute for Nuclear Power Engineering, Obninsk, 249039, Russian Federation, sc 6701907645, https://orcid.org/0000-0002-0420-7856, sergeystarkov56@mail.ru

Tatev A. Agababian — PhD (Medicine), Head of Department, A. Tsyb Medical Radiological Research Centre — branch of the National Medical Research Radiological Center of the Ministry of Health of the Russia, Obninsk, 249036, Russian Federation, sc 57202285176, https://orcid.org/0000-0002-9971-3451, tatevik.05@mail.ru

Valeria A. Kukarskaya — Clinical Resident, A. Tsyb Medical Radiological Research Centre — branch of the National Medical Research Radiological Center of the Ministry of Health of the Russia, Obninsk, 249036, Russian Federation, https://orcid.org/0009-0001-4872-3182, kukarely@mail.ru

Статья поступила в редакцию 21.11.2023
Одобрена после рецензирования 25.12.2023
Принята к печати 24.01.2024

Received 21.11.2023
Approved after reviewing 25.12.2023
Accepted 24.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

МАТЕМАТИЧЕСКОЕ И КОМПЬЮТЕРНОЕ МОДЕЛИРОВАНИЕ

MODELING AND SIMULATION

doi: 10.17586/2226-1494-2024-24-1-133-143
УДК 531.383

Обеспечение режимов функционирования кориолисовых вибрационных гироскопов с низкодобротными резонаторами

Валерий Владимирович Матвеев[1], Владимир Владимирович Лихошерст[2], Алексей Владимирович Каликанов[3], Максим Георгиевич Погорелов[4]✉, Максим Дмитриевич Кирсанов[5], Сергей Владимирович Телухин[6]

[1,2,3,4,5,6] Тульский государственный университет, Тула, 300012, Российская Федерация

[1] matweew.valery@yandex.ru, https://orcid.org/0000-0001-8997-0297
[2] lvv_01@inbox.ru, https://orcid.org/0009-0009-4315-5997
[3] kalikanov.aleksei@mail.ru, https://orcid.org/0000-0002-1590-4278
[4] mgpogoreloff@yandex.ru✉, https://orcid.org/0009-0005-9731-3066
[5] kirsanovMD@yandex.ru, https://orcid.org/0009-0002-2872-9322
[6] teluhinserg@yandex.ru, https://orcid.org/0009-0001-2649-8420

Аннотация

Введение. Кориолисовые вибрационные гироскопы — класс перспективных инерциальных датчиков первичной информации, реагирующих на вращение основания резонатора посредством сил инерции Кориолиса, возникающих в вибрирующей оболочке. В настоящее время получили развитие два направления производства резонаторов таких гироскопов: из кварцевого стекла — материала с чрезвычайной малым внутренним трением и на основе обработки металлического сплава. При использовании первого направления, благодаря высокой добротности кварца, удается создавать интегрирующие гироскопы навигационного класса. Существующие образцы кориолисовых вибрационных гироскопов с металлическими резонаторами, как правило, представляют собой датчики угловой скорости. Проблема создания интегрирующего режима гироскопа с металлическим резонатором связана с низкой добротностью металлических сплавов, которая обычно не превышает 35 000. При таком значении добротности длительность функционирования гироскопа в режиме датчика угловых отклонений составит несколько секунд. В работе представлены способы обеспечения функционирования кориолисовых вибрационных гироскопов, включая режим интегрирующего гироскопа. **Метод.** Приведено математическое описание кориолисовых вибрационных гироскопов с цилиндрическим объемным резонатором на основе динамической модели доктора Д. Линча методом огибающих амплитуд колебаний. Математическая модель дополнена поправками, обеспечивающими компенсацию диссипации энергии колебаний резонатора для реализации интегрирующего режима гироскопа. Показаны условия полной компенсации рассеяния энергии колебаний. Представлено описание способов возбуждения стоячей волны в резонаторе с помощью вынуждающего периодического воздействия и путем создания автоколебаний. Показано, что длительность переходного процесса возбуждения определяется постоянной времени резонатора. **Основные результаты.** Приведены результаты экспериментальных исследований кориолисовых вибрационных гироскопов с низкодобротным металлическим резонатором, подтверждающие возможность реализации интегрирующего режима функционирования гироскопа. Первоначальное возбуждение колебаний резонатора осуществлялось автоколебательным контуром. Согласно результатам экспериментальных исследований, добротность металлического резонатора удалось повысить схемотехнически в 17 раз. В равной мере увеличено и время функционирования кориолисовых вибрационных гироскопов. **Обсуждение.** Теоретически и экспериментально показана возможность построения кориолисовых вибрационных гироскопов в режиме интегрирующего гироскопа на основе низкодобротного металлического резонатора. Решение данной проблемы основывалось на схемотехническом способе повышения добротности. Принципиально добротность резонатора может быть значительно повышена по сравнению с достигнутым в эксперименте показателем. Это обеспечит более длительное время функционирования кориолисовых вибрационных гироскопов в интегрирующем режиме.

Ключевые слова

кориолисовый вибрационный гироскоп, низкодобротный резонатор, режим функционирования, способы возбуждения

© Матвеев В.В., Лихошерст В.В., Каликанов А.В., Погорелов М.Г., Кирсанов М.Д., Телухин С.В., 2024

Благодарности
Работа выполнена при финансовой поддержке Министерства науки и высшего образования Российской Федерации в рамках государственного задания по теме FEWG-2022-0002.

Ссылка для цитирования: Матвеев В.В., Лихошерст В.В., Каликанов А.В., Погорелов М.Г., Кирсанов М.Д., Телухин С.В. Обеспечение режимов функционирования кориолисовых вибрационных гироскопов с низкодобротными резонаторами // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 133–143. doi: 10.17586/2226-1494-2024-24-1-133-143

Providing operating modes for Coriolis vibration gyroscopes with low-Q resonators

Valery V. Matveev[1], Vladimir V. Likhosherst[2], Alexey V. Kalikanov[3], Maxim G. Pogorelov[4]✉, Maxim D. Kirsanov[5], Sergey V. Telukhin[6]

[1,2,3,4,5,6] Tula State University, Tula, 300012, Russian Federation

[1] matweew.valery@yandex.ru, https://orcid.org/0000-0001-8997-0297
[2] lvv_01@inbox.ru, https://orcid.org/0009-0009-4315-5997
[3] kalikanov.aleksei@mail.ru, https://orcid.org/0000-0002-1590-4278
[4] mgpogoreloff@yandex.ru✉, https://orcid.org/0009-0005-9731-3066
[5] kirsanovMD@yandex.ru, https://orcid.org/0009-0002-2872-9322
[6] teluhinserg@yandex.ru, https://orcid.org/0009-0001-2649-8420

Abstract
Coriolis vibration gyroscopes are a class of promising inertial primary information sensors that respond to the rotation of the resonator base through Coriolis inertial forces arising in the vibrating shell. Currently, two directions for the production of resonators for such gyroscopes have been developed: from quartz glass, a material with extremely low internal friction, and based on the processing of a metal alloy. When using the first direction, thanks to the high quality factor of quartz, it is possible to create navigation-class integrating gyroscopes. Existing samples of Coriolis vibration gyroscopes with metal resonators, as a rule, are angular velocity sensors. The problem of creating an integrating mode of a gyroscope with a metal resonator is associated with the low quality factor of metal alloys which usually does not exceed 35,000. With this value of quality factor, the duration of operation of the gyroscope in the angular deviation sensor mode will be several seconds. The paper presents methods for ensuring the functioning of Coriolis vibration gyroscopes, including the integrating gyroscope mode. A mathematical description of Coriolis vibration gyroscopes with a cylindrical cavity resonator is given based on the dynamic model of Dr. D. Lynch using the method of envelope amplitudes of oscillations. The mathematical model is supplemented with corrections that provide compensation for the dissipation of the resonator oscillations energy to implement the integrating mode of the gyroscope. The conditions for complete compensation of vibration energy dissipation are shown. A description of methods for exciting a standing wave in a resonator using periodic forcing and by creating self-oscillations is presented. It is shown that the duration of the transient excitation process is determined by the time constant of the resonator. The results of experimental studies of Coriolis vibration gyroscopes with a low-Q metal resonator are presented confirming the possibility of implementing an integrating mode of operation of the gyroscope. The initial excitation of the resonator oscillations is carried out by a self-oscillating circuit. According to the results of experimental studies, the quality factor of the metal resonator was increased by a factor of 17. The operating time of Coriolis vibrating gyroscopes has been equally increased. The possibility of constructing Coriolis vibration gyroscopes in the integrating gyroscope mode based on a low-Q metal resonator has been shown theoretically and experimentally. The solution to this problem was based on a circuitry method for increasing the quality factor. In principle, the quality factor of the resonator can be significantly increased compared to the figure achieved in the experiment. This will ensure a longer operating time of Coriolis vibration gyroscopes in the integrating mode.

Keywords
Coriolis vibration gyroscope, low-Q resonator, operating mode, excitation methods

Acknowledgements
The work was supported financially by the Ministry of Science and Higher Education of the Russian Federation within the framework of the state assignment on the topic FEWG-2022-0002.

For citation: Matveev V.V., Likhosherst V.V., Kalikanov A.V., Pogorelov M.G., Kirsanov M.D., Telukhin S.V. Providing operating modes for Coriolis vibration gyroscopes with low-Q resonators. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 133–143 (in Russian). doi: 10.17586/2226-1494-2024-24-1-133-143

Введение

В настоящее время перспективными инерциальными датчиками первичной информации, чувствительными к вращению основания резонатора посредством сил инерции Кориолиса, являются вибрационные гироскопы, выполненные на основе кремниевой технологии микро-электромеханических систем (МЭМС) [1, 2], высокоточной обработки кварцевого стекла [3, 4] или металлического сплава [5–8]. Для производства МЭМС-гироскопов требуются специализированные технологии поверхностной микрообработки, глубинного травления кремния, литографии, осаждения и др. [9]. Обработка кварцевого стекла весьма трудоемка и для достижения

высокой добротности резонатора требуется применение бесконтактных датчиков и устройств управления стоячей волной [10, 11]. Простота механической обработки металлических сплавов и отсутствие необходимости применения специализированного технологического оборудования делает привлекательным применение металлических резонаторов в кориолисовых вибрационных гироскопах (КВГ) [12].

Проблема создания КВГ с металлическим объемным резонатором связана с низкой добротностью металлических сплавов, которая обычно не превышает 35 000. Для сравнения добротность кварцевых полусферических резонаторов составляет от единиц миллионов и выше [13]. В связи с этим на базе КВГ с низкодобротными резонаторами, как правило, реализуются датчики угловой скорости (ДУС). При этом теоретическое обоснование режимов возбуждения и получения информации в основном направлено на КВГ с кварцевым полусферическим резонатором [14]. Отметим, что аналитическое описание с экспериментальным подтверждением возбуждения свободной волны автоколебательным контуром в научных работах весьма ограничено. Публикации по реализации КВГ с низкодобротным резонатором, функционирующим в качестве интегрирующего гироскопа (ИГ), также практически отсутствуют. В связи с этим целью настоящей работы является систематизация описания режимов функционирования КВГ с низкодобротным резонатором в качестве дифференцирующего гироскопа и ИГ в виде дифференциальных уравнений, передаточных функций и характеристик, связывающих значение угловой скорости основания и выходного сигнала.

Типовая конструкция и принцип действия КВГ

Типовая конструкция КВГ с объемным резонатором содержит, как правило, цилиндрический чувствительный элемент (рис. 1, *a*), на дно которого приклеиваются четыре пары пьезоэлементов. Резонатор устанавливается на основание, содержащее выводы с пьезоэлементов [7]. Для обнаружения угловой скорости (или угла) поворота основания Ω используется свойство инерции стоячей волны, возбуждаемой в резонаторе путем подачи переменного напряжения на одну из пар пьезоэлементов, например, находящихся вдоль оси x_1 (рис. 1, *b*). Точки стоячей волны с максимальной амплитудой радиальных колебаний являются пучностями, а с минимальной — узлами. Стоячая волна для второй формы колебаний, являющейся рабочей формой колебаний в КВГ, имеет четыре узла и четыре пучности.

Если основание, на котором находится резонатор, не вращается ($\Omega = 0$), то волновая картина неизменна (рис. 1, *b*). При этом сигналы, снимаемые с пары пьезоэлементов, расположенных вдоль оси x_2 и находящихся под углом 45° к оси возбуждения, близки к нулю. Если основание вращается с угловой скоростью Ω, то волновая картина прецессирует (запаздывает) от направления возбуждения на некоторый угол, и сигнал, снимаемый с пьезоэлементов, установленных вдоль оси x_2, становится отличен от нуля и определяется величиной угловой скорости. В зависимости от способа возбуждения КВГ может являться ДУС или ИГ. Две оставшиеся пары пьезоэлементов служат для контроля резонансного режима возбуждения и компенсации прецессии стоячей волны.

Математическая модель КВГ

Строгая теория КВГ с объемными резонаторами достаточно полно изложена в фундаментальных работах [4, 10, 11]. Для инженерного анализа функционирования КВГ с различной формой чувствительного элемента широко применяется математическая модель, описывающая колебания резонатора вдоль осей контроля x_1 и x_2. Модель при постоянной угловой скорости основания может быть представлена в виде [12, 15]:

$$\begin{cases} \ddot{x}_1 + \dfrac{\omega_0}{Q}\dot{x}_1 + \omega_0^2 x_1 - 2K\Omega\dot{x}_2 = f_1\left(t, x_1, \dot{x}_1, \int\limits_t x_1 dt, \ldots\right) \\ \ddot{x}_2 + \dfrac{\omega_0}{Q}\dot{x}_2 + \omega_0^2 x_2 + 2K\Omega\dot{x}_1 = f_2\left(t, x_2, \dot{x}_2, \int\limits_t x_2 dt, \ldots\right) \end{cases}, \quad (1)$$

где ω_0 — круговая частота собственных колебаний; Ω — угловая скорость вращения основания; K — коэффициент прецессии волны; $f_i\left(t, x_i, \dot{x}_i, \int\limits_t x_i dt, \ldots\right)$ — функция управления ($i = 1, 2$); Q — добротность колебаний.

Математическая модель (1) записана для постоянной угловой скорости основания и при отсутствии разночастотности и разнодобротности.

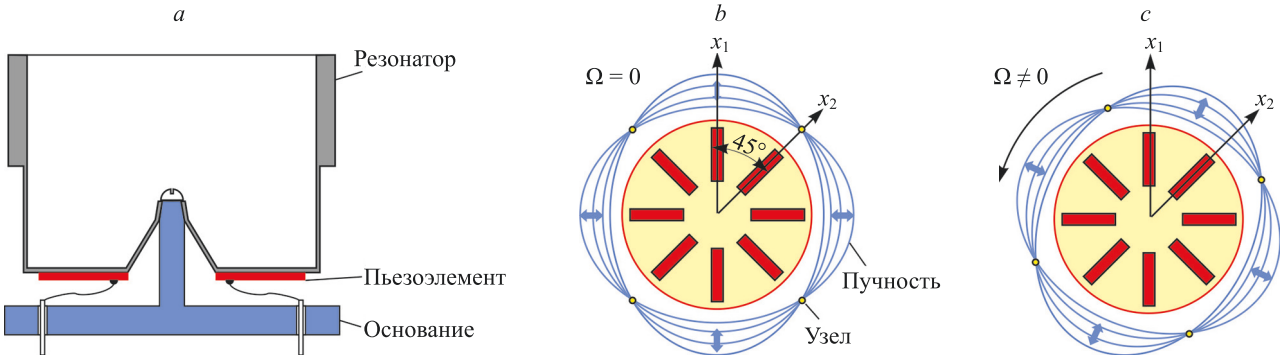


Рис. 1. Типовая схема кориолисового вибрационного гироскопа (*a*), положение стоячей волны при неподвижном (*b*) и повороте основания (*c*) резонатора

Fig. 1. Typical diagram of a Coriolis vibration gyroscope (*a*), position of a standing wave when the resonator base is stationary (*b*) and rotated (*c*)

Возбуждение стоячей волны

Возбуждение стоячей волны осуществляется на собственной частоте второй формы колебаний. С учетом конструктивных особенностей цилиндрических резонаторов КВГ возможны следующие способы возбуждения:

— формированием вынужденных колебаний резонатора генератором сигнала синусоидальной или прямоугольной формы с фазовой автоподстройкой частоты (рис. 2, *a*);
— путем формирования автоколебаний, реализуемых нелинейной обратной связью (рис. 2, *b*).

В первом случае на дно резонатора со стороны пьезоэлементов прикладывается воздействие на собственной частоте . В этом случае функции управления имеют вид:

$$f_1(\ldots) = f_0 \sin \omega_0 t, f_2(\ldots) = 0. \quad (2)$$

Для способа управляющих воздействий в форме (2) и отсутствии вращения основания ($\Omega = 0$), решение первого уравнения системы (1) при $Q >> 1$ в резонансном режиме с высокой степенью точности можно представить в виде:

$$x_1(t) \approx f_0 \frac{Q}{\omega_0^2}\left(1 - e^{-\frac{\omega_0}{2Q}t}\right)\sin(\omega_0 t - \pi/2). \quad (3)$$

Отметим, что для контроля резонансного режима необходимо контролировать фазовое запаздывание отклика резонатора, которое должно составлять минус 90° по отношению к сигналу возбуждения (рис. 2, *a*).

Из решения выражения (3) следует, что огибающая амплитуд колебаний при возбуждении резонатора на собственной частоте имеет апериодический характер, определяемый постоянной времени

$$T = 2Q/\omega_0. \quad (4)$$

Практическую ценность при исследовании КВГ имеет построение математической модели по огибающим амплитуд колебаний (*amplitude envelope*) резонатора. Подобные модели позволяют проводить анализ и синтез КВГ без учета высокочастотной несущей колебаний резонатора [16]. На основе решения (3) огибающая амплитуд колебаний при совпадении частоты возбуждения с собственной частотой резонатора соответствует апериодическому звену первого порядка с передаточной функцией вида (рис. 3):

$$W_1^{env}(s) = \frac{x_1^{env}(s)}{f_0(s)} = \frac{T}{2\omega_0(Ts+1)} = \frac{Q}{\omega_0^2(Ts+1)}, \quad (5)$$

где s — аргумент в преобразовании Лапласа.

Исходя из свойств передаточной функции (5), время переходного процесса может быть оценено по следующей зависимости

$$t_{\Pi} \approx 3T = 6Q/\omega_0.$$

Для рассматриваемого класса резонаторов собственная частота обычно составляет величину порядка 6000 Гц, а добротность — 30 000. Для данных параметров постоянная времени резонатора принимает значение $T \approx 1{,}6$ с, тогда время переходного процесса составляет около 5 с. Таким образом, без специальных мер длительность процесса возбуждения может составлять значительную величину.

Вместо синусоидального сигнала, стоячую волну можно возбуждать прямоугольными импульсами амплитудой f_0 и круговой частотой ω_0. Известно, что при разложении прямоугольных импульсов в ряд Фурье, амплитуда первой гармоники будет составлять величину $4f_0/\pi$. Допустим, что высшие гармоники ослабляются резонатором, тогда огибающую амплитуду колебаний резонатора опишем передаточной функцией вида:

$$W_1^{env}(s) = \frac{x_1^{env}(s)}{f_0(s)} = \frac{4Q}{\pi\omega_0^2(Ts+1)}. \quad (6)$$

При организации автоколебательного режима возбуждения в контуре должен присутствовать источник питания постоянного тока и нелинейный элемент, регулирующий приток энергии. Возможны два варианта

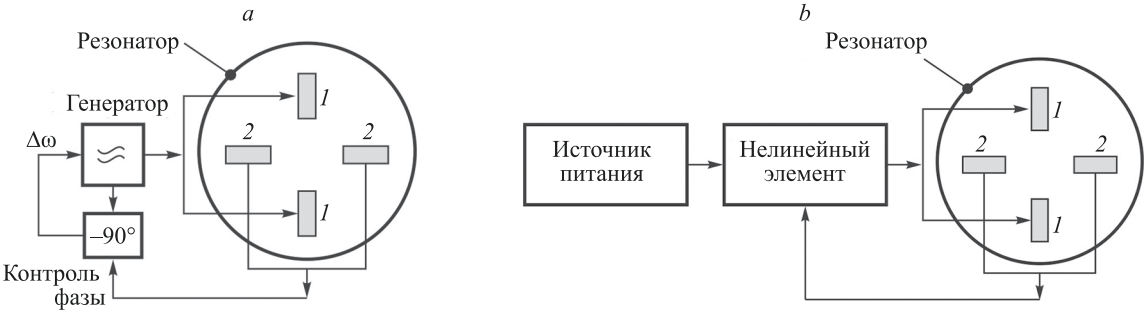


Рис. 2. Способы возбуждения стоячей волны в резонаторе кориолисового вибрационного гироскопа при формировании: вынужденных колебаний резонатора (*a*) и автоколебаний (*b*).

1 — пьезоэлементы возбуждения; *2* — пьезоэлементы обратной связи; Δω — приращение частоты сигнала генератора

Fig. 2. Methods for excitation of a standing wave in the resonator of a Coriolis vibration gyroscope during the formation of forced oscillations of the resonator (*a*) and self-oscillations (*b*).

1 — piezoelectric excitation elements; *2* — feedback piezoelements; Δω — generator signal frequency increment

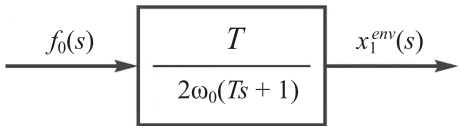


Рис. 3. Структурная схема контура первичных колебаний кориолисового вибрационного гироскопа в условиях резонансной настройки

Fig. 3. Block diagram of the primary oscillation circuit of a Coriolis vibrating gyroscope under resonant tuning conditions

реализации автоколебательного режима возбуждения КВГ. В первом случае возбуждение автоколебаний реализуется в соответствии с уравнением вида:

$$\ddot{x}_1 + \frac{2}{T}\dot{x}_1 + \omega_0^2 x_1 - f_0 \operatorname{sign}(\dot{x}_1) = 0, \qquad (7)$$

где f_0 — амплитуда удельной силы трения.

Исследование уравнения (7) показало [17], что частота автоколебаний практически равна собственной частоте резонатора ω_0, а огибающая амплитуд колебаний носит апериодический характер с постоянной времени, определяемой соотношением (4).

Амплитуду установившихся автоколебаний оценим по следующей зависимости

$$x_1^{\text{уст}} \approx \frac{4f_0 Q}{\pi\omega_0^2}, \qquad (8)$$

что равным образом соответствует амплитуде вынужденных резонансных колебаний, возбуждаемых прямоугольными импульсами. В связи с этим динамику огибающей амплитуд автоколебаний можно также описать передаточной функцией апериодического звена первого порядка (6).

К аналогичному результату приводит возбуждение автоколебаний резонатора в соответствии с уравнением:

$$\ddot{x}_1 + \frac{2}{T}\dot{x}_1 + \omega_0^2 x_1 + f_0 \operatorname{sign}\left(\int_0^t x_1 dt\right) = 0, \qquad (9)$$

отличающимся от (7) наличием интеграла от перемещения резонатора под знаком сигнатуры, что является более рациональным с точки зрения снижения шумов в цепи возбуждения. При этом частота автоколебаний также соответствует собственной частоте ω_0, а амплитуда может быть оценена по зависимости (8).

Режим КВГ–ДУС

В режиме ДУС к резонатору прикладывается возбуждающее воздействие на собственной частоте колебаний. В разделе «Возбуждение стоячей волны» показано, что возбуждение резонатора прямоугольными импульсами или автоколебаниями не вносит существенных отличий по сравнению с возбуждением синусоидальным сигналом. В связи с этим уравнения КВГ на вращающемся с угловой скоростью Ω основании запишем в виде:

$$\begin{cases} \ddot{x}_1 + \frac{2}{T}\dot{x}_1 + \omega_0^2 x_1 - 2K\Omega\dot{x}_2 = f_0 \sin\omega_0 t, \\ \ddot{x}_2 + \frac{2}{T}\dot{x}_2 + \omega_0^2 x_2 + 2K\Omega\dot{x}_1 = 0. \end{cases} \qquad (10)$$

При необходимости учета автоколебательного способа возбуждения к величине f_0 необходимо добавить множитель $4/\pi$. Структурная схема КВГ по огибающим амплитуд колебаний резонатора приведена на рис. 4.

Структурная схема соответствует случаю постоянства угловой скорости вращения основания и обеспечению резонансной настройки, что позволяет исследовать КВГ без учета высокочастотной несущей колебаний резонатора. Блок с коэффициентом передачи $4/\pi$ отмечен пунктирной линией, так как должен присутствовать при возбуждении стоячей волны прямоугольными импульсами или в автоколебательном режиме. При организации компенсационного режима функционирования КВГ вводится обратная связь с передаточной функцией $W_k(s)$. В этом случае выходной сигнал КВГ x_{out}^{env} привлекается из цепи обратной связи.

