УНИВЕРСИТЕТ ИТМО

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ
март–апрель 2022 Том 22 № 2 http://ntv.ifmo.ru/
SCIENTIFIC AND TECHNICAL JOURNAL OF INFORMATION TECHNOLOGIES, MECHANICS AND OPTICS
March–April 2022 Vol. 22 No 2 http://ntv.ifmo.ru/en/
ISSN 2226-1494 (print) ISSN 2500-0373 (online)

НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК УНИВЕРСИТЕТ ИТМО
ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ, МЕХАНИКИ И ОПТИКИ

КРАТКИЕ СООБЩЕНИЯ

BRIEF PAPERS

doi: 10.17586/2226-1494-2022-22-2-410-414
УДК 004.5; 004.93

Метод детектирования пространственного положения рук по данным глубинных камер для малопроизводительных вычислительных устройств

Дмитрий Сергеевич Медведев[1]✉, Андрей Дмитриевич Игнатов[2]

[1] Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация
[2] Федеральный исследовательский центр «Информатика и управление» РАН, Москва, 119333, Российская Федерация

[1] dsmedvedev@itmo.ru✉, https://orcid.org/0000-0001-5596-3988
[2] rayignatov@outlook.com, https://orcid.org/0000-0003-4343-3407

Аннотация
Предложен метод определения направления рук в пространстве с применением малопроизводительных устройств интернета вещей. Метод основан на применении алгоритмов, предназначенных для решения задач оценки позы человека (данный класс задач известен как Human Pose Estimation, HPE). На базе алгоритмов построена модель обнаружения направлений рук. Выполнено тестирование и сравнение известных алгоритмов PoseNet и OpenPose, положенных в основу разработанного метода, по среднему углу ошибки. В качестве аппаратной платформы применен одноплатный компьютер Raspberry Pi 4B и глубинный сенсор Intel RealSense D435i. Разработанный метод может быть применен в системах управления жестами для системы «умный дом».

Ключевые слова
алгоритм оценки позы человека, human pose estimation, человеко-машинное взаимодействие, карты глубин, интернет вещей, управление жестами

Ссылка для цитирования: Медведев Д.С., Игнатов А.Д. Метод детектирования пространственного положения рук по данным глубинных камер для малопроизводительных вычислительных устройств // Научно-технический вестник информационных технологий, механики и оптики. 2022. Т. 22, № 2. С. 410–414. doi: 10.17586/2226-1494-2022-22-2-410-414

Method for discovering spatial arm positions with depth sensor data at low-performance devices

Dmitrii S. Medvedev[1]✉, Andrei D. Ignatov[2]

[1] ITMO University, Saint Petersburg, 197101, Russian Federation
[2] Federal Reserch Center “Computer Science and Control” of RAS, Moscow, 119333, Russian Federation

[1] dsmedvedev@itmo.ru✉, https://orcid.org/0000-0001-5596-3988
[2] rayignatov@outlook.com, https://orcid.org/0000-0003-4343-3407

Abstract
A method of arm aiming direction estimation for low performance Internet of Things devices is proposed. It uses Human Pose Estimation (HPE) algorithms for retrieving human skeleton key points. Having these key points, arm aiming directions model is calculated. Two well-known HPE methods (PoseNet and OpenPose) are examined. These algorithms have been tested and compared by the average angle of error. The system includes a Raspberry Pi 4B single-board computer and an Intel RealSense D435i depth sensor. The developed approach may be utilized in “smart home” gesture control systems.

Keywords
human pose estimation, human computer interaction, human machine interaction, depth maps, Internet of Things, gesture control

For citation: Medvedev D.S., Ignatov A.D. Method for discovering spatial arm positions with depth sensor data at low-performance devices. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2022, vol. 22, no. 2, pp. 410–414 (in Russian). doi: 10.17586/2226-1494-2022-22-2-410-414

В настоящее время решения для человеко-машинного взаимодействия находятся в стадии активного развития. Проблема улучшения пользовательского опыта при взаимодействии человека со сложными системами подталкивает исследователей к разработке все большего количества новых и необычных решений в данной области. К традиционным физическим интерфейсам (например, кнопкам или сенсорным экранам) добавляются совершенно новые, не требующие от пользователя прямого управляющего воздействия. К таким относятся, например, голосовые интерфейсы [1] и системы управления жестами.