Передаточные функции по огибающим амплитуд колебаний резонатора прямого измерения по осям контроля x_1 и x_2, полученные в соответствии со структурной схемой, при отсутствии блока $W_k(s)$, имеют вид:

$$W_1(s) = \frac{x_1^{env}(s)}{f_0(s)} = \frac{T(Ts+1)}{2\omega_0[(Ts+1)^2 + K^2T^2\Omega^2]}, \qquad (11)$$

$$W_2(s) = \frac{x_2^{env}(s)}{f_0(s)} = -\frac{KT^2\Omega^2}{2\omega_0[(Ts+1)^2 + K^2T^2\Omega^2]}. \qquad (12)$$

Из анализа знаменателей передаточных функций (11) и (12) сделаем следующие выводы:

— наличие величины Ω^2 показывает нелинейную зависимость характеристик КВГ от угловой скорости основания;

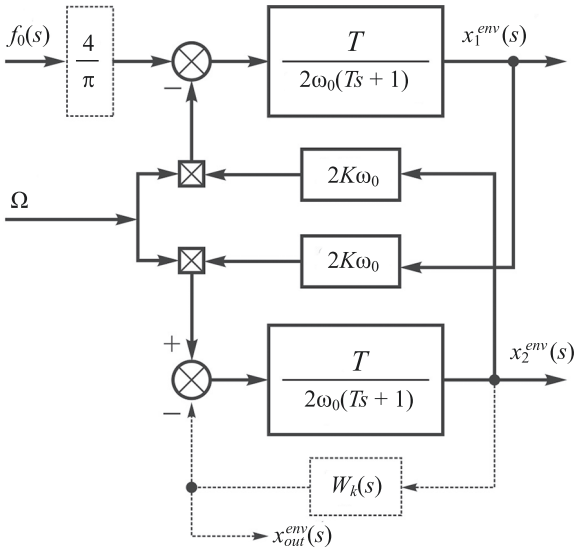


Рис. 4. Структурная схема кориолисового вибрационного гироскопа по огибающим амплитуд колебаний резонатора

Fig. 4. Block diagram of a Coriolis vibration gyroscope based on the envelopes of the resonator oscillation amplitudes

— переходный процесс огибающей амплитуд колебаний по координате x_2 описывается колебательным звеном с постоянной времени

$$T_2 = T/\sqrt{1 + K^2T^2\Omega^2},$$

которая для угловых скоростей основания до 50 °/с сопоставима с постоянной времени T, что делает процесс измерения угловой скорости достаточно инерционным.

В соответствии с передаточными функциями (11), (12), установившиеся амплитуды колебаний резонатора по координатам x_1 и x_2 определим выражениями

$$x_1^a = \frac{Tf_0}{2\omega_0(1 + K^2T^2\Omega^2)}, \quad (13)$$

$$x_2^a = \frac{KT^2f_0}{2\omega_0}\frac{\Omega}{1 + K^2T^2\Omega^2}. \quad (14)$$

Рассмотрим нелинейную зависимость амплитуд колебаний резонатора (13) и (14) (рис. 5) с помощью графиков зависимостей установившихся амплитуд колебаний резонатора КВГ с параметрами: $f_0 = 0{,}1$ м/с^2, $\omega_0 = 6000$ Гц, $Q = 30\,000$, $K = 0{,}4$, $T = 1{,}6$ с. Таким образом, по мере возрастания угловой скорости Ω амплитуда колебаний в цепи возбуждения снижается. При этом амплитуда колебаний по оси x_2 возрастает при достижении максимальной угловой скорости Ω_m, а затем начинает снижаться. Выполним анализ соотношения (14) на экстремум. В результате получим, что амплитуда колебаний резонатора КВГ максимальна при угловой скорости основания

$$\Omega_m = \frac{1}{TK}. \quad (15)$$

При угловой скорости (15) амплитуды колебаний x_1^a и x_2^a равны. Отношение амплитуд линейно зависит от угловой скорости основания

$$x_2^a/x_1^a = KT\Omega. \quad (16)$$

Таким образом, при функционировании КВГ прямого измерения в режиме ДУС угловую скорость основания можно определить как отношение амплитуд (16) установившихся колебаний резонатора.

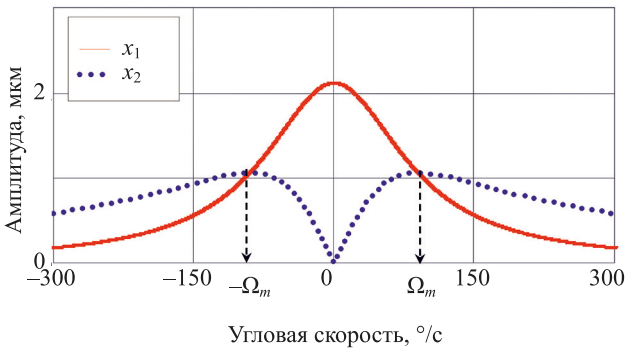


Рис. 5. Зависимость установившихся амплитуд колебаний кориолисового вибрационного гироскопа по осям контроля x_1 и x_2

Fig. 5. Dependence of the steady-state oscillation amplitudes of the Coriolis vibration gyroscope along the control axes x_1 and x_2

Для улучшения динамических характеристик КВГ и обеспечения линейности выходной характеристики обычно реализуется компенсационный режим функционирования ДУС, при котором за счет реализации цепи обратной связи осуществляется компенсация действия кориолисовых сил инерции, вызывающих прецессию стоячей волны.

В работе [18] показано, что «при построении контура компенсации с управлением по огибающей модулированного сигнала при сохранении резонансной настройки может быть обеспечена требуемая полоса рабочих частот без снижения чувствительности к угловой скорости основания». При этом при построении контура компенсации кориолисова момента гироскопа использован пропорциональный регулятор (П-регулятор).

В [12] в контуре компенсации кориолисового ускорения и квадратурной составляющей применен пропорционально-интегральный регулятор (ПИ-регулятор), а в [19] для управления вторичными колебаниями резонатора — пропорционально-интегрально-дифференцирующий регулятор (ПИД-регулятор), с помощью которого возможно реализовать полосу пропускания в 100–200 Гц. Однако в работах [12, 18, 19] математическая модель для огибающих амплитуд колебаний резонатора рассматривается без учета члена $2K\Omega\dot{x}_2$ в уравнениях движения (10).

Для учета компенсационного режима КВГ–ДУС структурная схема (рис. 4) должна быть дополнена цепью обратной связи, включающей звено с передаточной функцией $W_k(s)$.

Рассмотрим случай коррекции КВГ ПИ-регулятором с передаточной функцией

$$W_k(s) = k_p + \frac{k_i}{s} = \frac{k_i(T_ps + 1)}{s},$$

где k_p, k_i — коэффициенты передачи ПИ-регулятора; $T_p = k_p/k_i$ — постоянная времени корректирующего устройства.

В этом случае передаточные функции огибающих колебаний резонатора примут вид

$$W_1(s) = \frac{x_1^{env}(s)}{f_0(s)} = \frac{T[2\omega_0 s(Ts + 1) + Tk_i(T_ps + 1)]}{2\omega_0\{(Ts + 1)[2\omega_0 s(Ts + 1) + Tk_i(T_ps + 1)] + 2\omega_0K^2T^2\Omega^2s\}},$$

$$W_2(s) = \frac{x_2^{env}(s)}{f_0(s)} = \frac{\Omega KT^2s}{(Ts + 1)[2\omega_0 s(Ts + 1) + Tk_i(T_ps + 1)] + 2\omega_0K^2T^2\Omega^2s},$$

из которых следует, что слагаемые, содержащие квадрат угловой скорости основания резонатора Ω^2, снабжены множителем в виде оператора Лапласа s. В установившемся режиме значения амплитуд колебаний резонатора в точках контроля определим соотношениями:

$$x_1^a = \frac{f_0T}{2\omega_0}, \; x_2^a = 0. \quad (17)$$

Из соотношений (17) получим, что установившееся значение амплитуды колебаний по координате x_1 не зависит от угловой скорости основания, а по координате x_2 стремится к нулю. Если в качестве выходного сигнала КВГ привлекается сигнал x_{out}^{env} (рис. 4), то ему соответствует передаточная функция:

$$W_{out}(s) = \frac{x_{out}^{env}(s)}{f_0(s)} = \\ = \frac{\Omega K T^2 k_i (T_p s + 1)}{(Ts + 1)[2\omega_0 s(Ts + 1) + Tk_i(T_p s + 1)] + 2\omega_0 K^2 T^2 \Omega^2 s}.$$

Установившееся значение амплитуды данного выходного сигнала определим в виде:

$$x_{out}^{a} = f_0 KT\Omega. \qquad (18)$$

На основании выражения (18) получим линейную зависимость от измеряемой угловой скорости основания.

Предположим, что $T_p = T$, то передаточную функцию выходного сигнала КВГ–ДУС можно приближенно записать в виде

$$W_{out}(s) = \frac{x_{out}(s)}{\Omega(s)} \approx \frac{f_0 KT}{T_c s + 1},$$

где $T_c = \dfrac{2\omega_0}{Tk_i}$ — постоянная времени КВГ с контуром компенсации кориолисова ускорения, которая может быть достаточно малой за счет соответствующего выбора коэффициента k_i.

Отметим, что использование П-регулятора (пропорционального закона регулирования) в цепи обратной связи не устраняет зависимость выходного сигнала от квадрата угловой скорости основания.

Режим КВГ–ИГ

Рассмотрим свободные колебания резонатора, при которых отсутствует вынуждающее воздействие со стороны пьезоэлементов

$$f_1(\ldots) = 0, f_2(\ldots) = 0.$$

В этом случае математическая модель КВГ приобретает вид:

$$\begin{cases} \ddot{x}_1 + \dfrac{2}{T}\dot{x}_1 - 2K\Omega\dot{x}_2 + \omega_0^2 x_1 = 0, \\ \ddot{x}_2 + \dfrac{2}{T}\dot{x}_2 + 2K\Omega\dot{x}_1 + \omega_0^2 x_2 = 0. \end{cases} \qquad (19)$$

Уравнения движения (19) идентичны уравнениям колебаний маятника Фуко [20] с учетом диссипации энергии. Введением комплексной переменной $z = x_1 + jx_2$ система (19) может быть преобразована к одному дифференциальному уравнению

$$\ddot{z} + \left(\frac{2}{T} + 2jK\Omega\right)\dot{z} + \omega_0^2 z = 0. \qquad (20)$$

Примем угловую скорость основания постоянной и решим уравнение (20) при начальных условиях:

$$z(0) = z_0 = x_1^0,$$

$$\dot{z}(0) = 0,$$

где x_1^0 — начальные отклонения резонатора по координате x_1.

Переходя затем от переменной z к x_1 и x_2, решение системы уравнений (19) запишем в виде:

$$x_1(t) \approx x_1^0 e^{-t/T}\cos\omega_0 t\cos K\Omega t,$$

$$x_2(t) \approx -x_1^0 e^{-t/T}\cos\omega_0 t\sin K\Omega t.$$

Найдем отношение переменных $x_2(t)/x_1(t)$ и определим угол поворота основания резонатора:

$$\Omega t = -\frac{1}{K}\operatorname{arctg}\frac{x_2(t)}{x_1(t)}, \qquad (21)$$

из которого возможно получить принципиальную реализацию КВГ в режиме ИГ. Предположим, что для условия $K = 1$ соотношение (21) соответствует повороту плоскости колебаний маятника Фуко. Так как ось x_2 колебаний резонатора располагается под углом 45° к оси x_1 (а не под 90°, как у маятника Фуко), то для правильной интерпретации измерений соотношение (21) необходимо умножить на 1/2.

Для обеспечения заданной продолжительности работы в режиме КВГ–ИГ осуществим компенсацию диссипации энергии, связанной с наличием членов $2\dot{x}_1/T$, $2\dot{x}_2/T$ в уравнениях (19). Возможны два варианта реализации такой компенсации. При первом варианте — пьезоэлементы КВГ должны формировать сигналы, пропорциональные скоростям движения резонатора $\dot{x}_1$ и $\dot{x}_2$, но с противоположным знаком. Тогда математическая модель КВГ примет вид:

$$\begin{cases} \ddot{x}_1 + \dfrac{2}{T}\dot{x}_1 + \omega_0^2 x_1 - 2K\Omega\dot{x}_2 - k_d\dot{x}_1 = 0, \\ \ddot{x}_2 + \dfrac{2}{T}\dot{x}_2 + \omega_0^2 x_2 + 2K\Omega\dot{x}_1 - k_d\dot{x}_2 = 0, \end{cases} \qquad (22)$$

где k_d — коэффициент пропорциональности производных от перемещения резонатора.

Полная компенсация диссипации энергии будет выполнена при равенстве:

$$k_d = 2T^{-1}.$$

Недостатком модели (22) является необходимость формирования производных от перемещений резонатора $\dot{x}_1$ и $\dot{x}_2$, что может привести к возрастанию шума в цепях компенсации.

Второй вариант компенсации диссипации энергии связан с формированием воздействия на резонатор КВГ сигналами, пропорциональными интегралам от перемещений x_1 и x_2. Математическую модель КВГ представим системой уравнений:

$$\begin{cases} \ddot{x}_1 + \dfrac{2}{T}\dot{x}_1 + \omega_0^2 x_1 - 2K\Omega\dot{x}_2 + k_i\int\limits_0^t x_1 dt = 0, \\ \ddot{x}_2 + \dfrac{2}{T}\dot{x}_2 + \omega_0^2 x_2 + 2K\Omega\dot{x}_1 + k_i\int\limits_0^t x_2 dt = 0, \end{cases} \quad (23)$$

где k_i — коэффициент пропорциональности интегральных членов.

При этом полная компенсация диссипации энергии имеет вид:

$$k_i = \frac{2\omega_0^2}{T} = \frac{\omega_0^3}{Q}.$$

Отметим, что при $k_i > \omega_0^3/Q$ колебательный процесс будет расходящимся.

Результаты экспериментальных исследований

Выполним исследования на натурном макете КВГ с металлическим резонатором, собственная частота которого составила около 5750 Гц, при возбуждении свободной волны за счет реализации автоколебаний в соответствии с уравнением (9). После получения автоколебаний КВГ был переведен в режим свободной волны с компенсацией диссипации энергии колебаний в соответствии с уравнениями (23). Для реализации требуемого интегрирующего устройства использована аналоговая схема на операционном усилителе (рис. 6).

Интегрирующее устройство описано передаточной функцией вида:

$$W(s) = \frac{ks}{(T_1 s + 1)(T_2 s + 1)},$$

где $T_1 = R_1C_1$, $T_2 = R_2C_2$ — постоянные времени; $k = R_2C_1$ — коэффициент усиления.

Для данной собственной частоты колебаний резонатора обеспечивались постоянные времени: $T_1 = 0{,}1$ мс, $T_2 = 1{,}6$ мс.

На рис. 7 приведены амплитудная и фазовая частотные характеристики используемого интегрирующего устройства.

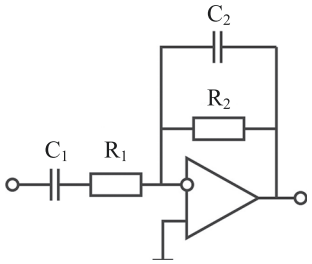


Рис. 6. Схема интегрирующего устройства
Fig. 6. Integrating device diagram

Собственная частота резонатора лежит в области спада амплитудной частотной характеристики, имеющей наклон минус 20 дБ/дек, что соответствует интегрирующему звену. Функция sign () реализована также на операционном усилителе с помощью схемы компаратора. Получены экспериментальная осциллограмма процесса возбуждения автоколебаний резонатора (рис. 8, *а*), установившиеся автоколебания резонатора и сигнал на выходе компаратора (рис. 8, *b*).

На рис. 9 приведены экспериментальные графики затухающих колебаний по координате x_1 при неподвижном основании без и с компенсацией диссипации энергии колебаний. Кроме того, на рис. 9 приведены огибающие амплитуд колебаний резонатора, построенные на основе импульсной переходной характеристики передаточной функции (5).

Постоянная времени металлического резонатора без компенсации энергии составила $T \approx 1$ с, что при собственной частоте резонатора 5750 Гц соответствует добротности около 18 000. Вместе с тем постоянная времени резонатора с контурами компенсации диссипации энергии имеет значение $T_e \approx 17$ с, что соответствует эквивалентной добротности 307 091. Таким образом, в 17 раз возросло эффективное увеличение добротности, а также увеличение времени функционирования до значения порядка $3T_e \approx 54$ с.

На рис. 10 приведены фигуры Лиссажу, полученные после возбуждения свободной волны, и ожидания в течение 3 с (основание было неподвижно) с последую-

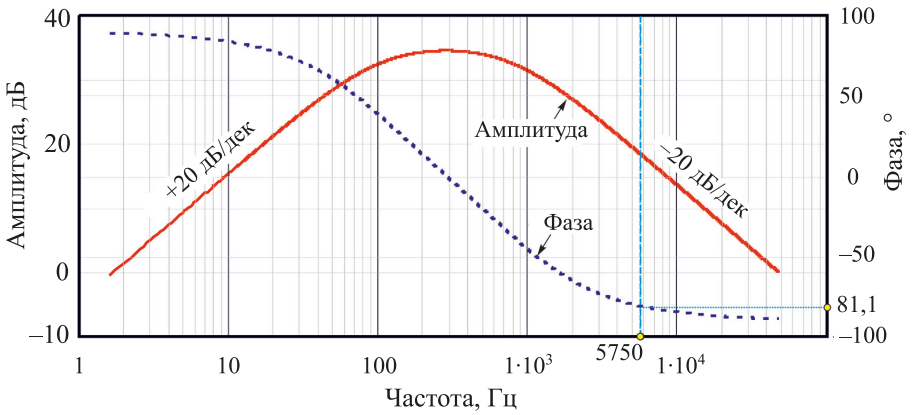


Рис. 7. Амплитудная и фазовая частотные характеристики интегрирующего устройства и его фазовое запаздывание 81°6′ на собственной частоте резонатора 5770 Гц

Fig. 7. Amplitude and phase frequency characteristics of the integrating device and its phase lag of 81°6′ at the natural frequency of the resonator 5770 Hz

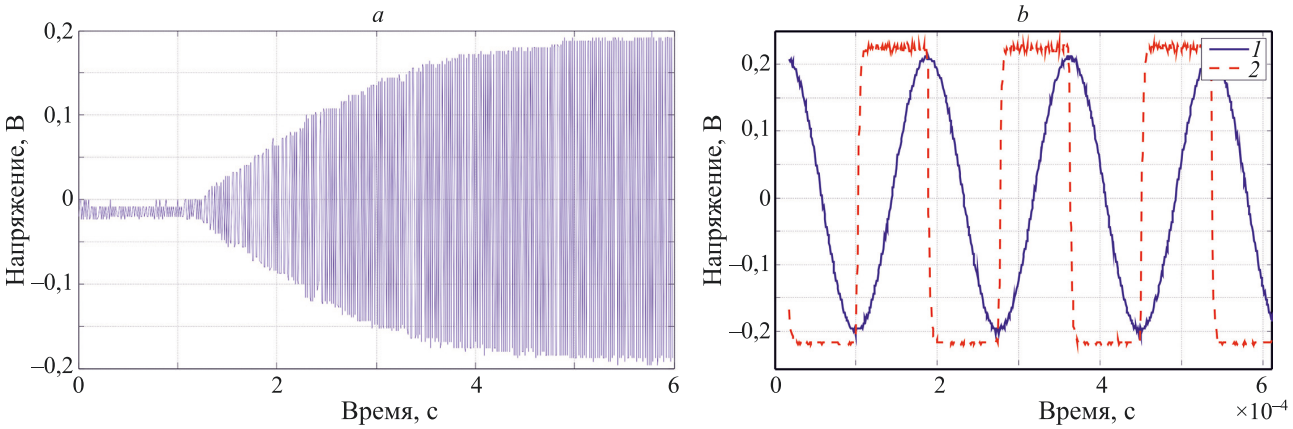


Рис. 8. Экспериментальная осциллограмма процесса возбуждения автоколебаний резонатора (*a*); установившиеся автоколебания резонатора (кривая *1*) и сигнал на выходе компаратора (кривая *2*) (*b*)

Fig. 8. Experimental oscillogram of the process of excitation of self-oscillations of the resonator (*a*); steady-state self-oscillations of the resonator (curve *1*) and the signal at the output of the comparator (curve *2*) (*b*)

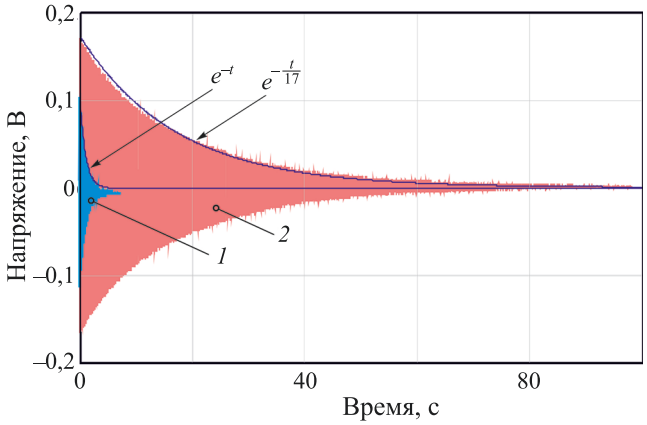


Рис. 9. Графики затухающих колебаний резонатора: без (кривая *1*) и с компенсацией (кривая *2*) диссипации энергии

Fig. 9. Graphs of damped oscillations of the resonator: without compensation (*1*) and with compensation for energy dissipation (*2*)

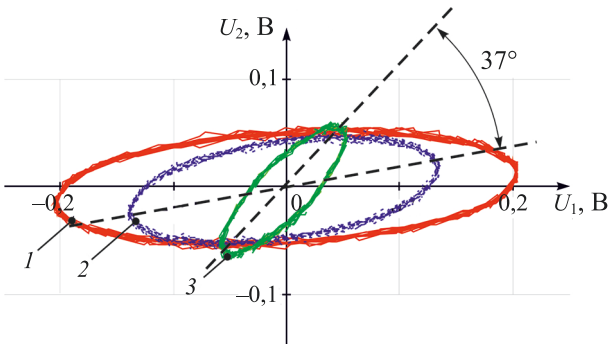


Рис. 10. Фигуры Лиссажу для кориолисового вибрационного гироскопа в режиме интегрирующего гироскопа: *1* — в момент запуска; *2* — после ожидания в течение 3 с; *3* — после поворота основания на угол 45°

Fig. 10. Lissajous figures for a Coriolis vibrating gyroscope in the integrating gyroscope mode: *1* — at the moment of launch; *2* – after waiting for 3 s; *3* — after rotating the base at an angle of 45°

щим поворотом основания на угол 45°. Напряжение U_1 соответствует сигналу, снимаемому с пьезоэлементов, установленных вдоль оси x_1, U_2 — с пьезоэлементов вдоль оси x_2.

Для коэффициента прецессии волны $K \approx 0{,}4$ измеренный угол поворота основания составил $37°/(2 \cdot 0{,}4) \approx 46°15'$. Данный результат следует считать предварительным, так как получен при реализации контура управления КВГ на макетных платах и требует дополнительной проработки. Тем не менее, проведенный эксперимент подтверждает возможность создания свободной волны в низкодобротном резонаторе путем компенсации диссипации энергии колебательного процесса и функционирования КВГ в режиме ИГ.

Заключение

Выполнено математическое описание режимов функционирования кориолисового вибрационного гироскопа с низкодобротным металлическим резонатором в качестве датчика угловой скорости и интегрирующего гироскопа посредством дифференциальных уравнений, структурных схем и передаточных функций. Представлены оценки амплитуд установившихся колебаний при возбуждении стоячей волны на резонансной частоте с помощью генератора синусоидальных колебаний и методом автоколебаний. Установлена нелинейная зависимость амплитуд колебаний резонатора от угловой скорости основания при функционировании режима датчика угловой скорости. Показано, что при достижении угловой скорости Ω_m амплитуды установившихся колебаний резонатора в точках контроля равны между собой. Установлено, что при функционировании гироскопа в режиме прямого измерения в качестве информационного параметра, функционально зависящего от угловой скорости основания, целесообразно использовать отношение амплитуд колебаний x_2^a/x_1^a. Приведено теоретическое обоснование использования пропорционально-интегрирующего регулятора при реализации гироскопа в режиме датчика угловой скорости. Дано математическое описание гироскопа в качестве

измерителя угловых перемещений при свободных колебаниях резонатора. Показано, что без специальных мер функционирование гироскопа с низкодобротным резонатором в интегрирующем режиме возможно только в течение нескольких секунд. Рассмотрены способы компенсации диссипации энергии при свободных колебаниях резонатора. Получены результаты экспериментальных исследований гироскопа с низкодобротным металлическим резонатором в режиме датчика угловых перемещений. Согласно результатам экспериментальных исследований, добротность металлического резонатора удалось повысить схемотехнически в 17 раз. В равной мере увеличено и время функционирования кориолисовых вибрационных гироскопов в интегрирующем режиме.

Литература

1. Вавилов В.Д., Тимошенков С.П., Тимошенков А.С. Микросистемные датчики физических величин: монография. М.: Техносфера, 2018. 550 с.
2. Пешехонов В.Г. Перспективы развития гироскопии // Гироскопия и навигация. 2020. Т. 28. № 2(109). С. 3–10. https://doi.org/10.17285/0869-7035.0028
3. Лунин Б.С. Научно-технологические основы разработки полусферических резонаторов волновых твердотельных гироскопов: автореферат диссертации на соискание ученой степени доктора технических наук. М.: Изд-во МГУ, 2006. 32 с.
4. Лунин Б.С., Матвеев В.А., Басараб М.А. Волновой твердотельный гироскоп. Теория и технология. М.: Радиотехника, 2014. 176 с.
5. Распопов В.Я., Егоров С.В. Проектирование резонатора волнового твердотельного гироскопа и верификация конечно-элементной модели // Известия ТулГУ. Технические науки. 2021. № 10. С. 227–235.
6. Распопов В.Я., Лихошерст В.В. Волновые твердотельные гироскопы с металлическим резонатором // Гироскопия и навигация. 2023. Т. 31. № 1(120). С. 26–44.
7. Распопов В.Я., Волчихин И.А., Волчихин А.И., Ладонкин А.В., Лихошерст В.В., Матвеев В.В. Волновой твердотельный гироскоп с металлическим резонатором: монография / под ред. В.Я. Распопова. Тула: Изд-во ТулГУ, 2018. 189 с.
8. Chikovani V.V., Yatsenko Yu.A., Barabashov A.S., Kovalenko V.A., Tewksbury P. A set of high accuracy low cost metallic resonator CVG // Proc. of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2007). 2007. P. 1345–1350.
9. Acar C., Shkel A. MEMS Vibratory Gyroscopes: Structural Approaches to Improve Robustness. Springer, 2009. 256 p. https://doi.org/10.1007/978-0-387-09536-3
10. Матвеев В.А., Липатников В.И., Алехин А.В. Проектирование волнового твердотельного гироскопа: учеб. пособие для студентов высш. техн. учеб. заведений. М.: Изд-во МГТУ им. Н Э. Баумана, 1998. 166 с.
11. Климов Д.М., Журавлёв В.Ф., Жбанов Ю.К. Кварцевый полусферический резонатор (волновой твердотельный гироскоп). М.: Ким Л.А., 2017. 193 с.
12. Lynch D.D. Coriolis vibratory gyroscope // IEEE Standard Specification Format Guide and Test Procedure for Coriolis Vibratory Gyros. IEEE std.1431 annex B. 2004. P. 56–66.
13. Шарма Н.Г., Сундарараджан Т., Сингх Г.С. Гибридный резонатор твердотельного волнового гироскопа с высокой добротностью: конструкция с использованием термоупругого демпфирования, исследование чувствительности и определение характеристик // Гироскопия и навигация. 2021. Т. 29. № 1(112). С. 70–96. https://doi.org/10.17285/0869-7035.0057
14. Маслов А.А., Маслов Д.А., Ниналалов И.Г., Меркурьев И.В. Волновые твердотельные гироскопы: обзор публикаций // Гироскопия и навигация. 2023. Т. 31. № 1(120). С. 3–25.
15. Lynch D.D. Vibratory gyro analysis by the method of averaging // Proc. of the 2nd St. Petersburg conference on gyroscopic technology and navigation. St. Petersburg, 1995. P. 26–34.
16. Северов Л.А., Золотарев С.К., Овчинникова Н.А., Панферов А.И., Пономарев В.К. Информационные характеристики микромеханических гироскопов на основе кремниевой технологии микроэлектромеханических систем // Известия высших учебных заведений. Приборостроение. 2011. Т. 54. № 8. С. 12–22.
17. Пановко Я.Г. Введение в теорию механических колебаний: учеб. пособие для втузов. М.: Наука, 1971. 239 с.

References

1. Vavilov V.D., Timoshenkov S.P., Timoshenkov A.S. *Microsystem Sensors of Physical Quantities*. Moscow, Technosphera Publ., 2018, 550 p. (in Russian)
2. Peshekhonov V.G. The author of precision gyroscope. *Gyroscopy and Navigation*, 2020, vol. 11, no. 2, pp. 188–192. https://doi.org/10.1134/s2075108720020042
3. Lunin B.S. *Science and technology basis for the development of the hemispherical resonators for solid-state wave gyroscopes*. Abstract of the dissertation for the degree of Doctor of Technical Sciences. Moscow, MSU, 2006, 32 p. (in Russian)
4. Lunin B.S., Matveev V.A., Basarab M.A. *Solid-State Wave Gyroscope*. Moscow, Radiotehnika Publ., 2014, 176 p. (in Russian)
5. Raspopov V.Ya., Egorov S.V. Design and verification of the finite element model of a wave solid-state gyroscope resonator. *Izvestiya Tula State University. Technical sciences*, 2021, no. 10, pp. 227–235. (in Russian)
6. Raspopov V.Y., Likhosherst V.V. HRG with a metal resonator. *Gyroscopy and Navigation*, 2023, vol. 14, no. 1, pp. 14–26. https://doi.org/10.1134/S2075108723010066
7. Raspopov V.Ia., Volchikhin I.A., Volchikhin A.I., Ladonkin A.V., Likhosherst V.V., Matveev V.V. *Solid-State Wave Gyroscope with a Metal Resonator*. Tula, Tula State University, 2018, 189 p. (in Russian)
8. Chikovani V.V., Yatsenko Yu.A., Barabashov A.S., Kovalenko V.A., Tewksbury P. A set of high accuracy low cost metallic resonator CVG. *Proc. of the 20th International Technical Meeting of the Satellite Division of The Institute of Navigation (ION GNSS 2007)*, 2007, pp. 1345–1350.
9. Acar C., Shkel A. *MEMS Vibratory Gyroscopes: Structural Approaches to Improve Robustness*. Springer, 2009, 256 p. https://doi.org/10.1007/978-0-387-09536-3
10. Matveev V.A., Lipatnikov V.I., Alekhin A.V. *Design of the Solid State Wave Gyro*. Moscow, Bauman Moscow State Technical University, 1998, 166 p. (in Russian)
11. Klimov D.M., Zhuravlev V.F., Zhbanov Iu.K. *Hemispherical Quartz Resonator (Solid-State Wave Gyroscope)*. Moscow, Kim L.A. Publ., 2017, 193 p. (in Russian)
12. Lynch D.D. Coriolis vibratory gyroscope. *IEEE Standard Specification Format Guide and Test Procedure for Coriolis Vibratory Gyros. IEEE std.1431 annex B*. 2004, pp. 56–66.
13. Sharma G.N., Sundararajan T., Gautam S.S. Thermoelastic damping based design, sensitivity study and demonstration of a functional hybrid gyroscope resonator for high quality factor. *Gyroscopy and Navigation*, 2021, vol. 12, no. 1, pp. 69–85. https://doi.org/10.1134/S2075108721010107
14. Maslov A.A., Maslov D.A., Ninalalov I.G., Merkuryev I.V. Hemispherical resonator gyros (an overview of publications). *Gyroscopy and Navigation*, 2023, vol. 14, no. 1, pp. 1–13. https://doi.org/10.1134/S2075108723010054
15. Lynch D.D. Vibratory gyro analysis by the method of averaging. *Proc. of the 2nd St. Petersburg conference on gyroscopic technology and navigation*. St. Petersburg, 1995, pp. 26–34.
16. Severov L.A., Zolotarev S.K., Ovchinnikova N.A., Panferov A.I., Ponomarev V.K. Data characteristics of the micromechanical gyroscopes based on silicon technology microelectromechanical systems. *Journal of Instrument Engineering*, 2011, vol. 54, no. 8, pp. 12–22. (in Russian)
17. Panovko Ia.G. *Introduction to Mechanical Vibration Theory*. Moscow, Nauka Publ., 1971, 239 p. (in Russian)
18. Kucherkov S.G. Using the integrating properties of a vibrating micromechanical gyroscope with resonant tuning when constructing

18. Кучерков С.Г. Использование интегрирующих свойств вибрационного микромеханического гироскопа с резонансной настройкой при построении датчика угловой скорости компенсационного типа // Гироскопия и навигация. 2002. № 2. С. 12–18.
19. Chikovani V. Secondary wave control system of the Coriolis vibratory gyroscope resonator // Electronics and Control Systems. 2013. N 3. P. 58–61. https://doi.org/10.18372/1990-5548.35.5805
20. Ландау Л.Д., Лифшиц Е.М. Теоретическая физика: Учеб. пособие. В 10 т. Т. 1. Механика / 4-е изд., испр. М.: Наука. гл. ред. физ.-мат. лит, 1988. 216 с.

a compensation-type angular velocity sensor. *Giroskopiya i Navigatsiya*, 2002, no. 2, pp. 12–18. (in Russian)
19. Chikovani V. Secondary wave control system of the Coriolis vibratory gyroscope resonator. *Electronics and Control Systems*, 2013, no. 3, pp. 58–61. https://doi.org/10.18372/1990-5548.35.5805
20. Landau L., Lifshitz E. *Course of Theoretical Physics. Vol. 1. Mechanics*. Elsevier, 1988, 224 p.

Авторы

Матвеев Валерий Владимирович — доктор технических наук, доцент, ведущий научный сотрудник, заведующий кафедрой, Тульский государственный университет, Тула, 300012, Российская Федерация, sc 56382006100, https://orcid.org/0000-0001-8997-0297, matweew.valery@yandex.ru

Лихошерст Владимир Владимирович — кандидат технических наук, старший научный сотрудник, доцент, Тульский государственный университет, Тула, 300012, Российская Федерация, sc 57190385346, https://orcid.org/0009-0009-4315-5997, lvv_01@inbox.ru

Каликанов Алексей Владимирович — младший научный сотрудник, ассистент кафедры, Тульский государственный университет, Тула, 300012, Российская Федерация, sc 5755680050, https://orcid.org/0000-0002-1590-4278, kalikanov.aleksei@mail.ru

Погорелов Максим Георгиевич — кандидат технических наук, заведующий лабораторией, доцент, Тульский государственный университет, Тула, 300012, Российская Федерация, sc 55983327300, https://orcid.org/0009-0005-9731-3066, mgpogoreloff@yandex.ru

Кирсанов Максим Дмитриевич — младший научный сотрудник, ассистент, Тульский государственный университет, Тула, 300012, Российская Федерация, sc 79539670944, https://orcid.org/0009-0002-2872-9322, kirsanovMD@yandex.ru

Телухин Сергей Владимирович — кандидат технических наук, доцент, Тульский государственный университет, Тула, 300012, Российская Федерация, sc 56770129600, https://orcid.org/0009-0001-2649-8420, teluhinserg@yandex.ru

Authors

Valery V. Matveev — D.Sc., Associate Professor, Leading Researcher, Head of Department, Tula State University, Tula, 300012, Russian Federation, sc 56382006100, https://orcid.org/0000-0001-8997-0297, matweew.valery@yandex.ru

Vladimir V. Likhosherst — PhD, Senior Researcher, Associate Professor, Tula State University, Tula, 300012, Russian Federation, sc 57190385346, https://orcid.org/0009-0009-4315-5997, lvv_01@inbox.ru

Alexey V. Kalikanov — Junior Researcher, Assistant, Tula State University, Tula, 300012, Russian Federation, sc 5755680050, https://orcid.org/0000-0002-1590-4278, kalikanov.aleksei@mail.ru

Maxim G. Pogorelov — PhD, Head of Laboratory, Associate Professor, Tula State University, Tula, 300012, Russian Federation, sc 55983327300, https://orcid.org/0009-0005-9731-3066, mgpogoreloff@yandex.ru

Maxim D. Kirsanov — Junior Researcher, Associate Professor, Tula State University, Tula, 300012, Russian Federation, sc 79539670944, https://orcid.org/0009-0002-2872-9322, kirsanovMD@yandex.ru

Sergey V. Telukhin — PhD, Associate Professor, Tula State University, Tula, 300012, Russian Federation, sc 56770129600, https://orcid.org/0009-0001-2649-8420, teluhinserg@yandex.ru

Статья поступила в редакцию 04.04.2023
Одобрена после рецензирования 09.12.2023
Принята к печати 13.01.2024

Received 04.04.2023
Approved after reviewing 09.12.2023
Accepted 13.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

doi: 10.17586/2226-1494-2024-24-1-144-155
УДК 53.082.79

Сбор и обработка экологической информации в районах нефтегазодобычи и решение других прикладных задач методами активного поиска (обзорная статья)

Игорь Владимирович Свитнев[1], Александр Фотеевич Найданов[2]✉, Алексей Владимирович Вилков[3], Дмитрий Александрович Соколов[4], Михаил Юрьевич Лебедев[5], Елена Александровна Харитонова[6], Людмила Анатольевна Лукьянова[7]

[1,2,3,4,5] Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация
[6,7] Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация

[1] isvitnev@mail.ru, https://orcid.org/0000-0002-8862-1797
[2] naydanov@bk.ru✉, https://orcid.org/0009-0003-4242-1700
[3] Leha_kostroma@mail.ru, https://orcid.org/0000-0002-9300-3547
[4] falcon.85@bk.ru, https://orcid.org/0009-0001-0111-403X
[5] Lebedev.m1ha1l@yandex.ru, https://orcid.org/0009-0003-8876-7846
[6] Xaritonova_ea@mail.ru, https://orcid.org/0000-0002-4851-2756
[7] lalukianova@mail.ru, https://orcid.org/0000-0001-5458-3019

Аннотация

Введение. Исследованы способы мониторинга экологической обстановки, проблемы решения смежных прикладных эколого-ресурсных задач в труднодоступных районах нефтегазодобычи, а также в других отраслях народного хозяйства с использованием беспилотных летательных средств. Рассмотрены методы исследования типов и толщин слоев подстилающей поверхности путем их зондирования электромагнитными импульсами радиочастотного диапазона и гамма-излучения. **Метод.** На основе существующих теоретических зависимостей взаимодействия электромагнитного излучения с поверхностью Земли представлены диаграммы прохождения электромагнитных волн дециметрового и сантиметрового диапазонов через различные структуры ландшафта (снег-лед-вода-мерзлая почва). **Основные результаты.** Показано, что использование гамма-излучения позволяет решить проблему определения эффективной высоты летательного аппарата при экологическом мониторинге за счет большой энергии фотонного излучения и альбедо от различных поверхностей, в том числе от снежного покрова. Приведена методика расчета содержания загрязнителя на подстилающей поверхности с заданной вероятностью его достоверного обнаружения. Отмечено, что на достоверность показаний измерительных средств значительное влияние оказывает их геометрическое расположение на транспортной платформе. Показано, что предложенное решение целесообразно реализовать с использованием двух беспилотных летательных аппаратов или малоразмерного беспилотного дирижабля. На основании сделанного обзора предложен состав технических средств комплекса распознавания типов и толщин слоев загрязнений подстилающей поверхности. Приведена возможная методика оценивания экологической обстановки. **Обсуждение.** Результаты работы могут найти применение при проведении экологической разведки, при транспортировке нефтяных и газовых ресурсов в условиях затруднения доступа к ним. Расширены возможности решения подобных военно-прикладных и инженерно-строительных задач. Впервые предложено совместное использование электромагнитных волн радиочастотного диапазона и гамма-излучения. Электромагнитное излучение радиочастотного диапазона позволяет исследовать структуру ландшафта. Гамма-излучение по обратному рассеянному ионизирующему излучению дает возможность определить тип загрязнителя и обеспечить высокую точность измерения расстояния от модуля до верхнего слоя подстилающей поверхности.

Ключевые слова

гамма-излучение, доверительный интервал, загрязнитель, окружающая среда, определение высоты, радиоимпульс, толщина слоя, экологическая обстановка, электромагнитные волны

© Свитнев И.В., Найданов А.Ф., Вилков А.В., Соколов Д.А., Лебедев М.Ю., Харитонова Е.А., Лукьянова Л.А., 2024

Ссылка для цитирования: Свитнев И.В., Найданов А.Ф., Вилков А.В., Соколов Д.А., Лебедев М.Ю., Харитонова Е.А., Лукьянова Л.А. Сбор и обработка экологической информации в районах нефтегазодобычи и решение других прикладных задач методами активного поиска (обзорная статья) // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 144–155. doi: 10.17586/2226-1494-2024-24-1-144-155

Collection and processing of environmental information in oil and gas production areas and solving other applied problems using active search methods (Review article)

Igor V. Svitnev[1], Alexander F. Naydanov[2]✉, Alexey V. Vilkov[3], Dmitry A. Sokolov[4], Mikhail Yu. Lebedev[5], Elena A. Kharitonova[6], Lyudmila A. Lukyanova[7]

[1,2,3,4,5] Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation
[6,7] St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation

[1] isvitnev@mail.ru, https://orcid.org/0000-0002-8862-1797
[2] naydanov@bk.ru✉, https://orcid.org/0009-0003-4242-1700
[3] Leha_kostroma@mail.ru, https://orcid.org/0000-0002-9300-3547
[4] falcon.85@bk.ru, https://orcid.org/0009-0001-0111-403X
[5] Lebedev.m1ha1l@yandex.ru, https://orcid.org/0009-0003-8876-7846
[6] Xaritonova_ea@mail.ru, https://orcid.org/0000-0002-4851-2756
[7] lalukianova@mail.ru, https://orcid.org/0000-0001-5458-3019

Abstract

The methods of monitoring the environmental situation as well as the problems of solving related applied environmental and resource problems in hard-to-reach areas of oil and gas production and also in other sectors of the national economy using unmanned aerial vehicles, are investigated. The methods of studying the types and thicknesses of the layers of the underlying surface by probing them with electromagnetic pulses of the radiofrequency range and gamma radiation are considered. Based on the existing theoretical dependencies of the interaction of electromagnetic radiation with the Earth's surface, diagrams of the passage of electromagnetic waves in the decimeter and centimeter ranges through various landscape structures (snow-ice-water-frozen soil) are presented. It is shown that the use of gamma radiation makes it possible to solve the problem of determining the effective altitude of an aircraft during environmental monitoring due to the high energy of photon radiation and albedo from various surfaces including snow cover. A method for calculating the pollutant content on the underlying surface with a given probability of its reliable detection is presented. It is noted that the reliability of the readings of measuring instruments is significantly influenced by their geometric location on the transport platform. It is shown that the proposed solution is advisable to implement using two unmanned aerial vehicles or as mall-sized unmanned airship. Based on the review, the composition of the technical means of the complex for recognizing the types and thicknesses of layers of contamination of the underlying surface is proposed. A possible methodology for assessing the environmental situation is presented. The results of the work can be used in conducting environmental exploration of infrastructure used for transporting oil and gas resources in conditions of difficult access to it as well as for solving similar military-applied and engineering-construction tasks. At the same time, for the first time, the joint use of the radio frequency range of electromagnetic waves and gamma radiation was proposed. The radio frequency range makes it possible to study the structure of the landscape, and gamma radiation from backscattered ionizing radiation is a type of pollutant, as well as to ensure high accuracy in measuring the distance from the module to the upper layer of the underlying surface.

Keywords

gamma radiation, confidence interval, pollutant, environment, height determination, radio pulse, layer thickness, environmental situation, electromagnetic waves

For citation: Svitnev I.V., Naydanov A.F., Vilkov A.V., Sokolov D.A., Lebedev M.Yu., Kharitonova E.A., Lukyanova L.A. Collection and processing of environmental information in oil and gas production areas and solving other applied problems using active search methods (Review article). *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 144–155 (in Russian). doi: 10.17586/2226-1494-2024-24-1-144-155

Введение

В настоящее время основными задачами государственной политики в сфере обеспечения экологической безопасности являются: совершенствование системы государственного экологического мониторинга; научное и информационно-аналитическое обеспечение охраны окружающей среды, в том числе в отношении земельных ресурсов и территорий районов нефтегазодобычи. В последние годы проекты по добыче нефти и газа становятся более требовательными с точки зрения технологий и рисков, и нефтяники все большее внимание уделяют трудноизвлекаемым запасам углеводородов. В результате появляется необходимость разработки новых месторождений и строительства транспортной инфраструктуры в труднодоступных районах. Развитие сектора нефтегазодобычи приводит к многократному увеличению площадей территорий, подлежащих экологическому мониторингу. Приоритетными направлениями при этом являются повышение технического потенциала и оснащенности сил, участвующих в мероприятиях по предотвращению и ликвидации негативных экологических последствий, и обеспечение развития программ обработки данных, позволяющих

своевременно получать достоверную информацию о состоянии окружающей среды[1].

Одной из основных форм контроля выполнения природоохранных мероприятий являются плановые и внеплановые инспекционные проверки, в рамках которых экологическое состояние территорий в нефтепромышленных районах оценивают преимущественно путем визуального осмотра с использованием, при необходимости, средств инструментального контроля.

Данный подход обусловливает наличие ряда объективных обстоятельств, которые ограничивают возможности органов управления, отвечающих за контрольно-надзорную деятельность. С одной стороны — наличие большого количества объектов и территорий, в том числе труднодоступных и значительных по площади, на фоне возрастания угроз экологической безопасности при наличии накопленного вреда окружающей среде. С другой стороны — необходимость контроля этих территорий при условии, что сбор данных, подлежащих анализу и учету, характеризующих их экологическое состояние, является достаточно трудоемким [1].

Нивелирование данного противоречия при решении задач экологического мониторинга территорий возможно путем привлечения средств аэрокосмической съемки, работа которых, как правило, основана на визуальных методах. Они позволяют охватить значительные и труднодоступные территории, но автоматизация обработки получаемых изображений относится к весьма сложной задаче. Дешифрирование полученной информации осуществляется вручную специалистом-дешифровщиком, либо экспертами. Кроме того, применение этих средств довольно ограничено условиями видимости, метеорологическими условиями и требованиями к безопасности полетов, что напрямую влияет на качество получаемых материалов. Также на результат мониторинга значительное влияние оказывает профессионализм дешифровщика и экспертов.

Вместе с тем указанные средства и реализуемые ими визуальные методы не предназначены для разведки структуры слоистой подстилающей поверхности, тем более в сложных метеорологических условиях. В настоящее время для решения подобных задач применяются контактные методы многократного бурения для определения толщин слоев подстилающей поверхности. Однако в эпоху развития робототехники такой метод не соответствует концепциям успешных добывающих компаний.

Одним из наиболее очевидных вариантов решения обозначенной проблемы является использование эффекта отражения и преломления электромагнитных волн при прохождении различных сред.

Техническая реализация данного принципа возможна, например, с беспилотного летательного аппарата. Однако метод по некоторым причинам не нашел применения. Основной из них является недостижение необходимой точности из-за ряда технических ограничений, накладываемых на носитель аппаратуры, обусловленных условиями его применения, например учет нестабильности излучения при измерении яркости по спектральным каналам [2].

Таким образом, цель настоящей работы — повышение точности определения типов и толщин слоев подстилающей поверхности, обнаружение наличия на ней загрязнителей, и использование этих данных для оценивания экологического состояния земельных ресурсов в районах нефтегазодобычи. Также с применением такого подхода целесообразно рассматривать возможность решения смежных эколого-ресурсных задач, требующих получение информации о структуре и составе подстилающей поверхности.

Поставленная цель достигается решением задач по разработке методического аппарата, позволяющего определить содержание загрязнителя на подстилающей поверхности, оценить экологическую обстановку, а также обосновать технический облик комплекса распознавания типов и толщин слоев подстилающей поверхности.

Научная новизна работы заключается в том, что предлагаемый способ принципиально отличается от известных активных способов зондирования использованием в дополнение к радиочастотному диапазону, позволяющему определять толщины и типы слоев подстилающей поверхности (снег-лед-вода-мерзлая почва), гамма-излучение, благодаря которому по обратно рассеянному фотонному излучению возможно определять состав загрязнителя на подстилающей поверхности [3], а также обеспечивать устойчивое поддержание высоты полета в условиях рыхлого (снежного) покрова.

Обзор существующих методов зондирования земной поверхности

В настоящее время уделяется особое внимание на проблемы экологии в нефтегазодобывающих районах, появляются новые технологические решения для эффективного сбора и анализа экологической информации. Рассмотрим подробнее несколько перспективных направлений таких исследований.

Дистанционное зондирование и дроны в экологическом мониторинге нефтегазовых регионов. Исследования, проведенные с помощью дистанционного зондирования и беспилотных летательных аппаратов (дронов), подчеркивают важность их применения [4]. Данные, собранные этими средствами, обеспечивают детальный мониторинг труднодоступных участков, а также позволяют в реальном времени отслеживать изменения в экосистемах[2].

Дистанционное зондирование основано на использовании спутников и аэросъемочных платформ для получения данных о поверхности Земли. Рабочий диапа-

[1] Указ Президента Российской Федерации от 19.04.2017 г. № 176 «О Стратегии экологической безопасности Российской Федерации на период до 2025 года».

[2] Киселевская К.Е. Применения метода дистанционного зондирования Земли для экологического мониторинга // ГИАБ. 2009. № 1 [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/primeneniya-metoda-distantsionnogo-zondirovaniya-zemli-dlya-ekologicheskogo-monitoringa (дата обращения: 13.11.2023).

зон длин волн, принимаемых съемочной аппаратурой, составляет от долей микрометра (видимое оптическое излучение) до метров (радиоволны) [5].

Дроны предоставляют возможность собирать данные в реальном времени, обеспечивая более высокое пространственное и временное разрешения. Это позволяет эффективно обнаруживать и отслеживать утечки нефти, изменения в ландшафте и водных ресурсах [6, 7].

Сенсорные технологии в экологическом мониторинге нефтегазовых регионов. Сенсоры регулярно отслеживают выбросы и качество воды, предоставляя важные данные для контроля и предотвращения экологических аварий. Примером может служить использование сенсоров для измерения уровней метана, сероводорода и других газов вблизи нефтяных скважин, мониторинг качества воды в реках и озерах, а также реагирование на утечки нефти или химических веществ [8]. При этом выполняется автоматическая передача данных в центры управления, обеспечивая оперативный отклик на потенциальные экологические угрозы.

Использование сенсорных технологий в экологическом мониторинге нефтегазовых регионов предоставляет возможность оперативного реагирования на потенциальные угрозы и минимизацию вредного воздействия на природную среду[1].

Машинное обучение и анализ данных в экологическом мониторинге нефтегазового сектора. Нейронные сети применяют для выявления аномалий в данных и для автоматического реагирования на потенциальные угрозы. Главная причина внедрения нейронных сетей – возможность оперативного выявления аномалий в экологических данных. Методы глубокого обучения позволяют создавать модели, способные автоматически выявлять отклонения от нормы в параметрах окружающей среды, что создает условия для быстрого реагирования на потенциальные экологические угрозы[2].

Выводы по обзору методов зондирования земных поверхностей. В результате можно сделать следующие выводы. Современные технологии в экологическом мониторинге нефтегазовых регионов представляют собой мощный инструментарий для контроля и реагирования на экологические угрозы. Дистанционное зондирование и дроны обеспечивают широкий охват и точный мониторинг. Сенсорные технологии создают условия непрерывного контроля параметров окружающей среды, а машинное обучение и анализ данных позволяют обрабатывать огромные объемы информации для выявления тенденций и предупреждений об угрозах.

[1] Казьмина И.Г., Рязанцева Л.Т., Федянин В.И. Химические сенсоры в системе экологического мониторинга // Вестник ВГТУ. 2011. № 2 [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/himicheskie-sensory-v-sisteme-ekologicheskogo-monitoringa (дата обращения: 13.11.2023).

[2] Федутинов К.А. Машинное обучение в задачах поддержки принятия решений при управлении охраной природы // ИВД. 2021. № 9 (81) [Электронный ресурс]. URL: https://cyberleninka.ru/article/n/mashinnoe-obuchenie-v-zadachah-podderzhki-prinyatiya-resheniy-pri-upravlenii-ohranoy-prirody (дата обращения: 13.11.2023).

Решение задачи определения структуры подстилающей поверхности по отраженным электромагнитным сигналам

Известно, что различные среды, обладающие проводимостями, магнитными и диэлектрическими проницаемостями, являются для электромагнитных волн радиочастотного диапазона диэлектриками или проводниками, или полупроводниками. Отсюда следует разное соотношение между коэффициентами преломления, отражения и поглощения [9–11]. Производя измерение уровней отраженных сигналов E_1–E_5. относительно исходного E_0, можно судить о типах сред, находящихся на границе их раздела (рис. 1).

Пользуясь фундаментальными законами физики, установим необходимые соотношения для расчета толщины слоя [9].

Зависимость параметров (угол преломления и скорость) электромагнитной волны от среды, в которой она распространяется, получим в виде:

$$\sin\theta_i = \sqrt{\frac{\mu_{r_i}\varepsilon_{r_i}}{\mu_{r_{i-1}}\varepsilon_{r_{i-1}}}}\sin\theta_{i-1}, \quad (1)$$

$$v_i = \frac{c}{\sqrt{\mu_{r_i}\varepsilon_{r_i}}}, \quad (2)$$

где θ_i — угол преломления электромагнитной волны; μ_{r_i} и ε_{r_i} — магнитная и диэлектрическая проницаемости i-го слоя (после раздела сред); $\mu_{r_{i-1}}$ и $\varepsilon_{r_{i-1}}$ — магнитная и диэлектрическая проницаемости i–1 слоя (до раздела сред); θ_{i-1} и v_i — угол падения и скорость электромагнитной волны; c — скорость света в вакууме [12].

Формула (1) устанавливает зависимость угла преломления электромагнитной волны от среды, а (2) — скорость распространения электромагнитной волны в среде. Проецируя v_i на нормаль к границе раздела сред, получим:

$$v_{ni} = \frac{c}{\sqrt{\mu_{r_i}\varepsilon_{r_i}}}\sqrt{1 - \frac{\mu_{r_i}\varepsilon_{r_i}}{\mu_{r_{i-1}}\varepsilon_{r_{i-1}}}\sin^2\theta_{i-1}}. \quad (3)$$

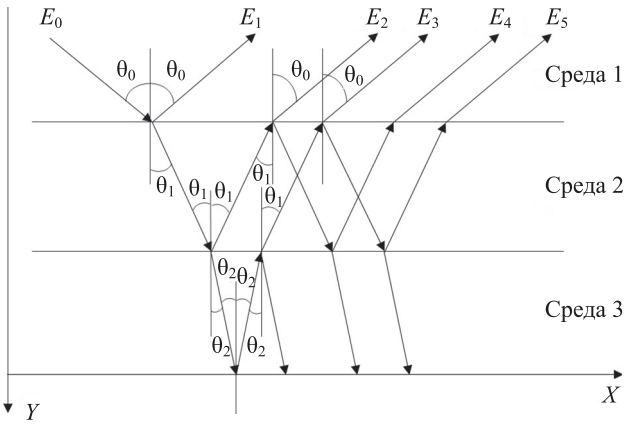


Рис. 1. Геометрическая модель распространения радиоволн в различных средах

Fig. 1. Geometric model of radio wave propagation in different environments

Определение толщин слоев выполним по временному интервалу между принимаемыми отраженными сигналами от границ раздела сред (рис. 1).

Определение глубины прохождения сигнала в средах H_n опишем системой уравнений:

$$\begin{cases} H_n = v_1\frac{\tau_1}{2} + v_2\frac{\tau_2}{2} + v_3\frac{\tau_3}{2} + \ldots + v_n\frac{\tau_n}{2} = \sum\limits_{i=1}^{n} v_i\frac{\tau_i}{2}, \\ H_n = h_1 + h_2 + h_3 + \ldots + h_n = \sum\limits_{i=1}^{n} h_i, \end{cases} \quad (4)$$

где τ_i — время прохождения отраженных сигналов от i-го слоя; h_i — толщина i-го слоя; n — количество слоев (сред); i — номер слоя (среды).

С учетом системы уравнений (4) получим выражение для расчета слоя i-ой среды:

$$h_{in} = \frac{c(\tau_i - \tau_{i-1})}{\sqrt{\mu_{r_i}\varepsilon_{r_i}}}\sqrt{1 - \frac{\mu_{r_i}\varepsilon_{r_i}}{\mu_{r_{i-1}}\varepsilon_{r_{i-1}}}\sin^2\theta_{i-1}}, \quad (5)$$

где τ_{i-1} — время регистрации от i–1-го слоя [12].

Из формулы (5) видно, что для расчета толщины слоя необходимо знать тип самого слоя. Для этого аналитически определим отношение принимаемого сигнала к излученному через коэффициенты отражения $P_{\text{отр}}$ и преломления $P_{\text{пр}}$:

$$P_{\text{отр}} = \frac{Z_i - Z_{i-1}}{Z_i + Z_{i-1}}, \quad (6)$$

$$P_{\text{пр}} = \frac{2Z_i}{Z_i + Z_{i-1}}, \quad (7)$$

где $Z_i = \sqrt{\frac{\sigma_i}{\omega\varepsilon_{r_i}}\cos\theta_i}$ — волновое сопротивление i-ой среды; σ_i — проводимость i-ой среды; ω — круговая частота падающей электромагнитной волны [12].

Для примера на рис. 2 показана диаграмма, иллюстрирующая потери энергии электромагнитной волны при отражении от различных типовых сред. Данная диаграмма позволяет по отношению амплитуд принимаемого E и излученного E_0 сигналов определить тип среды, а затем, сравнив с параметрами среды, определить толщину слоя, используя математическую модель (1)–(7).

Как видно из рис. 2 [12], отражение электромагнитной волны от раздела воздух-снег, в отличие от других границ, составляет менее 5 %, что оказывает влияние на определение высоты полета носителя аппаратуры относительно снежного покрова.

Использование гамма-излучения позволяет решить проблему определения высоты полета за счет высокой энергетики фотонного излучения и альбедо от различных поверхностей. Процессы рассеивания и поглощения гамма-излучения больше происходят в плотных слоях, поэтому на разделах воздуха и подстилающего основания, в том числе снежного покрова, будет фиксироваться значительное изменение альбедо.

Основной эффект взаимодействия гамма-квантов на границе раздела сред, позволяющий получить отраженное излучение — эффект Комптона [13]. При наличии данного эффекта энергия отраженных квантов зависит от угла падения.

Результаты расчетов показали, что оптимальным углом падения [14], при котором энергия принимаемых гамма-квантов наибольшая, является 30°–45° относительно нормали к поверхности. В результате получим, что дальность действия такой системы очень зависит от физических размеров носителя аппаратуры и геометрического расположения ее компонентов — передатчика и приемника.

Точное определение высоты полета — важная техническая составляющая, так как определение загрязнителя на поверхности требует высокой чувствительности приемной аппаратуры.

Для минимизации мощности источника гамма-излучения предлагается высоту полета обеспечивать на достаточно минимальном уровне (от двух до нескольких десятков метров). Однако данная задача по определению оптимальной высоты полета требует проведения дополнительных исследований и в материал настоящей работы не входит. При определении высоты полета гамма-источник необходимо использовать в импульсном режиме.

Методика расчета содержания загрязнителя на подстилающей поверхности с заданной вероятностью его достоверного обнаружения

Рассмотрим методику расчета концентрации и толщин слоев загрязнителя на подстилающей поверх-

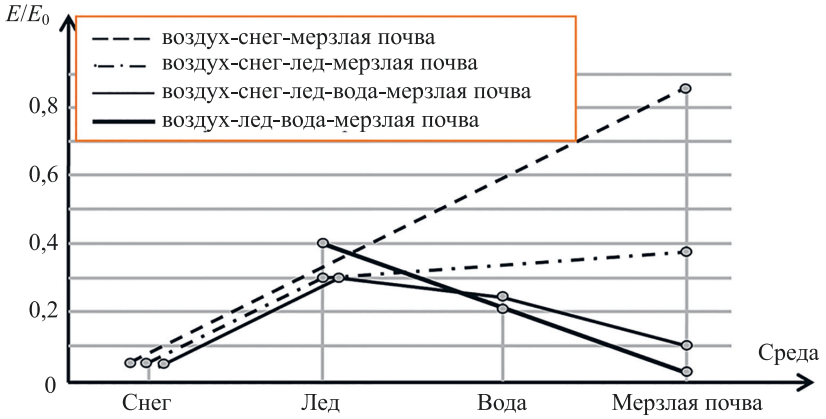


Рис. 2. Диаграмма, иллюстрирующая потери энергии электромагнитной волны при отражении от различных сред

Fig. 2. A diagram illustrating the electromagnetic wave energy losses during reflection from different environments

ности, в которой установление наличия (отсутствия) загрязнителя на подстилающей поверхности зависит от регистрации параметров обратно рассеянного ионизирующего излучения (ОРИИ) и погрешности средств измерения. Отметим, что на достоверность показаний измерительных средств значительное влияние оказывает геометрическое расположение их компонентов на носителе аппаратуры. Правильный расчет геометрического расположения компонентов аппаратуры позволяет далее говорить о подходах к точному нахождению загрязнителей. Так установлено, что исходя из высоты полета над подстилающей поверхностью не менее двух метров, база между передатчиком и приемником должна иметь размер от двух до четырех метров. Для современных беспилотных летательных аппаратов, выполненных на основе планера или квадрокоптера, такое условие трудновыполнимо. Однако это решение можно реализовать с использованием двух беспилотных летательных аппаратов или на малоразмерном беспилотном дирижабле [15].

Под термином «загрязнитель» понимается любой агент, имеющий природное или техногенное происхождение (физический агент или химическое вещество), попадающий в окружающую среду или возникающий в ней в количествах, которые выходят за рамки предельных естественных колебаний фона, и негативно влияющих на качество окружающей среды и здоровье человека. Исходя из определения, любое отклонение от нормированных показателей (границы которых известны) в окружающей среде на подстилающей поверхности быть не должно.

На рис. 3 представлена схема расчета содержания загрязнителя на подстилающей поверхности с заданной вероятностью его достоверного обнаружения [16].

На этапе 1 вводятся исходные данные, которыми являются интенсивности ОРИИ от чистой подстилающей поверхности I^{cle} и загрязнителя I^{pol}, а также относительная погрешность средства измерения, обусловленная в первую очередь техническими характеристиками применяемого блока детектирования ионизирующего излучения.

На этапе 2, зная исходные данные, рассчитываются абсолютные погрешности, которые формируют доверительные интервалы значений ОРИИ от чистой поверхности и загрязнителя, т. е. интервалы значений, которые может принимать ОРИИ I_0^{sys}, по закону нормального распределения (рис. 4).

Так для чистой подстилающей поверхности абсолютная погрешность измерения ОРИИ будет составлять ΔI^{cle}, а для загрязнителя — ΔI^{pol}.

Таким образом, доверительные интервалы измерений ОРИИ для чистой подстилающей поверхности составят $[I^{cle} - \Delta I^{cle}; I^{cle} + \Delta I^{cle}]$, для загрязнителя $[I^{pol} - \Delta I^{pol}; I^{pol} + \Delta I^{pol}]$. При проведении измерений, с учетом погрешности измерения детектора, интенсивность ОРИИ может принять одно из значений в указанных интервалах.

Доверительные интервалы могут пересекаться (рис. 4, *a*). Возможен случай, когда $I^{cle} > I^{pol}$, тогда пересечение расчетных доверительных интервалов составит от $[I^{cle} - \Delta I^{cle}]$ до $[I^{pol} + \Delta I^{pol}]$. В случае, когда $I^{cle} < I^{pol}$, пересечение доверительных интервалов — от $[I^{pol} - \Delta I^{pol}]$ до $[I^{cle} + \Delta I^{cle}]$.

Полученное детектором значение интенсивности в этих интервалах может говорить как о наличии загрязнителя на подстилающей поверхности, если $I_0^{sys} \in [I^{pol} - \Delta I^{pol}; I^{pol} + \Delta I^{pol}]$, так и о его отсутствии при соблюдении условия, что $I_0^{sys} \in [I^{cle} - \Delta I^{cle}; I^{cle} + \Delta I^{cle}]$.

Также возможен случай, когда расчетные значения доверительных интервалов ОРИИ от чистой подстилающей поверхности и загрязнителя не пересекаются (рис. 4, *b*). В этом случае значение интенсивности ОРИИ так же принимает одно из значений $I_0^{sys} \in [I^{pol} - \Delta I^{pol}; I^{pol} + \Delta I^{pol}]$ или $I_0^{sys} \in [I^{cle} - \Delta I^{cle}; I^{cle} + \Delta I^{cle}]$. Отсутствие пересечения доверительных интервалов говорит о том, что химические составы загрязнителя и чистой подстилающей поверхности отличаются настолько, что расчетные значения интенсивностей ОРИИ от них не пересекаются (с учетом погрешности средства измерения). Соответственно, при регистрации ОРИИ можно достаточно точно определить по полученному значению интенсивности, к чему она относится: к загрязнителю или чистой поверхности.

Таким образом, достоверность обнаружения наличия (отсутствия) загрязнения подстилающей поверхности (ЗПП) будет зависеть от того, с какой вероятностью значение интенсивности ОРИИ I_0^{sys} примет одно из следующих значений:

— $I_0^{sys} \in [I^{pol} - \Delta I^{pol}; I^{cle} - \Delta I^{cle}] \cup [I^{pol} + \Delta I^{pol}; I^{cle} + \Delta I^{cle}]$ при пересечении доверительных интервалов расчетных интенсивностей ОРИИ при $I^{cle} > I^{pol}$ (рис. 4, *a*);

— $I_0^{sys} \in [I^{cle} - \Delta I^{cle}; I^{pol} - \Delta I^{pol}] \cup [I^{cle} + \Delta I^{cle}; I^{pol} + \Delta I^{pol}]$ при пересечении доверительных интервалов расчетных интенсивностей ОРИИ при $I^{cle} < I^{pol}$ (рис. 4, *a*);

— $I_0^{sys} \in [I^{pol} - \Delta I^{pol}; I^{pol} + \Delta I^{pol}] \cup [I^{cle} - \Delta I^{cle}; I^{cle} + \Delta I^{cle}]$ при отсутствии пересечения доверительных интервалов расчетных интенсивностей ОРИИ при $I^{cle} < I^{pol}$ или $I^{cle} < I^{pol}$ (рис. 4, *b*).

Для расчета вероятности достоверного обнаружения наличия (отсутствия) загрязнителя принимаем, что I_0^{sys} по результатам проведенного измерения равновероятно примет одно из значений в диапазоне $[I^{pol} - \Delta I^{pol}; I^{cle} + \Delta I^{cle}]$, если $I^{cle} > I^{pol}$ или $[I^{cle} - \Delta I^{cle}; I^{pol} + \Delta I^{pol}]$, если $I^{cle} < I^{pol}$ (рис. 4, *a*). Либо для случая, когда доверительные интервалы не пересекаются, I_0^{sys} примет значение в интервале $[I^{pol} - \Delta I^{pol}; I^{pol} + \Delta I^{pol}] \cup [I^{cle} - \Delta I^{cle}; I^{cle} + \Delta I^{cle}]$ (рис. 4, *b*).

На этапе 3 рассчитаем области пересечения доверительных интервалов для всех случаев. В случае получения по результатам измерения значения интенсивности ОРИИ I_0^{sys} в области, которая относится к доверительным интервалам, как чистой подстилающей поверхности, так и загрязнителя, невозможно определить к какому именно интервалу это значение относится.

Для случая пересекающихся доверительных интервалов при $I^{cle} > I^{pol}$ (рис. 4, *a*) область пересечения рассчитаем следующим образом:

$$\Delta I^{err} = I^{pol} + \Delta I^{pol} - (I^{cle} - \Delta I^{cle}) = \\ = I^{pol} + \Delta I^{pol} - I^{cle} + \Delta I^{cle}, \quad (8)$$

где ΔI^{err} — область пересечения доверительных интервалов, при попадании значения интенсивности ОРИИ

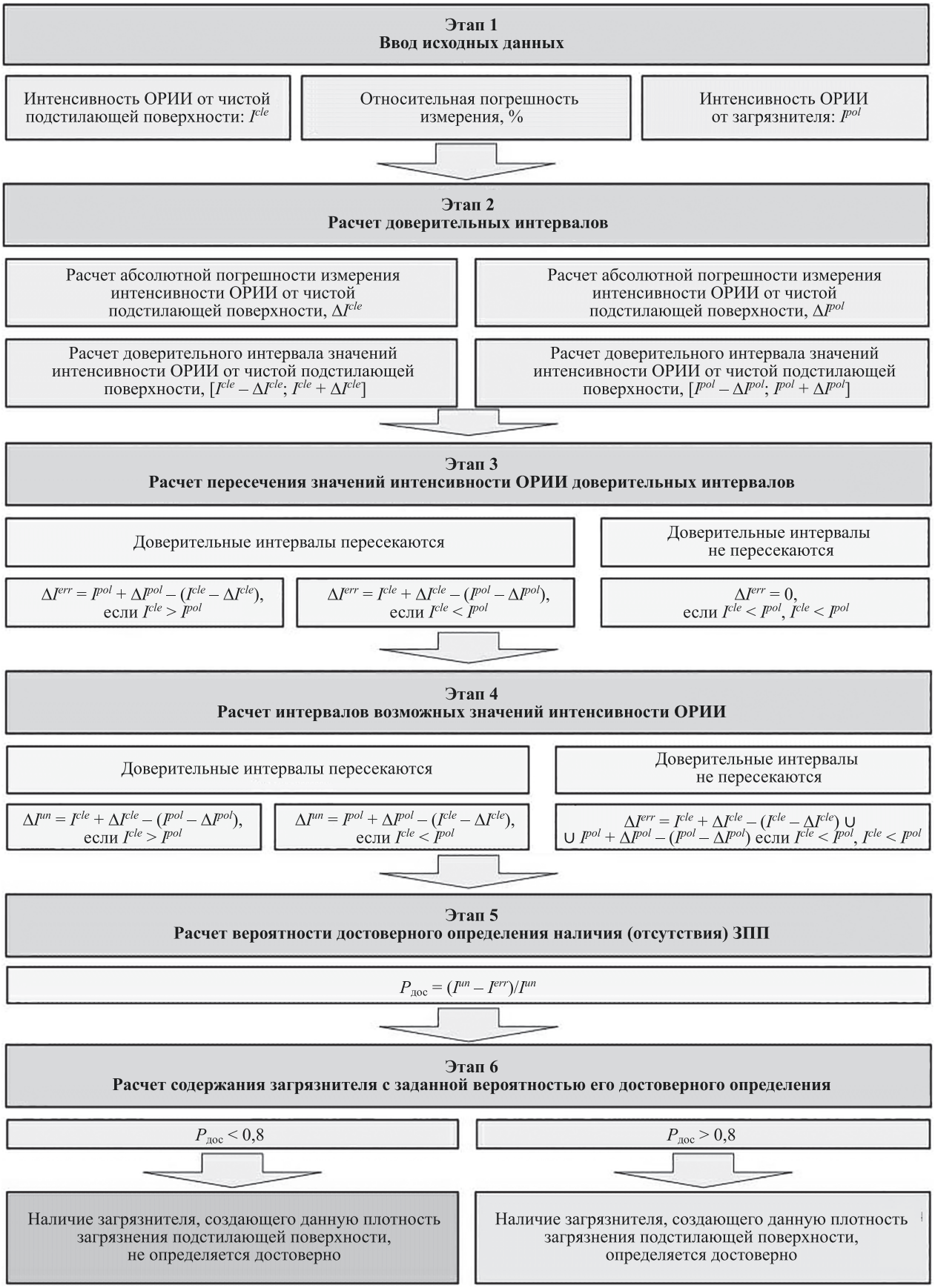


Рис. 3. Структурная схема расчета содержания загрязнителя на подстилающей поверхности с заданной вероятностью его достоверного обнаружения

Fig. 3. Block diagram for calculation the content of a pollutant on the underlying surface with a given probability of its reliable detection

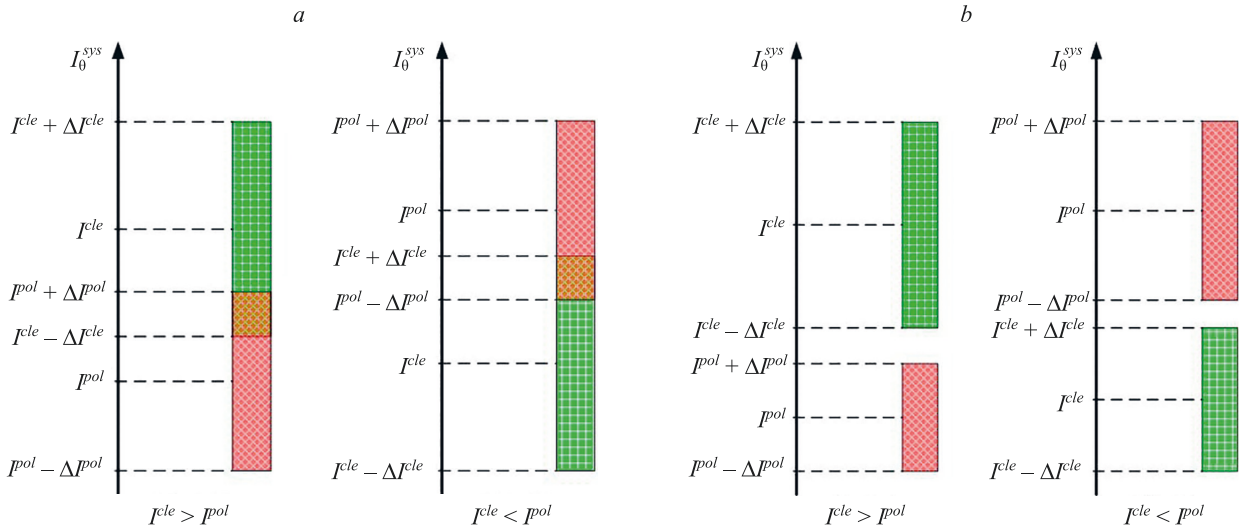


Рис. 4. Схемы доверительных интервалов интенсивностей обратно рассеянного ионизирующего излучения от чистой подстилающей поверхности (зеленый фон) и загрязнителя (красный фон), которые пересекаются (*a*) и не пересекаются (*b*)

Fig. 4. Schemes of confidence intervals for the intensities of backscattered ionizing radiation from a clean underlying surface and a pollutant that intersect (*a*) and do not intersect (*b*)

I_0^{sys} в которую, наличие (отсутствие) ЗПП достоверно не определяется.

Для случая пересекающихся доверительных интервалов при $I^{cle} < I^{pol}$ (рис. 4, *a*) расчет области пересечения имеет вид:

$$\Delta I^{err} = I^{cle} + \Delta I^{cle} - (I^{pol} - \Delta I^{pol}) = \\ = I^{cle} + \Delta I^{cle} - I^{pol} + \Delta I^{pol}. \quad (9)$$

Очевидно, что при отсутствии пересечения доверительных интервалов (рис. 4, *b*) интервал их пересечения $\Delta I^{err} = 0$.

На этапе 4, с учетом условия, что I_0^{sys} принимает любое значение из области доверительных интервалов, получим области возможных значений интенсивности.

Для случая пересекающихся доверительных интервалов и при $I^{cle} > I^{pol}$ (рис. 4, *a*) область возможных значений рассчитаем по формуле:

$$\Delta I^{un} = I^{cle} + \Delta I^{cle} - (I^{pol} - \Delta I^{pol}) = \\ = I^{cle} + \Delta I^{cle} - I^{pol} + \Delta I^{pol}, \quad (10)$$

где ΔI^{un} — область возможных значений I_0^{sys}.

Для случая пересекающихся доверительных интервалов и при $I^{cle} < I^{pol}$ (рис. 4, *a*) выражение (11) примет вид:

$$\Delta I^{un} = I^{pol} + \Delta I^{pol} - (I^{cle} - \Delta I^{cle}) = \\ = I^{pol} + \Delta I^{pol} - I^{cle} + \Delta I^{cle}. \quad (11)$$

Для случая непересекающихся доверительных интервалов (рис. 4, *b*) область возможных значений интенсивности I_0^{sys} составит:

$$\Delta I^{un} = [I^{cle} + \Delta I^{cle} - (I^{cle} - \Delta I^{cle}) + \\ + I^{pol} + \Delta I^{pol} - (I^{pol} - \Delta I^{pol}). \quad (12)$$

Исходя из условия, что I_0^{sys} по результатам проведенного измерения равновероятно примет одно из значений в указанных интервалах, предположим, что вероятность того, что значение I_0^{sys} попадет в тот или иной интервал пропорциональна размеру этих интервалов. Тогда вероятность достоверного обнаружения наличия (отсутствия) загрязнителя будет пропорциональна вероятности того, что значение I_0^{sys} примет одно из значений в интервалах $[I^{pol} - \Delta I^{pol}; I^{cle} - \Delta I^{cle}] \cup [I^{pol} + \Delta I^{pol}; I^{cle} + \Delta I^{cle}]$ или $[I^{cle} - \Delta I^{cle}; I^{pol} - \Delta I^{pol}] \cup [I^{cle} + \Delta I^{cle}; I^{pol} + \Delta I^{pol}]$ при пересечении доверительных интервалов (рис. 4, *a*), или в интервалах $[I^{pol} - \Delta I^{pol}; I^{pol} + \Delta I^{pol}] \cup [I^{cle} - \Delta I^{cle}; I^{cle} + \Delta I^{cle}]$ при отсутствии их пересечения (рис. 4, *b*).

Зная области возможных значений ΔI^{un}, которые может принимать I_0^{sys} области пересечения доверительных интервалов ΔI^{err}, при попадании I_0^{sys} в которые наличие (отсутствие) загрязнения не определяется, а также исходя из пропорциональности вероятности попадания значения I_0^{sys} в эти интервалы их размеру, на этапе 5 перейдем к вероятности достоверного определения наличия (отсутствия) ЗПП:

$$P_{\text{дос}} = \frac{\Delta I^{un} - \Delta I^{err}}{\Delta I^{un}}. \quad (13)$$

На заключительном этапе 6 рассчитаем содержание загрязнителя. При этом выполним оценку соответствия полученных значений вероятности обнаружения ЗПП критериальным методом [14].

Если $P_{\text{дос}} < 0{,}8$, то наличие загрязнителя, создающего данную плотность ЗПП, не определяется достоверно.

Если $P_{\text{дос}} \geq 0{,}8$, то наличие загрязнителя, создающего данную плотность ЗПП, определяется достоверно.

Значение 0,8, по достижении которого загрязнение определяется достоверно, выбрано как стандартное значение шкалы Харрингтона. Данное значение было применено в настоящей работе в связи с отсутствием требований по достоверности к предлагаемой авторами (8)–(13).

Представленный методический аппарат, позволяет рассчитывать содержание загрязнителя на подстилающей поверхности с заданной вероятностью его достоверного обнаружения, которая задается исследователем.

Таким образом, разработанная методика позволяет определить значения интервалов плотности ЗПП, при которых, с использованием методики дистанционного обнаружения ЗПП, основанной на регистрации ОРИИ, с необходимой достоверностью определяется наличие (отсутствие) загрязнителя.

Устройство и принцип действия комплекса распознавания типов и толщин слоев подстилающей поверхности

Принцип действия комплекса распознавания основан на регистрации обратно-отраженного радиочастотного видеоимпульса и импульса гамма-излучения от границ разнородных сред, схема которого представлена на рис. 5 [17].

Основываясь на принципе действия рассматриваемого комплекса была предложена структурная схема, описывающая работу приемной части и порядок обработки информации (рис. 6).

В качестве генератора импульсов гамма-излучения предложено использовать «револьверный» способ преобразования непрерывного гамма-излучения от радиоизотопного источника в импульсное. Данный способ заключается в следующем. Путем вращения источника над коллиматором, выполненным в виде свинцового кольца с одним или несколькими отверстиями, обеспечивается попадание пучка гамма-излучения в эти отверстия и, тем самым, имитируются импульсы. Вращать можно и наоборот: коллиматор под неподвижным источником.

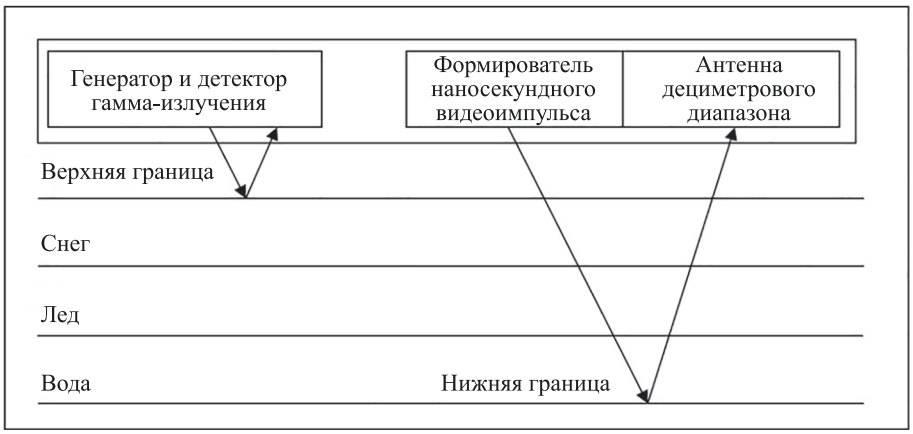


Рис. 5. Схема, поясняющая принцип действия комплекса распознавания сред

Fig. 5. The principle of operation of the environments recognition complex

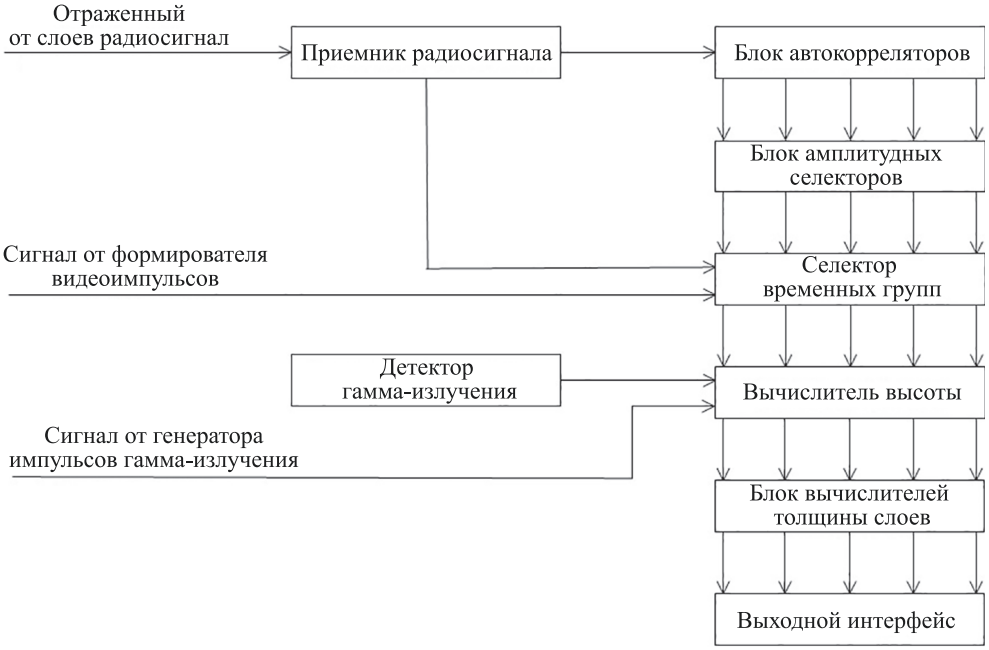


Рис. 6. Структурная схема приемной части комплекса распознавания сред

Fig. 6. Block diagram of the receiving part of environments recognition module

Рассмотрим алгоритм работы приемной части комплекса. Пусть передающая антенна испускает радиосигнал в дециметровом диапазоне длин волн по пяти различным частотам, под начальным углом падения 0,01° к нормали границы раздела сред. Вместе с этим сигналом синхронно, за счет строгой привязки к стандарту частоты для данной системы (порядка 10^{-10} с$^{-1}$), испускается пучок фотонов с энергией, например 0,66 МэВ, под углом к нормали 45°. Тогда приемная антенна примет отраженную часть от испущенной волны. На приемную антенну поступает радиоимпульс с входными данными для радиоволны: время регистрации сигнала, частота, амплитуда огибающей, фаза в момент регистрации; для гамма-излучения: энергия пришедших фотонов, время их регистрации, при этом детектор гамма-излучения находится под углом в 45° к нормали, т. е. зеркально углу падения.

Принимаемый радиосигнал проходит оцифровку и поступает в компьютер модуля, где происходит его регистрация с привязкой ко времени поступления относительно излученного. Затем временной маркер разделяется на два канала: один проходит с сигналом в автокоррелятор, а другой попадает в селектор временных групп. Сигнал, проходя через автокорреляторы, после соответствующего отклика на одном из них, попадает в амплитудный селектор, где происходит сравнение сигнала с табличным значением его амплитуды. Тем самым происходит распознавание типа среды. После амплитудного селектора сигнал подается на селектор временных групп, где маркер времени, представляющий собой видеосигнал из передающей части, и маркер времени, отделенный в приемнике радиоволны, сравниваются между собой. Результатом сравнения является выстраивание сигналов в хронологическом порядке их приема.

Одновременно с приемом радиосигнала принимается импульс интенсивности ОРИИ I^{cle} или I^{pol}, который так же проходит привязку ко времени и оцифровку в приемнике гамма-излучения. Затем, после формирования временных групп, данные радиоимпульса и гамма-излучения используются для вычисления высоты полета воздушного средства. Далее рассчитываются толщины остальных слоев ландшафта. Рассчитанные и структурированные данные отправляются на выходной интерфейс и визуализируются.

Анализируя принцип действия приемной части данного комплекса, можно сказать, что гибкая система анализа слоистых сред позволит решать целый класс задач инженерной разведки, прикладной геодезии, эколого-ресурсных служб. Однако такие факторы, как большие размеры модуля и не стабильность измеряемых параметров, в частности амплитуды, не позволяют использовать его во время ведения боевых действий, когда необходимо в режиме реального времени проложить безопасный путь военной и специальной технике, а также личному составу. Решением этой проблемы может быть: использование нескольких диапазонов радиоволн; измерение более устойчивых параметров, несущих информацию о среде (например, набег фазы, поляризация).

Методика оценивания экологической обстановки

Предложенный способ активного поиска позволяет идентифицировать признаки вреда, причиненного почвам как объекту охраны окружающей среды, который образовался в результате поступления в почву загрязняющих веществ, а также не ликвидирован вследствие низкого качества проведения мероприятий рекультивации загрязненной почвы. Учитывая действующие экологические стандарты, существует объективная необходимость при оценивании экологической обстановки использовать коэффициент оценки экологического состояния почвы, характеризующий несоблюдение нормативов качества окружающей среды для почв, включая нормативы предельно допустимых концентраций и ориентировочно допустимых концентраций химических веществ в почве. В результате:

— для оценивания степени загрязнения $С_{загр}$ и качества рекультивации загрязненных земель используем выражение:

$$С_{загр} = \frac{I_0^{sys} - I^{cle}}{I^{cle}};$$

— для оценивания ориентировочного экологического ущерба $У_{ор}$ (если $С_{загр} \in 0 \ldots 1$) применим выражение:

$$У_{ор} = С_{загр} S_{загр} h_{in} К_{исп} Т_{вр},$$

где $S_{загр}$ — площадь загрязнения (м^2); h_{in} — глубина загрязнения, порчи почв; $К_{исп}$ — показатель, учитывающий категорию земель и вид разрешенного использования земельного участка; $Т_{вр}$ — такса для исчисления размера вреда, причиненного почвам[1];

— для мониторинга экологического состояния почвы $М_{эсп}$ вблизи объектов промышленности и социальной инфраструктуры (рассчитывается, как частное в отношении средних характеристик распространения электромагнитной волны текущих и ранее проведенных исследований почвы):

$$М_{эсп} = I_{ср}^{t} / I_{ср}^{t-1},$$

где $I_{ср}^{t}$ и $I_{ср}^{t-1}$ — средние интенсивности ОРИИ на соответствующей территории для текущих и ранее проведенных исследований почвы.

В случае, если значения $С_{загр}$ или $М_{эсп}$ менее или равны 1, то экологическое состояние почвы оценивается как не испытывающее негативного воздействия.

Заключение

Предлагаемый комплекс сбора и обработки экологической информации может решать народно-хозяйственные задачи совместно с военно-прикладными.

[1] Методика исчисления размера вреда, причиненного почвам как объекту охраны окружающей среды. Утв. приказом Министерства природных ресурсов и экологии Российской Федерации 08.07.2010 № 238 (ред. от 18.11.2021): офиц. текст. М.: Маркетинг, 2021.11 с.

К подобным типовым задачам в том числе относятся: инженерно-геодезические изыскания различных участков, площадок и трасс с целью составления планов и профилей; инженерно-геодезическое проектирование — преобразование рельефа местности для инженерных целей, подготовка геодезических данных для строительных работ; детальная разбивка осей сооружений; выверка конструкций и технологического оборудования в плане и по высоте; наблюдение за деформациями зданий и сооружений; установление характера и объема разрушений, завалов, пожаров, затоплений, естественных препятствий и отыскание пути их обхода; определение проходимости местности, состояния дорог и мостов.

Отметим, что для решения данной задачи впервые было предложено совместное использование радиочастотного диапазона электромагнитных волн и гамма-излучения. При этом радиочастотный диапазон позволяет исследовать структуру ландшафта, а гамма-излучение по обратно-рассеянному ионизирующему излучению — тип загрязнителя, а также обеспечить высокую точность при измерении расстояния от модуля до верхнего слоя подстилающей поверхности. Кроме того, использование гамма-излучения обеспечит надежную работу данного комплекса при плохих погодных условиях.

Таким образом, решение военно-политических и эколого-ресурсных задач в районах Крайнего Севера и Дальнего Востока может быть осуществлено при существенном прорыве в вопросах совершенствования технического облика средств мониторинга окружающей среды.

Литература

1. Притужалова О.А., Радченко К.С. Проблемы проведения производственного экологического контроля в строительных организациях // Строительство и техногенная безопасность. 2019. № 14(66). С. 145–154.
2. Куклина П.П., Качукова А.А., Кварацхелия Е.В., Песня Е.И., Бляхарский Д.П. Беспилотное воздушное лазерное сканирование // Геодезия, картография, геоинформатика и кадастры. производство и образование: Сборник материалов IV всероссийской научно-практической конференции. 2021. С. 108–112.
3. Васильев В.Н. и др. Спектры излучения рентгеновских установок. М.: Энергоатомиздат, 1990. 143 с.
4. Храмов А.В., Ермолаев А.А., Шалашова А.И., Контрош Л.В. Возможности экологического мониторинга с применением БПЛА в России // Известия СПбГЭТУ ЛЭТИ. 2017. № 7. С. 79–84.
5. Хабаров Д.А., Адиев Т.С., Попова О.О., Чугунов В.А., Кожевников В.А. Анализ современных технологий дистанционного зондирования Земли // Московский экономический журнал. 2019. № 1. С. 181–190. https://doi.org/10.24411/2413-046X-2019-11068
6. Абрамов Н.С., Макаров Д.А., Талалаев А.А., Фраленко В.П. Современные методы интеллектуальной обработки данных ДЗЗ // Программные системы: теория и приложения. 2018. Т. 9. № 4(39). С. 417–442. https://doi.org/10.25209/2079-3316-2018-9-4-417-442
7. Григорьев А.Н., Шабаков Е.И., Дементьев А.Н., Романов А.А. Метод сокращения избыточности данных дистанционного зондирования из космоса // Известия вузов. Приборостроение. 2016. Т. 59. № 1. С. 38–44. https://doi.org/10.17586/0021-3454-2016-59-1-38-44
8. Севрюкова Е.А., Волкова Е.А., Дорошенко В.А., Солодков А.В. Разработка макета распределенной сенсорной системы экологического мониторинга // Известия вузов России. Радиоэлектроника. 2021. Т. 24. № 3. С. 98–108. https://doi.org/10.32603/1993-8985-2021-24-3-98-108
9. Трофимова Т.И. Физика. 500 основных законов и формул. М.: Высшая школа, 2005. 64 с.
10. Лукьянов Л.А., Свитнев И.В., Харитонова Е.А., Гаврилов И.Е. Метод дистанционного контроля радиационных параметров космических аппаратов на основе рентгенофлуоресцентного анализа // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22. № 4. С. 650–658. https://doi.org/10.17586/2226-1494-2022-22-4-650-658
11. Чернецкая И.Е., Спевакова С.В. Мультиспектральное оптико-электронное устройство для автономной мобильной платформы экологического мониторинга // Труды МАИ. 2020. № 114. С. 13. https://doi.org/10.34759/trd-2020-114-14
12. Дудкин Д.С., Найданов А.Ф., Свитнев И.В. Симулятор аппаратно-программного комплекса распознавания типов и толщин слоев подстилающей поверхности методом зондирования электромагнитным излучением дециметрового, сантиметрового и

References

1. Prituzhalova O.A., Radchenko K.S. Problems of industrial environmental control in building companies. *Construction and Industrial Safety*, 2019, no. 14(66), pp. 145–154. (in Russian)
2. Kuklina P.P., Kachukova A.A., Kvarachelija E.V., Pesnja E.I., Bljaharskij D.P. Unmanned aerial laser scanning. *Geodesy, cartography, geoinformatics and cadastre. From idea to application. Collection of abstracts of the IV Russian Conference on Applied Science*, 2021, pp. 108–112. (in Russian)
3. Vasilev V.N. et al. *Emission Spectra of X-ray Installations*. Moscow, Jenergoatomizdat Publ., 1990, 143 p. (in Russian)
4. Khramov A.V., Ermolaev A.A., Shalashova A.I., Kontrosh L.V. Possibilities of environmental monitoring with UAV use in Russia. *Proceedings of Saint Petersburg Electrotechnical University*, 2017, no. 7, pp. 79–84. (in Russian)
5. Khabarov D.A., Adiev T.S., Popova O.O., Chugunov V.A., Kozhevnikov V.A. Analysis of modern technologies for remote sensing of the Earth. *Moscow Economic Journal*, 2019, no. 1, pp. 181–190. (in Russian). https://doi.org/10.24411/2413-046X-2019-11068
6. Abramov N.S., Makarov D.A., Talalaev A.A., Fralenko V.P. Modern methods for intelligent processing of earth remote sensing data. *Program Systems: Theory and Applications*, 2018, vol. 9, no. 4(39), pp. 417–442. (in Russian). https://doi.org/10.25209/2079-3316-2018-9-4-417-442
7. Grigoriev A.N., Shabakov E.I., Dementiev A.N., Romanov A.A. Method to reduce data redundancy in remote sensing from space. *Journal of Instrument Engineering*, 2016, vol. 59, no. 1, pp. 38–44. (in Russian). https://doi.org/10.17586/0021-3454-2016-59-1-38-44
8. Sevryukova E.A., Volkova E.A., Doroshenko V.A., Solodkov A.V. A prototype unit of a distributed sensor system for ecological monitoring. *Journal of the Russian Universities. Radioelectronics*, 2021, vol. 24, no. 3, pp. 98–108. https://doi.org/10.32603/1993-8985-2021-24-3-98-108
9. Trofimova T.I. *500 Basic Laws and Formulas*. Moscow, Vysshaja shkola Publ., 2005, 64 p. (in Russian)
10. Lukyanova L.A., Svitnev I.V., Kharitonova E.A., Gavrilov I.E. Method for remote control of radiation parameters of spacecraft based on X-ray fluorescence analysis. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 4, pp. 650–658. (in Russian). https://doi.org/10.17586/2226-1494-2022-22-4-650-658
11. Chernetskaya I., Spevakova S. Multispectral electronic device for autonomous mobile platform of ecological monitoring. *Trudy MAI*, 2020, no. 114, pp. 13. (in Russian). https://doi.org/10.34759/trd-2020-114-14
12. Dudkin D.S., Naidanov A.F., Svitnev I.V. Simulator of a hardware-software complex for recognizing the types and layers thicknesses of the underlying surface by probing with electromagnetic radiation in the decimeter, centimeter and X-ray ranges. *Certificate of the state computer program registration RU2018614834*, 2018. (in Russian)

рентгеновского диапазонов. Свидетельство о государственной регистрации программы для ЭВМ № RU2018614834. 2018.
13. Алексеев Н.В. Новый взгляд на эффект Комптона // Прорывные научные исследования как двигатель науки: сборник статей международной научно-практической конференции. М.: Аэтерна, 2019. С. 9–15.
14. Кулганов В.А., Свитнев И.В., Соколов Д.А., Нигматуллин Р.А. Метод обнаружения загрязнений подстилающей поверхности по характеристикам обратно рассеянного ионизирующего излучения // Материалы IV Всероссийской научной конференции «Экология и космос» имени академика К.Я. Кондратьева / под общ.ред. Ю.В.Кулешова. СПб.: ВКА имени А.Ф.Можайского, 2020. С. 206–212.
15. Соколов Д.А., Косырев С.В., Кислицына И.А. Методика расчета содержания загрязнителя на подстилающей поверхности с заданной вероятностью его определения // Экология и развитие общества. 2021. № 1(35). С. 31–38.
16. Любушин Н.П., Брикач Г.Е. Использование обобщенной функции желательности Харрингтона в многопараметрических экономических задачах // Экономический анализ: теория и практика. 2014. № 18(369). С. 2–10.
17. Дудкин Д.С., Найданов А.Ф., Свитнев И.В. Модуль распознавания типов и толщин слоев подстилающей поверхности в условиях Крайнего Севера // Инновации. 2018. № 11(241). С. 116–121.

13. Alekseev N.V. New look at the Compton effect. *Breakthrough scientific research as the engine of science: Proc. of the International Scientific and Practical Conference*. Moscow, Ajeterna Publ., 2019, pp. 9–15. (in Russian)
14. Kulganov V.A., Svitnev I.V., Sokolov D.A., Nigmatullin R.A. Method for detecting the contamination of the underlying surface based on the characteristics of backscattered ionizing radiation. *Proc. of the All-Russian Scientific Conference "Ecology and Space" named after Academician K.Ya. Kondratiev*, 2020, pp. 206–212. (in Russian)
15. Sokolov D.A., Kosyrev S.V., Kislitsyna I.A. Rocedure for calculating the contaminant content on the substrate surface with a given probability of its reliable determination. *Ecology and Society Development*, 2021, no. 1(35), pp. 31–38. (in Russian)
16. Liubushin N.P., Brikach G.E. Harrington's desirability generalized function in multiple parameter economic tasks. *Economic Analysis: Theory and Practice*, 2014, no. 18(369), pp. 2–10. (in Russian)
17. Dudkin D.S., Naydanov A.F., Svitnev I.V. Module of recognition of types and thicknesses of the underlying surface under the conditions of the far north. *Innovations*, 2018, no. 11(241), pp. 116–121. (in Russian)

Авторы

Свитнев Игорь Владимирович — кандидат военных наук, доцент, доцент, Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация, SC 10040324500, https://orcid.org/0000-0002-8862-1797, isvitnev@mail.ru
Найданов Александр Фотеевич — кандидат технических наук, доцент, старший преподаватель, Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация, https://orcid.org/0009-0003-4242-1700, naydanov@bk.ru
Вилков Алексей Владимирович — кандидат технических наук, преподаватель, Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация, https://orcid.org/0000-0002-9300-3547, Leha_kostroma@mail.ru
Соколов Дмитрий Александрович — кандидат технических наук, преподаватель, Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация, https://orcid.org/0009-0001-0111-403X, falcon.85@bk.ru
Лебедев Михаил Юрьевич — кандидат технических наук, преподаватель, Военно-космическая академия имени А.Ф. Можайского, Санкт-Петербург, 197198, Российская Федерация, https://orcid.org/0009-0003-8876-7846, Lebedev.m1ha1l@yandex.ru
Харитонова Елена Александровна — кандидат медицинских наук, доцент, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, SC 57885000100, https://orcid.org/0000-0002-4851-2756, Xaritonova_ea@mail.ru
Лукьянова Людмила Анатольевна — старший преподаватель, Санкт-Петербургский государственный университет, Санкт-Петербург, 199034, Российская Федерация, https://orcid.org/0000-0001-5458-3019, lalukianova@mail.ru

Authors

Igor V. Svitnev — PhD (Military Sciences), Associate Professor, Associate Professor, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, SC 10040324500, https://orcid.org/0000-0002-8862-1797, isvitnev@mail.ru
Alexander F. Naydanov — PhD, Associate Professor, Senior Lecturer, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, https://orcid.org/0009-0003-4242-1700, naydanov@bk.ru
Alexey V. Vilkov — PhD, Lecturer, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, https://orcid.org/0000-0002-9300-3547, Leha_kostroma@mail.ru

Dmitry A. Sokolov — PhD, Lecturer, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, https://orcid.org/0009-0001-0111-403X, falcon.85@bk.ru

Mikhail Yu. Lebedev — PhD, Lecturer, Mozhaisky Military Aerospace Academy, Saint Petersburg, 197198, Russian Federation, https://orcid.org/0009-0003-8876-7846, Lebedev.m1ha1l@yandex.ru

Elena A. Kharitonova — PhD (Medicine), Associate Professor, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, SC 57885000100, https://orcid.org/0000-0002-4851-2756, Xaritonova_ea@mail.ru
Lyudmila A. Lukyanova — Senior Lecturer, St. Petersburg State University (SPbSU), Saint Petersburg, 199034, Russian Federation, https://orcid.org/0000-0001-5458-3019, lalukianova@mail.ru

Статья поступила в редакцию 27.06.2023
Одобрена после рецензирования 15.12.2023
Принята к печати 26.01.2024

Received 27.06.2023
Approved after reviewing 15.12.2023
Accepted 26.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

doi: 10.17586/2226-1494-2024-24-1-156-164
УДК 004.85

Использование технологий машинного обучения при решении задачи классификации сигналов мониторинга инфразвукового фона

Иван Николаевич Фролов[1], Николай Георгиевич Кудрявцев[2], Варвара Юрьевна Сафонова[3], Дмитрий Владимирович Кудин[4]✉

[1,2,3] Горно-Алтайский государственный университет, Горно-Алтайск, 649000, Российская Федерация
[4] Геофизический центр Российской академии наук, Москва, 119296, Российская Федерация

[1] xfin@bk.ru, https://orcid.org/0000-0001-9176-6965
[2] ngkudr@mail.ru, https://orcid.org/0000-0003-1327-5188
[3] safonova_varvara@mail.ru, https://orcid.org/0000-0002-8043-4014
[4] d.kudin@gcras.ru✉, https://orcid.org/0000-0002-9174-224X

Аннотация

Введение. Широко известно, что среди звуковых сигналов, порождаемых природными и техногенными явлениями, наиболее долгоживущими являются волны частотой менее 20 Гц, именуемые инфразвуком. Это свойство позволяет отслеживать путем инфразвукового мониторинга возникновение высокоэнергетических событий в региональных масштабах на расстоянии до 200–300 км. Вместе с тем выделение полезных инфразвуковых сигналов из фонового шума является нетривиальной задачей при обработке сигналов в реальном времени и постфактум. В работе предлагается новый метод классификации специфических сигналов в данных инфразвукового мониторинга с использованием перестановочной энтропии Шеннона и векторов распределения частот встречаемости перестановок последовательных выборочных значений ранга 3 (количество элементов перестановки). **Метод.** Для оценки применимости предлагаемого энтропийного метода классификации использованы два метода машинного обучения — метод случайных лесов и классический нейросетевой подход. Методы реализованы на языке Python с использованием библиотек Scikit-lean, TensorFlow и Keras. Выполнена оценка качества классификации по сравнению с традиционным частотным методом выделения классов, основанном на использовании преобразования Фурье. Распознавание проведено на подготовленных данных инфразвукового мониторинга на территории Республики Алтай. **Основные результаты.** Результаты вычислительного эксперимента по разделению пяти классов сигналов показали, что классификация предложенным методом дает одинаковые результаты распознавания нейронной сетью. При этом по сравнению с частотной классификацией исходных данных точность распознавания составила 51–58 %. Для метода случайных лесов точность распознавания частотных классов равна 51 %, а для метода перестановочной энтропии — 45 %. **Обсуждение.** Анализ результатов вычислительного эксперимента показал конкурентноспособность метода классификации перестановочной энтропии при распознавании инфразвуковых сигналов. Предлагаемый метод реализуется значительно проще при поточной обработке сигнала в низкопотребляющих микроконтроллерных системах. Планируется дальнейшая апробация метода на пунктах регистрации инфразвуковых сигналов, а также в составе системы обработки данных инфразвукового мониторинга для выделения событий в реальном режиме времени.

Ключевые слова

машинное обучение, модель случайного леса, искусственная нейронная сеть, инфразвук, перестановочная энтропия, классификация фрагментов временных рядов

Благодарности

Исследование выполнено за счет средств гранта Российского научного фонда и Министерства образования и науки Республики Алтай № 23-21-10087.

Ссылка для цитирования: Фролов И.Н., Кудрявцев Н.Г., Сафонова В.Ю., Кудин Д.В. Использование технологий машинного обучения при решении задачи классификации сигналов мониторинга инфразвукового фона // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 156–164. doi: 10.17586/2226-1494-2024-24-1-156-164

© Фролов И.Н., Кудрявцев Н.Г., Сафонова В.Ю., Кудин Д.В., 2024

Using machine learning technologies to solve the problem of classifying infrasound background monitoring signals

Ivan N. Frolov[1], Nikolai G. Kudryavtsev[2], Varvara Yu. Safonova[3], Dmitry V. Kudin[4]✉

[1,2,3] Gorno-Altaisk State University (GASU), Gorno-Altaisk, 649000, Russian Federation

[4] Geophysical Center of The Russian Academy of Sciences (GCRAS), Moscow, 119296, Russian Federation

[1] xfin@bk.ru, https://orcid.org/0000-0001-9176-6965

[2] ngkudr@mail.ru, https://orcid.org/0000-0003-1327-5188

[3] safonova_varvara@mail.ru, https://orcid.org/0000-0002-8043-4014

[4] d.kudin@gcras.ru✉, https://orcid.org/0000-0002-9174-224X

Abstract

It is widely known that among sound signals generated by natural and anthropogenic phenomena, the most long-lived are waves of frequency less than 20 Hz, called infrasound. This property allows tracking at a distance by infrasound monitoring the occurrence of high-energy events on regional scales (up to 200–300 km). At the same time, the separation of useful infrasound signals from background noise is a non-trivial task in real-time and post-facto signal processing. In this paper we propose a new method for classification of specific signals in infrasound monitoring data using Shannon permutation entropy and vectors of frequency distribution of occurrence frequencies of permutations of consecutive sample values of rank 3 (number of permutation elements). To evaluate the validity of the proposed entropy-based classification method, two machine learning methods — random forest method and classical neural network approach — implemented in Python language using Scikit-lean, TensorFlow and Keras libraries were used. The classification quality was evaluated against the traditional frequency-based method of class extraction based on Fourier transform. Recognition was performed on the prepared infrasound monitoring data in the Altai Republic. The results of computational experiment on the separation of 5 classes of signals showed that classification by the proposed method gives the same results of recognition by neural network with in comparison with frequency classification of the original data; the recognition accuracy was 51–58 %. For the random forests method, the recognition accuracy of frequency classes was slightly higher: 51 % vs. 45 % for classes using the permutation entropy method. The analysis of the results of the computational experiment shows sufficient competitiveness of the method of classification by permutation entropy in the recognition of infrasound signals. In addition, the proposed method is much easier to implement for inline signal processing in low-consumption microcontroller systems. The next step is to test the method at infrasound signal registration points and as part of the infrasound monitoring data processing system for real-time event detection.

Keywords

machine learning, random forest model, artificial neural network, infrasound, permutation entropy, classification of time series fragments

Acknowledgements

The research was carried out using a grant from the Russian Science Foundation (RSF) and the Ministry of Education and Science of the Altai Republic No. 23-21-10087.

For citation: Frolov I.N., Kudryavtsev N.G., Safonova V.Yu., Kudin D.V. Using machine learning technologies to solve the problem of classifying infrasound background monitoring signals. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 156–164 (in Russian). doi: 10.17586/2226-1494-2024-24-1-156-164

Введение

Многие события как природного, так и техногенного характера (сейсмическая или грозовая активность, извержения вулканов, лесные пожары, метеороиды, падение фрагментов ракет, химические и ядерные взрывы и т. п.) сопровождаются возникновением инфразвуковых волн, которые благодаря своим физическим характеристикам могут быть обнаружены на значительных расстояниях от источника событий. Мониторинг инфразвуковой обстановки позволяет исследовать, а иногда даже и предсказывать некоторые катастрофические явления. Для этого необходимо научиться принимать решения о наличии или отсутствии в исследуемых фрагментах временных рядов сигналов с заданными характеристиками, т. е. решать задачу классификации исходных данных. Одним из подходов к разработке автоматических классификаторов является использование методов машинного обучения.

Так, в работе [1] содержатся результаты исследования использования глубокого обучения со сверточными нейронными сетями (Cable News Network, CNN) для классификации сигналов электроэнцефалографии (ЭЭГ) при депрессии. Исследована эффективность метода CNN в сравнении с другими методами классификации и проанализирована его производительность на наборе данных сигналов ЭЭГ пациентов, страдающих депрессией. Результаты исследования показали, что CNN демонстрирует высокую точность классификации и может быть эффективным инструментом для диагностики депрессии на основе сигналов ЭЭГ.

Работа [2] посвящена исследованию методов классификации данных ЭЭГ с использованием техник машинного обучения. В работе рассмотрено применение различных алгоритмов машинного обучения для классификации сигналов ЭЭГ, таких как метод опорных векторов, случайный лес, нейронные сети и другие. Выполнен сравнительный анализ эффективности различных методов классификации на примере реальных данных ЭЭГ.

Разработчики методов машинного обучения в своих работах обращают особое внимание на важность подготовки наборов данных, подаваемых на вход выбранной модели машинного обучения. В большинстве работ в

качестве данных, подаваемых на вход применяемых моделей машинного обучения, исследователи используют векторы, полученные в результате преобразования Фурье.

Например, в работе [3] произведен анализ сигналов ЭЭГ пациентов с диагнозом эпилепсия, посредством которых в дальнейшем созданы обучающие датасеты для различных моделей нейронных сетей. С помощью применения языка Python, реализованы традиционные методы классификации, которые используют частотные данные исследований ЭЭГ, полученные путем извлечения признаков при помощи анализа Фурье с учетом диапазонов частот.

В [4] исследована возможность различения фокальных и нефокальных судорог на основе анализа сигналов ЭЭГ. Предложено применять метод скользящего режима сингулярного спектрального анализа с использованием спектра Фурье для обработки данных ЭЭГ и выделения характеристик, отличающих фокальные и нефокальные судороги.

В [5] описаны результаты исследования по классификации сигналов ЭЭГ посредством нейронной сети и последующему распознаванию признаков эпилепсии. Для классификации фокальных и нефокальных сигналов ЭЭГ предложен подход, основанный на частотной области (биспектр). Биспектр определяется как преобразование Фурье тройной корреляционной функции и используется для характеристики нелинейности и сбора информации, которая подходит для различения сигналов ЭЭГ.

Таким образом, изучение цифровых технологий, позволяющих автоматизировать решение задач классификации исследуемых явлений, позволяет выделить в качестве основных методов машинного обучения модели логистической регрессии, случайных лесов и искусственные нейронные сети [1–11]. Окончательный выбор подходящих для настоящей работы методов машинного обучения был определен после ознакомления с результатами экспериментов, описанных в работе [7], в которой авторами была проведена классификация акустических сигналов на примере шумов сердца. Классификация осуществлялась по особенностям спектра звукового сигнала с помощью методов логистической регрессии, случайных лесов и с использованием искусственной нейронной сети. Для каждого метода выполнен расчет метрики точности, полноты, F1, матрицы ошибок, а также время, затраченное на обучение и тестирование модели. В результате было установлено превосходство нейронной сети над методами случайного леса и логистической регрессии при классификации акустических сигналов. Отдельно отмечено, что доля неверно классифицированных больных сердец от общего количества больных сердец в тестовой выборке для искусственной нейронной сети составила всего 19 % (против 76 % и 31 % для методов логистической регрессии и случайных лесов). Исследователи установили, что улучшить качество распознавания патологий сердца по его звуковым характеристикам можно, увеличив объем набора данных, как больных, так и здоровых сердец, а также оценив важность признаков, применяемых при построении модели и используя признаки с большим качественным показателем. Кроме того, рассмотрена возможность применения CNN для повышения качества результатов классификации акустических сигналов.

В результате проведенного анализа авторы настоящей работы решили исследовать возможность использования метода случайных лесов и искусственных нейронных сетей, в качестве наиболее подходящих для разработки технологии классификации инфразвуковых сигналов. При этом для повышения эффективности вычислительных процедур, применяемых при подготовке классифицируемых наборов, было предложено использовать перестановочную энтропию Шеннона и векторы распределения частот встречаемости (оценок вероятности встречаемости) перестановок последовательных выборочных значений анализируемого временного ряда для $n = 3$ [12].

Цель работы — изучение возможности и получение сравнительных оценок эффективности использования метода случайных лесов и метода искусственных нейронных сетей при различных подходах к подготовке классифицируемых наборов в процессе решения задачи определения местоположения фрагментов временного ряда, содержащих инфразвуковые сигналы, имеющие заданные характеристики.

Метод

Современные технологии позволяют исследовать окружающий мир, используя возможности вычислительного эксперимента, который предполагает программную реализацию цифровых моделей исследуемых объектов или процессов, выполненных с высокой степенью детализации.

Для проведения вычислительного эксперимента использован язык программирования Python. С целью создания прогнозирующего классификатора на базе метода случайных лесов применены функции библиотеки Scikit-learn. Для реализации модели прогнозирующего классификатора на базе искусственных нейронных сетей использованы библиотеки TensorFlow и Keras. Особенности применения библиотеки Scikit-learn описаны в разделе «Основные результаты».

При подготовке классифицируемых наборов использованы специальные функции из библиотек Ordpy и Scipy. Для визуализации исследуемых фрагментов временных рядов применена библиотека Matplotlib.

В качестве первого метода машинного обучения, выбранного для исследования, был использован метод случайного леса. Благодаря своей высокой точности, простоте применения и защите от переобучения, данный метод является подходящим при проведении исследований и обработке любых объемов информации [2, 7].

Второй рассмотренный метод — нейронная сеть. В процессе эксперимента был выбран готовый класс нейронной сети, включающий все необходимые переменные и функции. Используемая нейронная сеть содержала четыре слоя. Первый или входной слой X соответствовал 7 классам входных признаков. Второй и третий внутренние слои H1 и H2 содержали по 10 нейронов. На выходе был определен слой Z, состоя-

щий из пяти классификаторов. На том основании, что выбранная нейронная сеть представляет собой классический вариант нейронной сети, был использован классический вариант функций прямого и обратного распространений ошибок.

Основные результаты

Для проведения вычислительного эксперимента по сравнительному анализу прогнозирующих классификаторов, реализованных на базе методов машинного обучения, на первом шаге экспертной группой были выделены и описаны пять хорошо различимых классов выборочных фрагментов временных рядов, содержащих данные инфразвукового мониторинга. В качестве иллюстрации получены графики классифицируемых сигналов во временной области и спектры этих же сигналов в частотной области. Введем названия классов и их буквенные обозначения: шум (nn); сигнал-шум (sn); хороший сигнал (gs); плохой сигнал (bs); смешанный сигнал (ds).

Подготовим обучающие наборы. Один набор[1] — «энтропийный» содержал 100 векторов по 20 на каждый распознаваемый класс. При этом компоненты (координаты) векторов представляли собой информацию о 7 признаках (первые 6 признаков — оценки вероятностей «встречаемости» соответствующих перестановок для $n = 3$ в заданном фрагменте временного ряда, 7-ой признак — значение «сложности» фрагмента временного ряда, которое вычисляется как дополнение до единицы значения энтропии Шеннона данного фрагмента). 8-ая компонента вектора обозначала тип распознаваемого класса. Второй набор[2] — «частотный», основанный на использовании преобразования Фурье. Любой вектор из этого набора содержал 10 признаков (фактически пять пар «частота-амплитуда», имеющих наибольшие значения амплитуд в спектре классифицируемого сигнала).

В первой части вычислительного эксперимента выполним применение и оценку точности метода случайных лесов.

Стандартная задача машинного обучения на начальном этапе предполагает подготовку и загрузку набора данных. Приведем фрагмент программного кода и пять строк из загруженного списка данных:

```
dataset = pd.read_csv('/content/dataset_entropy.csv', header = None)
dataset.columns = ['p1','p2','p3','p4','p5','p6','p7','class_type']
print('Список данных: ' + str(dataset.shape))
Список данных: (100, 8)
```

	p1	p2	p3	p4	p5	p6	p7	class_type
0	0.28163	0.11077	0.11044	0.10543	0.10585	0.28588	0.0577	bs
1	0.22336	0.13832	0.13557	0.14757	0.15033	0.20485	0.0109	bs
2	0.22272	0.14464	0.14014	0.13413	0.13864	0.21972	0.0137	bs
3	0.23183	0.13409	0.14160	0.12281	0.11654	0.25313	0.0266	bs
4	0.24525	0.13038	0.13138	0.12738	0.12621	0.23941	0.0255	bs

Представленные в листинге *p1-p7* данные являются классифицируемыми признаками, подающимися на вход модели машинного обучения, *class_type* — предопределенный класс в тестовой и тренировочной выборке.

Далее каждому текстовому обозначению распознаваемого класса сопоставлено число, а все признаки объединены в один «вектор». Таким образом, каждый «вектор» признаков будет принадлежать какому-либо классу, обозначенному в виде числа:

```
factor = pd.factorize(dataset['class_type'])
dataset.class_type = factor[0]
define = factor[1]
#Классы 0 – bs, 1 – ds, 2 – gs, 3 – nn, 4 – sn
```

Данное действие необходимо при подготовке данных, поскольку метод случайных лесов библиотеки Scikit-learn может работать только с числовыми обозначениями классов [13].

На следующем шаге все «векторы» и соответствующие им значения классов в отдельных списках были объединены и подготовлен тестовый и тренировочный наборы данных для обучения модели. Обучающая выборка была использована для «настройки» параметров применяемой модели, а тестовая выборка — для оценки качества обученной модели. Отметим, что в отдельных случаях применяют дополнительную выборку — проверочную (*Validationset*). Данная выборка нужна для нахождения наилучших гиперпараметров модели, т. е. таких параметров, которые не могут быть оптимизированы в процессе обучения модели, но влияют на ее качество. Например, в качестве гиперпараметров могут выступать скрытые слои в нейронной сети, скорость обучения моделей и любые другие параметры, которые можно настроить в «ручном» режиме.

Набор данных принято делить на обучающую и тестовую выборки, в соотношении 70/30 или 80/20, где большая часть используется для обучения модели. Также разделение данных должно выполняться случайным образом, для исключения возможности попадания более «качественных» или «необычных» значений в одну из выборок. В результате обеспечивается более точная оценка качества модели. Для обучения модели

[1] [Электронный ресурс]. https://cloud.mail.ru/public/rft2/bDr8X5Xsq (дата обращения: 28.11.2023).

[2] [Электронный ресурс]. https://cloud.mail.ru/public/dSqH/NjgfkMpUX (дата обращения: 28.11.2023).

на относительно небольшой выборке данных можно применить метод перекрестной проверки *Cross-Validation*. Данный метод позволяет распределить сравнительно небольшие данные более эффективно, чем, например, деление 80/20. Отметим, что в настоящей работе не использован метод *Cross-Validation*, так как цель описываемого вычислительного эксперимента заключается не в максимальном улучшении каких-либо методов, а в сравнении полученных результатов при «прочих равных условиях» [14].

Еще одним немаловажным моментом при подготовке обучающей выборки является тот факт, что распределение данных по классам должно быть сбалансировано, т. е. какой-либо из классов не желательно представлять значительно меньшим количеством данных. Соблюдение этого правила позволит значительно повысить качество обучения и точность модели [15].

После подготовки и разделения выборки данных выполняется масштабирование признаков. Данный шаг очень важен в машинном обучении, так как он «помогает» алгоритму быстро найти наилучшее решение. Например, представленный в библиотеке Scikit-learn метод *StandartScaler* просто преобразует данные для их центрирования, удаляя среднее значение каждой функции, а затем масштабирует ее делением непостоянных характеристик на их стандартное отклонение. После этого можно осуществить обучение модели. Опытным путем определим следующие важные параметры для обучения модели методом случайного леса класса *RandomForestClassifier* библиотеки Scikit-learn:

— *n_estimators*: количество деревьев в случайной классификации леса. Так как при значительном росте количества деревьев в диапазоне более 12 наблюдается незначительный рост точности модели [2], определим число деревьев в случайном лесу равным 12;
— *criterion*: функция потерь, которая используется для измерения качества разделения данных. Из трех возможных вариантов при выборе *энтропии Шеннона* были получены лучшие результаты, чем при выборе *Джини*;
— *random_state*: контроль случайности начальной загрузки выборок, используемых при построении деревьев.

Приведем фрагмент кода, который демонстрирует этапы подготовки тестового и тренировочного наборов, масштабирование, задание гиперпараметров и запуск обучения модели:

```
X = dataset.iloc[:,0:7].values
y = dataset.iloc[:,7].values
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size = 0.2, random_state = 42)
XX_test = X_test
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
classifier = RandomForestClassifier(n_estimators = 12,
criterion = 'entropy',
            random_state = 42, oob_score=True)
classifier.fit(X_train, y_train)
```

В результате выполнения приведенного фрагмента кода объект *classifier* является обученной моделью вида «случайный лес».

Следующим шагом произведем оценки производительности и точности данной модели. Оценку эффективности обучаемой модели разделим на следующие этапы.

Этап 1. Выполнение прогнозирования классов для выборки тестовых данных.

Этап 2. Выполнение обратного преобразования числовых сопоставлений классов в текстовый эквивалент, которое состоит из:

— создания словаря для сопоставления таблиц из класса в текст;
— преобразования таблицы данных из числовых классов в текстовые;
— оценки производительности классификатора с использованием матрицы неточностей.

В качестве примера приведем оценку точностей обученных моделей для наборов данных, подготовленных с помощью «энтропийного» и «частотного» наборов.

Возьмем проверочный вектор класса gs для «энтропийного» набора

[0.31617,0.08727,0.09371,0.09227, 0.08655, 0.32403, 0.0941]

и для «частотного» набора

[27.225,137.364,27.258,397.07,27.291,278.729,27.324, 265.731,27.357,92.933].

Рассмотрим программную реализацию проверочной процедуры:

```
#Матрица неточностей
print(pd.crosstab(y_test, y_pred, rownames=['Actual
class'],
colnames=['Predicted Class']))
#Точность модели
print(classifier.oob_score_)
#Предсказанный класс
observation = [[0.31617, 0.08727, 0.09371, 0.09227,
0.08655, 0.32403, 0.0941]]
print(classifier.predict(observation))
#Вывод "важности" каждого признака
print(list(zip(dataset.columns[0:7], classifier.feature_
importances_))).
```

Получим вид результатов для «энтропийного» набора:

```
Матрица неточностей:
Predicted Class  bs  ds  gs  nn  sn
Actual class
    bs           3   1   0   1   0
    ds           1   1   1   0   2
    gs           1   1   1   0   0
    nn           0   0   0   3   1
    sn           1   1   0   1   0
```

Точность модели:
0.4 #40%
Предсказанный класс:
[4]
Процент важности каждого признака:
[(‘p1’,0.13),(‘p2’,0.18),(‘p3’,0.16),(‘p4’,0.12), (‘p5’,0.14), (‘p6’,0.11),(‘p7’,0.16)]

Приведем результаты для «частотного» набора.

Матрица неточностей:
Predicted Class bs ds gs nn sn
Actual class
bs 3 2 0 0 0
ds 0 3 2 0 0
gs 0 1 2 0 0
nn 0 0 0 4 0
sn 0 1 1 0 1
Точность модели:
0.4625 #46%
Предсказанный класс:
[2]
Процент важности каждого признака:
[(‘p1’, 0.04), (‘p2’, 0.20), (‘p3’, 0.05), (‘p4’, 0.22), (‘p5’, 0.03), (‘p6’, 0.15), (‘p7’, 0.05), (‘p8’, 0.14), (‘p9’, 0.07), (‘p10’, 0.05)]

Для сохранения обученной модели используем библиотеку joblib, с помощью которой можно сохранить и в дальнейшем загрузить какой-либо объект программы. В данном случае сохранен настроенный и обученный объект класса метода случайного леса *RandomForestClassifier*:

```
#Сохранение модели
joblib.dump(classifier, ‘randomforestmodel.pkl’)
```

Таким образом, обученная модель может использоваться для предварительной классификации выборки инфразвуковых сигналов, после чего можно производить отбор «полезных» данных для дальнейшей обработки.

Рассмотрим метод машинного обучения на основе искусственной нейронной сети. Подготовка обучающей выборки данных этого метода не отличается от метода случайного леса, поэтому описание данного шага аналогично. Функция прямого распространения заключается в прямом проходе через все слои нейронной сети с прогнозированием выхода для каждой эпохи. После этого, на основе разницы между спрогнозированным выходом и фактическими данными уже в процессе обратного распространения, производится обновление весов. Функция обратного распространения ошибки позволяет обновлять веса узлов нейронной сети и определять их важность для системы в целом. После выбора класса и задания структуры сети можно начать обучение нейросети и оценить точность данной модели.

Далее, подобно методу случайных лесов, опытным путем определим гиперпараметры для обучения и проведем сборку модели и ее обучение:

```
#Создаем модель сети
model = Sequential()
model.add(Dense(1000, activation=‘relu’, kernel_initializer=‘he_normal’,
                kernel_regularizer=regularizers.l2(0.01), input_shape=(n_features,)))
model.add(Dense(1000, activation=‘relu’, kernel_initializer=‘he_normal’,
                kernel_regularizer=regularizers.l2(0.01)))
#Функция активации - softmax
model.add(Dense(10, activation=‘softmax’))
#Сборка модели
model.compile(optimizer=‘adam’, loss=‘sparse_categorical_crossentropy’,
              metrics=[‘accuracy’])
#Автоматическая остановка и сохранение модели при достижении наилучших
      # параметров в процессе обучения
callbacks = [EarlyStopping(monitor=“accuracy”, patience=45, mode=‘max’),
        ModelCheckpoint(filepath=“best_model.h5”,
        monitor=“accuracy”,
        save_best_only=True)]
#Запуск обучения и тренировки модели
model.fit(X_train, y_train, epochs=1000, batch_size=32000, verbose=0,
          callbacks=callbacks, validation_data=(X_test, y_test))
```

Приведем полученные результаты для «энтропийного» обучающего набора:

Точность модели:
0.55
Время обучения и тренировки:
1/1 [==============================] - 0s 96ms/step
Предсказанный класс:
(class =2)

На рис. 1 показано графическое представление отчета о качестве работы нейросетевой модели с «энтропийным» обучающим и тестовым наборами. При этом оценка точности вычисляется как отношение суммы правильно-положительных и правильно-отрицательных прогнозов к общему количеству прогнозов. Данное отношение на графике представлено в виде дробного числа, принимающего значение от 0 до 1 (ось ординат, рис. 1, *b*). Отметим, что оценка точности модели выполняется как для обучающих наборов (в процессе обучения), так и для тестовых (проверочных) наборов. В последнем случае говорят о процедуре тестирования или валидации. Вычисление отношения количества неправильных прогнозов к общему количеству прогнозов, умноженное на 100 позволяет оценивать потерю или ошибку при прогнозировании, формируя таким образом функцию потерь (ось ординат, рис. 1, *a*). В англоязычных источниках функцию потерь обозначают на графиках как Error Loss или просто Error. При этом надо иметь в виду, что уменьшение потерь с увеличением количества эпох при обучении (тренировке) говорит о

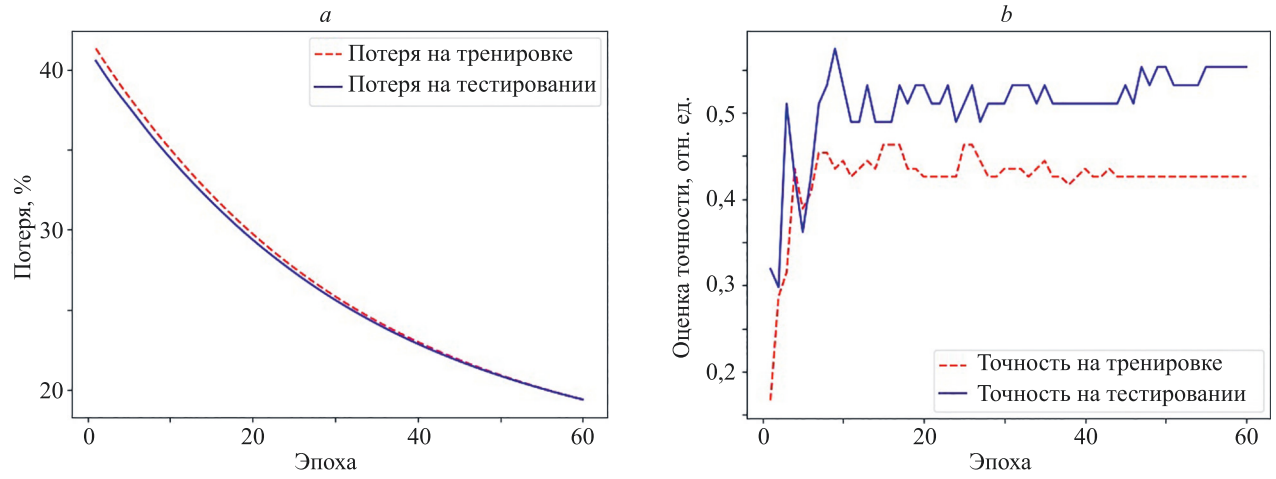


Рис. 1. Результаты для «энтропийного» обучающего набора: функции потерь на обучающем и тестовом наборах (*a*); функции оценок точности для обучающего и тестового наборов (*b*)

Fig. 1. Results for the entropy-based training set: error loss trend on the training and test sets (*a*); accuracy estimation trend for training and test sets (*b*)

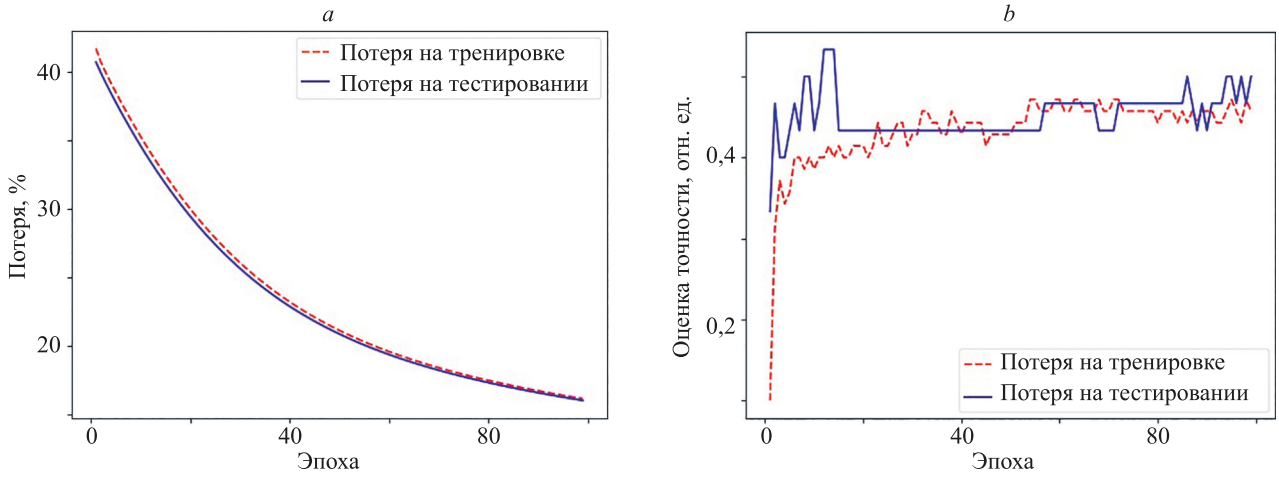


Рис. 2. Результаты для «частотного» обучающего набора: функции потерь на обучающем и тестовом наборах (*a*); функции оценок точности для обучающего и тестового наборов (*b*)

Fig. 2. Results for the frequency-based training set: error loss trend on the training and test sets (*a*); accuracy estimation trend for training and test sets (*b*)

хорошей динамике процесса обучения, который можно продолжать до тех пор, пока график функции потерь не станет параллелен оси абсцисс. Динамика, показывающая уменьшение значений функции потерь при тестировании (валидации), характеризует способность модели к «обобщению» новых данных.

Рассмотрим результаты работы нейронной сети при работе с «частотным» обучающим набором. На рис. 2 показано графическое представление отчета о качестве работы нейросети.

Точность модели:
0.5
Время обучения и тренировки:
1/1 [==============================] - 0s 73ms/step
Предсказанный класс:
(class=2)

В результате выполненного вычислительного эксперимента, можно сказать, что при обучении моделей использованы одинаковые по объему выборки данных, полученные на основании инфразвукового мониторинга в количестве 20 значений для каждого из пяти заданных классов. В итоге метод случайных лесов на «энтропийной» обучающей выборке набрал точность 45 %. Искусственная нейронная сеть при тонких настройках гиперпараметров получила точность в интервале от 51 до 58 %. Для сравнения эксперимент был повторен для «частотной» обучающей выборки. По результатам обучения, точность метода случайных лесов составила 51 %, а нейронная сеть показала такие же результаты — от 51 до 58 %.

Заключение

Результаты проведенного вычислительного эксперимента говорят о том, что выбор признаков методом

перестановочной энтропии работает, но объем самой выборки для обучающего набора стоит существенно увеличивать, поскольку по мере увеличения обучающего набора данных повышается точность обучаемой модели при распознавании.

Кроме того, результаты эксперимента позволяют отметить, что при практически одинаковом качестве распознавания «энтропийных» и «частотных» классифицируемых наборов, особенно для нейросетевой модели, использование «энтропийного» подхода позволяет реализовать более простую, а следовательно, и более эффективную вычислительную процедуру на этапе предварительной подготовки данных.

В дальнейшем планируется провести исследование работоспособности «энтропийного» подхода при реализации прогнозирующих классификаторов для большего числа классов распознаваемых сигналов с учетом повышения качества распознавания.

Литература

1. Sandheep P., Vineeth S., Poulose M., Subha D.P. Performance analysis of deep learning CNN in classification of depression EEG signals // Proc. of the TENCON 2019 — 2019 IEEE Region 10 Conference (TENCON). 2019. P. 1339–1344. https://doi.org/10.1109/TENCON.2019.8929254
2. Heyden M. Classification of EEG data using machine learning techniques: Master's thesis. Department of Automatic Control, Lund University, 2016.
3. Guerrero M.C., Parada J.S., Espitia H.E. EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks // Heliyon. 2021. V. 7. N 6. P. e07258. https://doi.org/10.1016/j.heliyon.2021.e07258
4. Siddharth T., Tripathy R., Pachori R. Discrimination of focal and non-focal seizures from EEG signals using sliding mode singular spectrum analysis // IEEE Sensors Journal. 2019. V. 19. N 24. P. 12286–12296. https://doi.org/10.1109/JSEN.2019.2939908
5. Sharma R., Sircar P., Pachori R.B. A new technique for classification of focal and nonfocal EEG signals using higher-order spectra // Journal of Mechanics in Medicine and Biology. 2019. V. 19. N 1. P. 1940010. https://doi.org/10.1142/S0219519419400104
6. Narayan Y. SEMG signal classification using KNN classifier with FD and TFD features // Materials Today: Proceedings. 2021. V. 37. Part 2. P. 3219–3225. https://doi.org/10.1016/j.matpr.2020.09.089
7. Shi Y., Davaslioglu K., Sagduyu Y.E., Headley W.C., Fowler M., Green G. Deep learning for RF signal classification in unknown and dynamic spectrum environments // Proc. of the 2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN). 2019. P. 1–10. https://doi.org/10.1109/DySPAN.2019.8935684
8. Shaker A.M., Tantawi M., Shedeed H.A., Tolba M.F. Generalization of convolutional neural networks for ECG classification using generative adversarial networks // IEEE Access. 2020. V. 8. P. 35592–35605. https://doi.org/10.1109/ACCESS.2020.2974712
9. Donmez H., Ozkurt N. Emotion classification from EEG signals in convolutional neural networks // Proc. of the 2019 Innovations in Intelligent Systems and Applications Conference (ASYU). 2019. P. 1–6. https://doi.org/10.1109/ASYU48272.2019.8946364
10. Кузин Д.А., Стаценко Л.Г., Анисимов П.Н., Смирнова М.М. Применение методов машинного обучения для классификации акустических сигналов по спектральным характеристикам // Известия СПбГЭТУ ЛЭТИ. 2021. № 3. С. 48–54.
11. Wu Y., Zhang J., Chen X., Zou S., Yang M. Research on the recognition of infrasound signal of nuclear explosion by SVM and CNN // IOP Conference Series: Earth and Environmental Science. 2020. V. 610. P. 012010. https://doi.org/10.1088/1755-1315/610/1/012010
12. Traversaro F., Redelico F.O., Risk M.R., Frery A.C., Rosso O.A. Bandt-Pompe symbolization dynamics for time series with tied values: A data-driven approach // Chaos. 2018. V. 28. N 7. P. 075502. https://doi.org/10.1063/1.5022021
13. Константинов А.В., Лештаев В.С. Подготовка признакового пространства нейросетевой модели для классификации сейсмоакустических событий системы PROGNOZ-ADS // 5-я конференция Международной научной школы академика РАН К.Н. Трубецкого «Проблемы и перспективы комплексного освоения и сохранения земных недр», 14–18 ноября 2022, Москва. М.: ИПКОН РАН, 2022. С. 108–112.
14. Jiao S., Geng B., Li Y., Zhang Q., Wang Q. Fluctuation-based reverse dispersion entropy and its applications to signal classification //

References

1. Sandheep P., Vineeth S., Poulose M., Subha D.P. Performance analysis of deep learning CNN in classification of depression EEG signals. *Proc. of the TENCON 2019 — 2019 IEEE Region 10 Conference (TENCON)*, 2019, pp. 1339–1344. https://doi.org/10.1109/TENCON.2019.8929254
2. Heyden M. *Classification of EEG data using machine learning techniques*. Master's thesis. Department of Automatic Control, Lund University, 2016.
3. Guerrero M.C., Parada J.S., Espitia H.E. EEG signal analysis using classification techniques: Logistic regression, artificial neural networks, support vector machines, and convolutional neural networks. *Heliyon*, 2021, vol. 7, no. 6, pp. e07258. https://doi.org/10.1016/j.heliyon.2021.e07258
4. Siddharth T., Tripathy R., Pachori R. Discrimination of focal and non-focal seizures from EEG signals using sliding mode singular spectrum analysis. *IEEE Sensors Journal*, 2019, vol. 19, no. 24, pp. 12286–12296. https://doi.org/10.1109/JSEN.2019.2939908
5. Sharma R., Sircar P., Pachori R.B. A new technique for classification of focal and nonfocal EEG signals using higher-order spectra. *Journal of Mechanics in Medicine and Biology*, 2019, vol. 19, no. 1, pp. 1940010. https://doi.org/10.1142/S0219519419400104
6. Narayan Y. SEMG signal classification using KNN classifier with FD and TFD features. *Materials Today: Proceedings*, 2021, vol. 37, part 2, pp. 3219–3225. https://doi.org/10.1016/j.matpr.2020.09.089
7. Shi Y., Davaslioglu K., Sagduyu Y.E., Headley W.C., Fowler M., Green G. Deep learning for RF signal classification in unknown and dynamic spectrum environments. *Proc. of the 2019 IEEE International Symposium on Dynamic Spectrum Access Networks (DySPAN)*, 2019, pp. 1–10. https://doi.org/10.1109/DySPAN.2019.8935684
8. Shaker A.M., Tantawi M., Shedeed H.A., Tolba M.F. Generalization of convolutional neural networks for ECG classification using generative adversarial network. *IEEE Access*, 2020, vol. 8, pp. 35592–35605. https://doi.org/10.1109/ACCESS.2020.2974712
9. Donmez H., Ozkurt N. Emotion classification from EEG signals in convolutional neural networks. *Proc. of the 2019 Innovations in Intelligent Systems and Applications Conference (ASYU)*, 2019, pp. 1–6. https://doi.org/10.1109/ASYU48272.2019.8946364
10. Kuzin D.A., Statsenko L.G., Anisimov P.N., Smirnova M.M. Applying machine learning methods to acoustic signal classification using spectrum characteristics. *Proceedings of Saint Petersburg Electrotechnical University*, 2021, no. 3, pp. 48–54. (in Russian)
11. Wu Y., Zhang J., Chen X., Zou S., Yang M. Research on the recognition of infrasound signal of nuclear explosion by SVM and CNN. *IOP Conference Series: Earth and Environmental Science*, 2020, vol. 610, pp. 012010. https://doi.org/10.1088/1755-1315/610/1/012010
12. Traversaro F., Redelico F.O., Risk M.R., Frery A.C., Rosso O.A. Bandt-Pompe symbolization dynamics for time series with tied values: A data-driven approach. *Chaos*, 2018, vol. 28, no. 7, pp. 075502. https://doi.org/10.1063/1.5022021
13. Konstantinov A.V., Leshtaev V.S. Preparing the feature space of a neural network model for the classification of seismoacoustic events of the PROGNOZ-ADS system. *Proc. of the 5th conference of the International Scientific School of Academician K.N. Trubetskoy from the Russian Academy of Sciences "Problems and prospects for the integrated development and conservation of the earth interior"*. Moscow, IPKON RAS, 2022, pp. 108–112. (in Russian)
14. Jiao S., Geng B., Li Y., Zhang Q., Wang Q. Fluctuation-based reverse dispersion entropy and its applications to signal classification. *Applied*

Applied Acoustics. 2021. V. 175. P. 107857. https://doi.org/10.1016/j.apacoust.2020.107857
15. Кузнецов И.А. Методы и алгоритмы машинного обучения для предобработки и классификации слабоструктурированных текстовых данных в научных рекомендательных системах: диссертация на соискание ученой степени кандидата технических наук. М.: МИФИ, 2019. 127 с.

Acoustics, 2021, vol. 175, pp. 107857. https://doi.org/10.1016/j.apacoust.2020.107857
15. Kuznetcov I.A. *Machine learning methods and algorithms for preprocessing and classification of semi-structured text data in scientific recommendation systems*. Dissertation for the degree of candidate of technical sciences. Moscow, MEPhI, 2019, 127 p. (in Russian)

Авторы

Фролов Иван Николаевич — инженер-программист, Горно-Алтайский государственный университет, Горно-Алтайск, 649000, Российская Федерация, https://orcid.org/0000-0001-9176-6965, xfin@bk.ru

Кудрявцев Николай Георгиевич — кандидат технических наук, старший научный сотрудник, Горно-Алтайский государственный университет, Горно-Алтайск, 649000, Российская Федерация, sc 57204835030, https://orcid.org/0000-0003-1327-5188, ngkudr@mail.ru

Сафонова Варвара Юрьевна — ассистент, Горно-Алтайский государственный университет, Горно-Алтайск, 649000, Российская Федерация, sc 57222578674, https://orcid.org/0000-0002-8043-4014, safonova_varvara@mail.ru

Кудин Дмитрий Владимирович — кандидат технических наук, старший научный сотрудник, Геофизический центр Российской академии наук, Москва, 119296, Российская Федерация, sc 56025952800, https://orcid.org/0000-0002-9174-224X, d.kudin@gcras.ru

Authors

Ivan N. Frolov — Software Engineer, Gorno-Altaisk State University (GASU), Gorno-Altaisk, 649000, Russian Federation, https://orcid.org/0000-0001-9176-6965, xfin@bk.ru

Nikolai G. Kudryavtsev — PhD, Senior Researcher, Gorno-Altaisk State University (GASU), Gorno-Altaisk, 649000, Russian Federation, sc 57204835030, https://orcid.org/0000-0003-1327-5188, ngkudr@mail.ru

Varvara Yu. Safonova — Assistant, Gorno-Altaisk State University (GASU), Gorno-Altaisk, 649000, Russian Federation, sc 57222578674, https://orcid.org/0000-0002-8043-4014, safonova_varvara@mail.ru

Dmitry V. Kudin — PhD, Senior Researcher, Geophysical Center of The Russian Academy of Sciences (GCRAS), Moscow, 119296, Russian Federation, sc 56025952800, https://orcid.org/0000-0002-9174-224X, d.kudin@gcras.ru

Статья поступила в редакцию 01.12.2023
Одобрена после рецензирования 08.01.2024
Принята к печати 26.01.2024

Received 01.12.2023
Approved after reviewing 08.01.2024
Accepted 26.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
январь–февраль 2024 Том 24 № 1 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
January–February 2024 Vol. 24 No 1 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

КРАТКИЕ СООБЩЕНИЯ

BRIEF PAPERS

doi: 10.17586/2226-1494-2024-24-1-165-169
УДК 621.373.826:537.312.5:681.7.068

Исследование влияния формы выходного торца оптического волокна на гидроакустические процессы в жидкости, стимулируемые микросекундными импульсами излучения Yb,Er:Glass-лазера

Раед Нассер[1], Сергей Николаевич Смирнов[2]✉

[1,2] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация

[1] raedrr9898@gmail.com, https://orcid.org/0009-0005-7557-1777
[2] sns@itmo.ru✉, https://orcid.org/0000-0003-1891-9122

Аннотация
Представлены результаты исследования гидроакустических процессов, возбуждаемых в объеме дистиллированной воды мощными микросекундными импульсами излучения Yb,Er:Glass-лазера, доставленными через оптические волокна с двумя различными формами выходного торца. Выполнено сравнение объема парогазовой полости, образовавшейся в жидкости, и перепадов давления, возникающих в момент действия лазерного импульса и в фазе «коллапс-возобновление» парогазовой полости. Полученные результаты полезны для развития технологий лазерных эндохирургических вмешательств, требующих эффективного разрушения патологических биотканей, например, лазерной экстракции катаракты.

Ключевые слова
Yb,Er:Glass-лазер, микросекундные импульсы, парогазовая полость, форма выходного торца волокна

Ссылка для цитирования: Нассер Р., Смирнов С.Н. Исследование влияния формы выходного торца оптического волокна на гидроакустические процессы в жидкости, стимулируемые микросекундными импульсами излучения Yb,Er:Glass-лазера // Научно-технический вестник информационных технологий, механики и оптики. 2024. Т. 24, № 1. С. 165–169. doi: 10.17586/2226-1494-2024-24-1-165-169

Study of the influence of the optical fiber output end shape on hydroacoustic processes in a liquid stimulated by microsecond pulses of Yb,Er:Glass laser radiation

Raed Nasser[1], Sergey N. Smirnov[2]✉

[1,2] ITMO University, Saint Petersburg, 197101, Russian Federation

[1] raedrr9898@gmail.com, https://orcid.org/0009-0005-7557-1777
[2] sns@itmo.ru✉, https://orcid.org/0000-0003-1891-9122

Abstract
The results of a study of hydroacoustic processes stimulated in a volume of distilled water by powerful microsecond pulses of Yb,Er:Glass laser radiation delivered through optical fibers with two different shapes of the output end are presented. A comparison of the volume of the steam-gas cavity formed in the liquid and the pressure drops that occur at the moment of the laser pulse action and in the “collapse-rebound” phase of the vapor-gas cavity is presented. The results obtained are useful for the development of technologies for laser endosurgical interventions that require effective destruction of pathological biological tissues, for example, laser cataract extraction.

Keywords
Yb,Er:Glass laser, microsecond pulses, steam-gas cavity, shape of the fiber output end

For citation: Nasser R., Smirnov S.N. Study of the influence of the optical fiber output end shape on hydroacoustic processes in a liquid stimulated by microsecond pulses of Yb,Er:Glass laser radiation. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2024, vol. 24, no. 1, pp. 165–169 (in Russian). doi: 10.17586/2226-1494-2024-24-1-165-169

© Нассер Р., Смирнов С.Н., 2024

Излучение Yb,Er:Glass-лазера с длиной волны 1,54 мкм является безопасным для сетчатки глаза и широко используется в дальнометрии, локации, а также в медицине. Благодаря сильному поглощению в воде и эффективной передаче через кварцевые оптические волокна, данное излучение успешно применяется в офтальмологии для проведения лазерной термокератопластики [1] и экстракции катаракты [2, 3]. В последнем случае обработка биоткани происходит в жидкой среде, а доставленные в переднюю камеру глаза через оптическое волокно лазерные импульсы стимулируют гидроакустические процессы, в том числе кавитацию, т. е. возникновение парогазовых полостей в жидкости. Эти процессы оказывают значительное влияние на эффективность разрушения катарактального хрусталика [2–4].

Параметры парогазовой полости, образующейся в результате взаимодействия лазерного излучения с водой, сильно зависят от длительности импульса, его временной структуры [5] и от объемной плотности поглощенной энергии [6], которая, в свою очередь, определяется формой выходного торца оптического волокна доставки излучения. При изменении формы выходного торца волокна изменяется распределение выходного излучения, таким образом, для различных применений разрабатываются оптические волокна со специальной оптимальной формой выходного торца [7–10].

В настоящей работе рассмотрены гидроакустические процессы, стимулируемые в объеме дистиллированной воды одиночными микросекундными импульсами излучения Yb,Er:Glass-лазера, доставленными через оптические волокна со ступенчатым профилем показателя преломления с двумя различными формами выходного торца — стандартной плоской (торец 1) и специальной (торец 2). Стандартное кварц-кварцевое волокно имело диаметр сердцевины/оболочки 470/500 мкм и числовую апертуру 0,18 (рис. 1, *a*). Торец 1 был подготовлен путем скалывания при помощи портативного скалывателя оптического волокна «SKL-6C» («Comptech Trading Co., Ltd.», Китай), обеспечивающего угол скола 90 ± 0,5°. Отличительная особенность данного метода подготовки торца — наличие дефекта в виде небольшого острого выступа, который не выходит за пределы оболочки волокна и мал по сравнению с его апертурой. Отметим, что данный дефект не оказал существенного влияния на распределение интенсивности на выходе волокна по сравнению с полированным плоским торцом без дефектов, однако имел бóльшую лучевую стойкость. Торец 2 (рис. 1, *b*) получен путем шлифовки и полировки стандартного оптического волокна одинаково с двух противоположных сторон (Wedge-Shaped Fiber) при помощи полировальной машины «ULTRAPOL Fiberlab Module System» («ULTRA TEC», США). Предварительно выполнено моделирование распространения излучения из волокна в прикладном программном пакете «TracePro 7.0» («Lambda Research Corp.», США). В результате установлено, что при угле раствора 65° обеспечивается высокая концентрация излучения в области под торцом 2 (рис. 1, *b*). Стоит отметить, что несимметричность изготовленного торца 2, заметная на рис. 1, *d*, не приводит к значительному изменению распределения интенсивности излучения на выходе волокна по отношению к идеальному симметричному торцу с

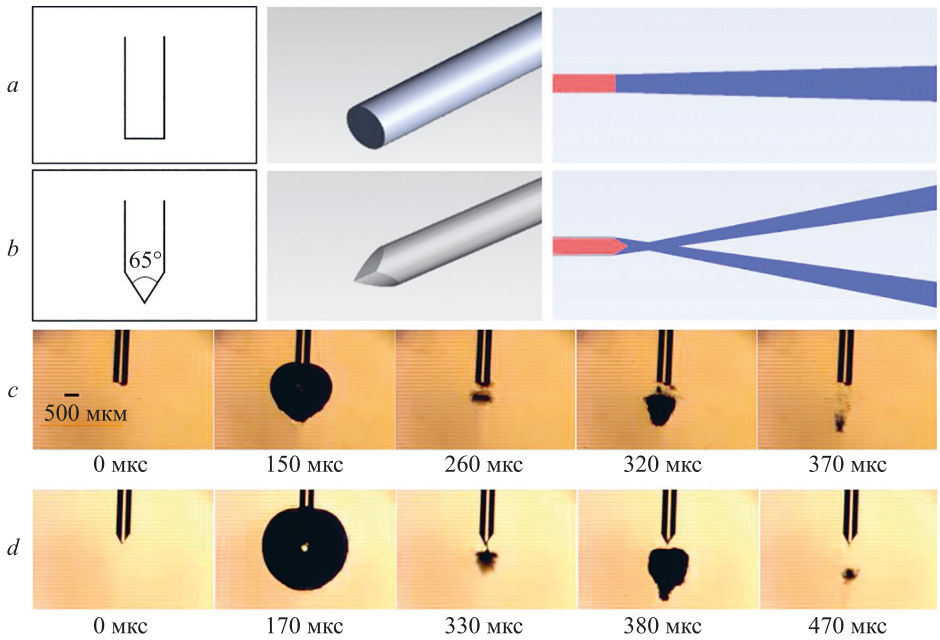


Рис. 1. Сравнения распространения выходного пучка в воде (результат моделирования) при использовании волокна с выходными торцами 1 (*a*) и 2 (*b*) для доставки излучения и формы парогазовой полости в различные моменты времени после действия лазерного импульса, доставленного через волокно с выходными торцами 1 (*c*) и 2 (*d*)

Fig. 1. Comparison of the propagation of the output beam in water (simulation result) when using a fiber with output end 1 (*a*) and end 2 (*b*) for radiation delivery. Comparison of the steam-gas cavity shape at different times after the action of a laser pulse delivered through the fiber by output end 1 (*c*) and end 2 (*d*)

углом раствора 65°, и, соответственно, существенно не изменяет объемную плотность поглощенного в воде излучения.

Цель работы — изучение и сравнение динамики парогазовой полости, а также акустического сигнала при возбуждении гидроакустических процессов с использованием для доставки излучения оптических волокон с двумя различными формами выходного торца.

Импульсы излучения доставлялись в объем дистиллированной воды на глубину 30 ± 1 мм, а расстояние от торца волокна до стенок кюветы превышало 50 мм. Динамика формы и размеров формируемой парогазовой полости регистрировалась высокоскоростной камерой «FASTCAM SA4 RV» («Photron», Япония), скорость записи составила 10^5 кадр/с. Акустический сигнал, сопровождающий взаимодействие импульсов излучения с водой, регистрировался гидрофоном «HGL-0200» («Onda», США), установленным на расстоянии 2,4 мм от выходного торца волокна под углом 45° к его нормали. Гидрофон фиксировал первичный перепад давления (ΔP_1), который возникал в момент действия лазерного импульса и был связан с термоупругим расширением жидкости, а также вторичный перепад давления (ΔP_2), связанный с процессом «коллапсом-возобновление» парогазовой полости.

На рис. 1, *c* представлены фотографии парогазовых полостей, отражающие характерные размеры парогазовой полости, возникающей после воздействия на жидкость одного лазерного импульса с энергией 85 ± 5 мДж, доставленного через волокно с торцом 1. Формирование полости происходит после действия лазерного импульса, полость достигает своего максимального размера на 150 ± 10 мкс, затем на 260 ± 10 мкс происходит ее «коллапс», после чего полость осциллирует еще два раза. В работе [10] подробно описана динамика парогазовой полости при Yb,Er:Glass-лазерном

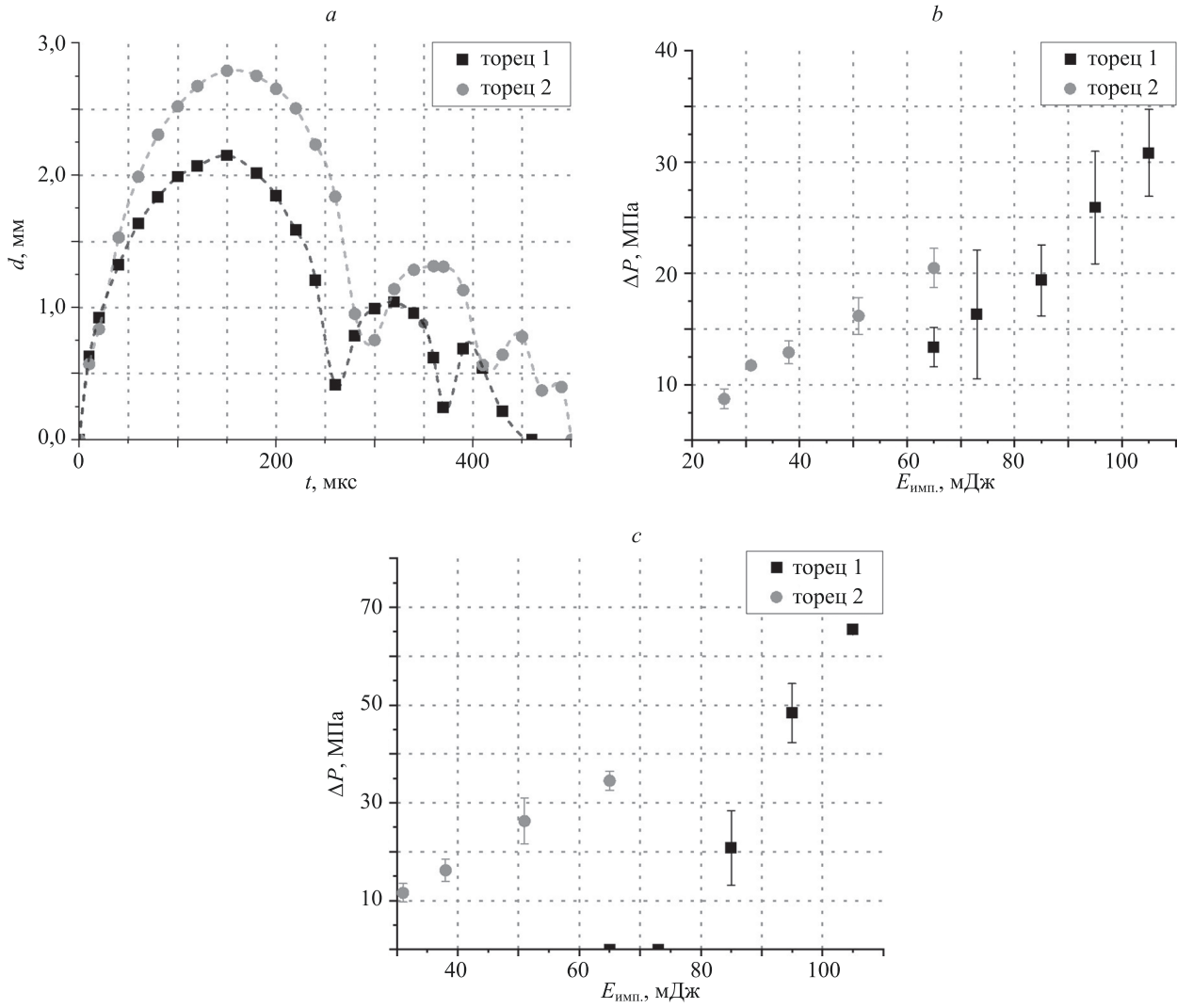


Рис. 2. Зависимости диаметра пузыря (*d*) от времени (*t*) при воздействии одиночного лазерного импульса, доставленного через оптическое волокно с разной формой выходного торца (*a*); первичного (ΔP_1) (*b*) и вторичного (ΔP_2) (*c*) перепадов давления от энергии лазерного импульса ($E_{имп.}$)

Fig. 2. Dependence of the bubble diameter (*d*) on time (*t*) when irradiating water with a single laser pulse delivered through an optical fiber with a different shape of the output end (*a*). Dependence of the primary (ΔP_1, thermoelastic expansion of water) (*b*) and secondary (ΔP_2, collapse-rebound of steam-gas cavity) (*c*) pressure drop on the laser pulse energy ($E_{имп.}$)

Таблица. Сравнение максимальных значений диаметра и объема пузыря, стимулируемого при доставке лазерного импульса через волокна с разными формами выходного торца

Table. Comparison of the maximum diameter and maximum volume of steam-gas cavity stimulated when delivering a laser pulse through fibers with different output end shapes

Форма торца	$E_{имп.}$, мДж	$d_{макс.}$, мм	$V_{макс.}$, мм^3	$\Delta P_{1\ макс.}$, МПа	$\Delta P_{2\ макс.}$, МПа
Торец 1	85 ± 2	2,150	6,96	18 ± 5	20 ± 7
Торец 2	65 ± 2	2,791	12,12	20 ± 3	34 ± 2

возбуждении. На рис. 1, *d* представлены фотографии парогазовых полостей, отражающие характерные размеры парогазовой полости, возникающей после воздействия на жидкость одного лазерного импульса с энергией 65 ± 5 мДж, доставленного через волокно с торцом 2. В этом случае формирование полости также происходит после действия лазерного импульса, полость достигает своего максимального размера на 150 ± 10 мкс, затем на 300 ± 10 мкс происходит ее «коллапс», после чего полость осциллирует еще три раза. Полученные в результате обработки кадров высокоскоростной съемки графики изменения вертикального размера парогазовой полости представлены на рис. 2, *a*.

В таблице представлено сравнение максимальных значений диаметра $d_{макс.}$ и объема пузыря $V_{макс.}$ (формируемого лазерным импульсом при его доставке через волокна с рассматриваемыми формами выходного торца оптического волокна), а также амплитуд перепадов давления $\Delta P_{1\ макс.}$ и $\Delta P_{2\ макс.}$.

Видно, что при использовании оптического волокна с выходным торцом 2, несмотря на меньшую на 20 мДж (в 1,31 раза) энергию лазерного импульса $E_{имп.}$, максимальный объем парогазовой полости $V_{макс.}$ более чем в 1,74 раза больше, значение первичного перепада давления $\Delta P_{1\ макс.}$ сопоставимо, а вторичного $\Delta P_{2\ макс.}$ — на 14 МПа (в 1,7 раза) больше. Полученные результаты могут быть полезны для развития технологий лазерных эндохирургических вмешательств, требующих эффективного разрушения патологических биотканей, например, лазерной экстракции катаракты, поскольку установлено, что применение для доставки излучения специального оптического волокна позволяет уменьшить суммарную энергетическую нагрузку на биообъект при проведении операции.

Отметим, что торец 2 имел острую грань, поэтому его механическая прочность была меньше, чем у торца 1. Вследствие этого, максимальная энергия на выходе волокна с торцом 2 была ограничена. В дальнейшем необходимо проведение исследования, направленного на повышение механической прочности выходных торцов специальных оптических волокон.

Литература

1. Куликова И.Л., Паштаев Н.П. Особенности энергетического воздействия на роговицу Glass-Yb:Er лазера с длиной волны 1,54 мкм // Вестник Оренбургского государственного университета. 2007. № S(78). С. 132–135.
2. Belikov A.V., Smirnov S.N., Kopaev S.Yu., Nemsitsveridze M.N., Batov Yu.N., Gubin A.B., Pirozhkov Yu.B. Use of bursts of 1.54-μm microsecond laser pulses for cataract destruction // Quantum Electronics. 2021. V. 51. N 1. P. 2. https://doi.org/10.1070/qel17473
3. Belikov A.V., Smirnov S.N., Batov Yu.N., Gubin A.B., Pirozhkov Yu.B., Boiko E.V., Nemsitsveridze M.N. In vitro study of cataract extraction by bursts of microsecond 1.54-μm laser pulses // Quantum Electronics. 2022. V. 52. N 1. P. 69–77. https://doi.org/10.1070/qel17959
4. Беликов А.В., Гагарский С.В., Сергеев А.Н., Смирнов С.Н., Загорулько А.М. Исследование акустического сигнала при лазерной гидроакустической обработке биоткани микросекундными импульсами излучения лазера на иттербий-эрбиевом стекле // Научно-технический вестник информационных технологий, механики и оптики. 2018. Т. 18. № 6. С. 939–945. https://doi.org/10.17586/2226-1494-2018-18-6-939-945
5. Asshauer T., Rink K., Delacrétaz G. Acoustic transient generation by holmium-laser-induced cavitation bubbles // Journal of Applied Physics. 1994. V. 76. N 9. P. 5007–5013. https://doi.org/10.1063/1.357212
6. Gregorčič P., Jezeršek M., Možina J. Optodynamic energy-conversion efficiency during an Er:YAG-laser-pulse delivery into a liquid through different fiber-tip geometries // Journal of Biomedical Optics. 2012. V. 17. N 7. P. 0750061. https://doi.org/10.1117/1.jbo.17.7.075006
7. Lippert B.M., Teymoortash A., Folz B.J., Werner J.A. Coagulation and temperature distribution in Nd: YAG interstitial laser thermotherapy: An in vitro animal study // Lasers in Medical Science. 2003. V. 18. N 1. P. 19–24. https://doi.org/10.1007/s10103-002-0246-2

References

1. Kulikova I.L., Pashtaev N.P. Peculiarities of energy influence of glass-Yb:Er laser with wave length 1,54 mkm on cornea. *Vestnik of the Orenburg State University*, 2007, no. S(78), pp. 132–135. (in Russian)
2. Belikov A.V., Smirnov S.N., Kopaev S.Yu., Nemsitsveridze M.N., Batov Yu.N., Gubin A.B., Pirozhkov Yu.B. Use of bursts of 1.54-μm microsecond laser pulses for cataract destruction. *Quantum Electronics*, 2021, vol. 51, no. 1, pp. 2. https://doi.org/10.1070/qel17473
3. Belikov A.V., Smirnov S.N., Batov Yu.N., Gubin A.B., Pirozhkov Yu.B., Boiko E.V., Nemsitsveridze M.N. In vitro study of cataract extraction by bursts of microsecond 1.54-μm laser pulses. *Quantum Electronics*, 2022, vol. 52, no. 1, pp. 69–77. https://doi.org/10.1070/qel17959
4. Belikov A.V., Gagarsky S.V., Sergeev A.N., Smirnov S.N., Zagorulko A.M. Study of acoustic signal during laser hydroacoustic processing of biological tissue by microsecond pulses of ytterbium-erbium glass laser radiation. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2018, vol. 18, no. 6, pp. 939–945. (in Russian). https://doi.org/10.17586/2226-1494-2018-18-6-939-945
5. Asshauer T., Rink K., Delacrétaz G. Acoustic transient generation by holmium-laser-induced cavitation bubbles. *Journal of Applied Physics*, 1994, vol. 76, no. 9, pp. 5007–5013. https://doi.org/10.1063/1.357212
6. Gregorčič P., Jezeršek M., Možina J. Optodynamic energy-conversion efficiency during an Er:YAG-laser-pulse delivery into a liquid through different fiber-tip geometries. *Journal of Biomedical Optics*, 2012, vol. 17, no. 7, pp. 0750061. https://doi.org/10.1117/1.jbo.17.7.075006
7. Lippert B.M., Teymoortash A., Folz B.J., Werner J.A. Coagulation and temperature distribution in Nd: YAG interstitial laser thermotherapy: An *in vitro* animal study. *Lasers in Medical Science*,

8. Russo V., Righini G., Sottini S., Trigari S. Optical fibres for medical applications: output beam shaping // Proceedings of SPIE. 1985. V. 0522. P. 166–173. https://doi.org/10.1117/12.946237
9. Clarkin J.P., Timmerman R.J., Shannon J.H. Shaped fiber tips for medical and industrial applications // Proceedings of SPIE. 2004. V. 5317. P. 70. https://doi.org/10.1117/12.540734
10. Беликов А.В., Гагарский С.В., Сергеев А.Н., Смирнов С.Н. Исследование гидродинамических процессов в жидкости при воздействии мощных микросекундных импульсов Yb,Er:Glass-лазера // Известия вузов. Приборостроение. 2017. Т. 60. № 4. С. 367–374. https://doi.org/10.17586/0021-3454-2017-60-4-367-374

2003, vol. 18, no. 1, pp. 19–24. https://doi.org/10.1007/s10103-002-0246-2
8. Russo V., Righini G., Sottini S., Trigari S. Optical fibres for medical applications: output beam shaping. *Proceedings of SPIE*, 1985, vol. 0522, pp. 166–173. https://doi.org/10.1117/12.946237
9. Clarkin J.P., Timmerman R.J., Shannon J.H. Shaped fiber tips for medical and industrial applications. *Proceedings of SPIE*, 2004, vol. 5317, pp. 70. https://doi.org/10.1117/12.540734
10. Belikov A.V., Gagarsky S.V., Sergeev A.N., Smirnov S.N. Study of hydrodynamic processes in liquids under the influence of powerful microsecond Yb,Er:glass laser pulses. *Journal of Instrument Engineering*, 2017, vol. 60, no. 4, pp. 367–374. (in Russian). https://doi.org/10.17586/0021-3454-2017-60-4-367-374

Авторы

Нассер Раед — аспирант, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0009-0005-7557-1777, raedrr9898@gmail.com
Смирнов Сергей Николаевич — кандидат технических наук, ассистент, ведущий инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, sc 7202271405, https://orcid.org/0000-0003-1891-9122, sns@itmo.ru

Authors

Raed Nasser — PhD Student, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0009-0005-7557-1777, raedrr9898@gmail.com
Sergey N. Smirnov — PhD, Assistant, Leading Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, sc 7202271405, https://orcid.org/0000-0003-1891-9122, sns@itmo.ru

Статья поступила в редакцию 30.10.2023
Одобрена после рецензирования 16.12.2023
Принята к печати 16.01.2024

Received 30.10.2023
Approved after reviewing 16.12.2023
Accepted 16.01.2024

Работа доступна по лицензии
Creative Commons
«Attribution-NonCommercial»

Уважаемые подписчики научно-технической литературы!

Журнал выходит 6 раз в год.

На журнал «Научно-технический вестник информационных технологий, механики и оптики» можно оформить подписку в почтовых отделениях по объединенному каталогу «Пресса России», подписные индексы 47197 (полугодовая подписка),

а также online
по объединенному каталогу «Пресса России» и по каталогу «Пресса по подписке»,
подписные индексы Э47197 (полугодовая подписка)
и по электронному каталогу Почты России, подписной индекс ПС543.

Корпоративная подписка и подписка физических лиц возможна
по каталогу компаний «Урал-Пресс»,
подписные индексы 47197 (полугодовая подписка) и 70522 (годовая подписка).

Сведения о подписке можно уточнить в редакции журнала по адресу:

Санкт-Петербург, ул. Ломоносова., д.9, литера А, комн. 2136.

Тел.: +7(812) 480 02 75

ТРЕБОВАНИЯ К ОФОРМЛЕНИЮ СТАТЕЙ

Журнал публикует научные обзоры, полнотекстовые статьи и краткие сообщения.

Полнотекстовая статья должна иметь четкую структуру, включающую в себя аннотацию, ключевые слова, введение, несколько содержательных разделов и заключение.

В аннотации, рассчитанной на самый широкий круг читателей, необходимо в объеме 200–400 слов в форме краткого реферата (Abstract) изложить научное содержание статьи: предмет, цель работы, метод или методология проведения работы, краткое описание эксперимента, полученные результаты, рекомендации по их применению.

Ключевые слова должны отражать информацию, представленную в статье. Основное ключевое слово указывается первым.

Во введении необходимо представить содержательную постановку рассматриваемого вопроса, провести краткий анализ известных из научной литературы решений (со ссылками на источники), дать критику их недостатков, показать научную новизну и преимущество (особенности) предлагаемого подхода.

В основном тексте статьи должна быть представлена строгая постановка решаемой задачи, изложены и обстоятельно разъяснены (доказаны) полученные утверждения и выводы, приведены результаты экспериментальных исследований или математического моделирования, иллюстрирующие сделанные утверждения. Основной текст статьи должен быть разбит на содержательные разделы (2–3 раздела).

В заключении необходимо кратко сформулировать основные результаты, прокомментировать их и, если возможно, указать направления дальнейших исследований и области применения.

Пристатейный список литературы (рекомендуется): для обзорной статьи — не менее 50, для полнотекстовой статьи — не менее 15, для краткого сообщения — не менее 8 литературных источников.

Ссылки на стандарты и иные нормативные документы, а также на неперсонифицированные интернет-ресурсы в список литературы не вносятся, а оформляются в виде сносок.

Объем обзорной статьи предварительно согласовывается с редакцией.

Объем полнотекстовой статьи, включая иллюстрации, таблицы и список литературы, не должен превышать 8 страниц машинописного текста (с рисунками и таблицами), шрифт 12 pt, один интервал.

Объем краткого сообщения — до 3 страниц, шрифт 12 pt, один интервал. Рубрикация текста не требуется. Статьи принимаются в электронном виде (ntvitmo@itmo.ru). В распечатанном виде с подписями авторов материалы представляются в редакцию по запросу редакции после прохождения процедуры рецензирования.

Комплект документов должен включать:

- текст статьи с заверстанными рисунками и таблицами;
- формы сведений об авторах (на каждого автора); допускается указание нескольких мест работы автора;
- файлы с рисунками к статье в оригинальном формате (предпочтительно JPEG) с максимальным разрешением; допускается представление цветных рисунков, если в черно-белом варианте теряется полезная информация;
- лицензионное соглашение;
- согласие на обработку персональных данных.

REQUIREMENTS FOR EXECUTION OF PAPERS

Scientific reviews, full-text and brief papers are published.

A full-text paper should have a well-defined structure, including an abstract, keywords, introduction, several substantive chapters and a conclusion.

An abstract is intended for general public. It is necessary to set forth the scientific content of the paper limited by 200–400 words in the form of a brief abstract: the subject, the work objective, the method or methodology of work, a brief description of the experiment, obtained results, recommendations for their application.

Keywords should be connected with the information presented in the paper. The main keyword is given first.

The introduction should contain a meaningful statement of the issue in question, a brief analysis of the solutions known from scientific literature (with references to sources), a criticism of their shortcomings, scientific novelty and advantages (features) of the proposed approach.

The main body of the paper should represent exact statement of the problem being solved, obtained assertions and conclusions are to be set forth and explained in detail (proved), the results of experimental studies or mathematical modeling should be given to illustrate the statements made. The main text of the paper needs to be split up into meaningful sections (2–3 sections).

In conclusion it is necessary to summarize the main results, comment on them and, if possible, to indicate the areas of future research and application.

A list of references is recommended: for a review paper — at least 50 items, for full-text paper — not less than 15, for a brief paper — at least 8 references. References to standards and other regulatory documents, as well as to Internet resources, are not included in a reference list and are documented as footnotes.

The length of a review paper is coordinated with the editorial board beforehand.

The length of a full-text paper, including illustrations, tables and references, should not exceed 8 pages of typewritten text (figures and tables included), font 12 pt, single-spaced.

The length of a brief paper is up to 3 pages, font 12 pt, single-spaced.

No rubrication is required.

Papers are accepted in electronic form (ntvitmo@itmo.ru).

Materials are submitted to the editors on request after the procedure of reviewing in a printed form signed by the authors.

A set of documents should include:

- The text of the paper with make-up figures and tables;
- The forms with information about the authors (for each author), several employers can be specified by the author;
- Files with pictures to the paper in their original format (preferably JPEG) with the maximum resolution; should useful information is lost in the black and white variant, color drawings are permitted;
- Publishers license agreement;
- Consent to the processing of personal data.

SCIENTIFIC AND TECHNICAL JOURNAL OF ITMO UNIVERSITY

INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS

2024, VOLUME 24, NUMBER 1 (JANUARY–FEBRUARY) ISSN 2226-1494 (PRINT), 2500-0373 (ONLINE)

OPTICAL ENGINEERING

Structural and spectral properties of YAG:Nd, YAG:Ce and YAG:Yb nanocrystalline powders synthesized via modified Pechini method — *Moussaoui A., Bulyga D.V., Ignatiev A.I., Evstropiev S.K., Nikonorov N.V.* — 1

Computational prediction in the problem of stereo image identification — *Samoilenko M.V., Hachikian V.A.* — 11

Comparison of application results of two speckle methods for study multi-cycle fatigue of structural steel — *Vladimirov A.P., Kamantsev I.S., Drukarenko N.A., Myznov K.E., Naumov K.V.* — 20

Laser-induced thermal effect on the electrical characteristics of photosensitive PbSe films — *Olkhova A.A., Patrikeeva A.A., Butyaeva M.A., Pushkareva A.E., Avilova E.A., Moskvin M.K., Sergeev M.M., Veiko V.P.* — 30

COMPUTER SCIENCE

Homograph recognition algorithm based on Euclidean metric — *Izrailova E.S., Astemirov A.V., Badaeva A.S., Sultanov Z.A., Umarkhadzhiev S.M., Khekhaev M.-S.L., Yasaeva M.L.* — 41

An improved performance of RetinaNet model for hand-gun detection in custom dataset and real time surveillance video (in English) — *Khin P.P., Htaik N.M.* — 51

Solving the problem of preliminary partitioning of heterogeneous data into classes in conditions of limited volume — *Sharamet A.V.* — 62

Correction of single error bursts beyond the code correction capability using information sets — *Isaeva M.N., Ovchinnikov A.A.* — 70

A novel strategic trajectory-based protocol for enhancing efficiency in wireless sensor networks (in English) — *Gopalakrishnan R., Sentil Kumar A.* — 81

Automation of complex text CAPTCHA recognition using conditional generative adversarial networks — *Zadorozhnyy A.S., Korepanova A.A., Abramov M.V., Sabrekov A.A.* — 90

Deep attention based Proto-oncogene prediction and Oncogene transition possibility detection using moments and position based amino acid features (in English) — *Vijayalakshmi M., Vallinayagi M.* — 101

A method of storing vector data in compressed form using clustering — *Tomilov N.A., Turov V.P., Babayants A.A., Platonov A.V.* — 112

ARTIFICIAL INTELLIGENCE AND COGNITIVE INFORMATION TECHNOLOGIES

Monocular depth estimation for 2D mapping of simulated environments (in English) — *Barhoum M., Pyrkin A.A.* — 118

Segmentation of muscle tissue in computed tomography images at the level of the L3 vertebra — *Teplyakova A.R., Shershnev R.V., Starkov S.O., Agababian T.A., Kukarskaya V.A.* — 124

MODELING AND SIMULATION

Providing operating modes for Coriolis vibration gyroscopes with low-Q resonators — *Matveev V.V., Likhosherst V.V., Kalikanov A.V., Pogorelov M.G., Kirsanov M.D., Telukhin S.V.* — 133

Collection and processing of environmental information in oil and gas production areas and solving other applied problems using active search methods (Review article) — *Svitnev I.V., Naydanov A.F., Vilkov A.V., Sokolov D.A., Lebedev M.Yu., Kharitonova E.A., Lukyanova L.A.* — 144

Using machine learning technologies to solve the problem of classifying infrasound background monitoring signals — *Frolov I.N., Kudryavtsev N.G., Safonova V.Yu., Kudin D.V.* — 156

BRIEF PAPERS

Study of the influence of the optical fiber output end shape on hydroacoustic processes in a liquid stimulated by microsecond pulses of Yb,Er:Glass laser radiation — *Nasser R., Smirnov S.N.* — 165