В данной работе управление жестами рассматривается как перспективный и более удобный по сравнению со всеми остальными способ человеко-машинного взаимодействия в окружении интернета вещей. В зависимости от возможностей восприятия системы, пользователь может совершать действия с любого места в пространстве, ему не нужна единая физическая точка взаимодействия для управления «умным домом».

Предполагаемый сценарий использования системы управления жестами: пользователь, авторизованный системой, находится в фокусе (отслеживается визуальными сенсорами либо носимыми устройствами); указывая жестами рук на различные актуаторы, он способен управлять ими без использования голоса или сторонних предметов, таких как смартфон или физическая кнопка.

Конечная цель заключается в имплементации системы, способной полностью поддержать данный сценарий взаимодействия. Для достижения данной цели выбраны хорошо известные алгоритмы решения задач оценки позы человека (Human Pose Estimation, HPE).

Подходы к имплементации алгоритмов решения задач HPE могут быть разделены на два больших класса: методы, основанные на данных с носимых устройств («умных» браслетов, часов и т. п.) [2] и методы, основанные на данных с видеокамер [3–9].

Новизна разработанного метода детектирования направления рук заключается в использовании алгоритмов, базирующихся на глубинных и RGB-сенсорах. Предполагается, что такой подход обладает рядом преимуществ по сравнению с методами, основанными на применении носимых устройств:

— не требуется постоянного использования посторонних предметов, которое сопряжено с дополнительными неудобствами для пользователя: носить устройство на себе, поддерживать его в заряженном состоянии и т. п.
— глубинные и RGB-сенсоры могут быть использованы в той же экосистеме для решения иных, не связанных с управлением жестами задач, например, для распознавания лиц.

Визуальные алгоритмы решения задач HPE в большинстве своем основаны на классификаторах Random Forest [3–5] или сверточных нейронных сетях [6–9]. Ввиду более высокой производительности и точности последние получили широкое распространение в науке и индустрии [8, 9].

Разработанный в настоящей работе метод детектирования пространственного положения рук по данным глубинных камер для малопроизводительных вычислительных устройств отличается от известных алгоритмов. Данное отличие состоит в применении данных глубинных сенсоров совместно с RGB-кадрами. Заметим, что другие алгоритмы используют данные обычных видеокамер [3–9] или носимых устройств [2]. Метод базируется на реализации двух алгоритмов: OpenPose [8] и PoseNet [9].

Разработана методика тестирования для оценки качества предложенного метода.

Методика предполагает измерение и расчет метрик качества (отклонение от эталонного значения планарного угла) и производительности (Frames Per Second (FPS) — количество кадров в секунду). Для этого требуется наличие двух непрерывных и синхронных потоков кадров: изображения RGB и матрицы глубин. Методика обработки каждой пары кадров включает следующие шаги.

Шаг 1. Получение кадра RGB и матрицы глубины.

Шаг 2. Выравнивание кадров друг относительно друга — практическая имплементация данного шага зависит от конструктивных особенностей устройства, а именно от взаимного расположения глубинного сенсора и объектива RGB-камеры.

Шаг 3. Получение двумерных координат ключевых точек с применением метода HPE (OpenPose или PoseNet).

Шаг 4. Если фигура человека не выявлена, выполняется возврат к шагу 1.

Шаг 5. Получение трехмерных координат из матрицы глубин — по координатам, полученным на шаге 3, из матрицы глубин извлекается соответствующая третья компонента координат.

Шаг 6. Нормализация — трехмерные координаты приводятся к базису с началом отсчета в груди оператора.

Шаг 7. Нахождение направления рук — рассчитывается как планарные углы в двух плоскостях — *OXY* и *OXZ* (ось *Z* направлена прямо от камеры).

Получение метрики качества подразумевает проведение серии экспериментов. В каждом эксперименте генерируются 8 наборов жестов, включающих в себя различные направления рук. Для обеих рук это направления: вверх, вниз, вперед, вперед-вверх (под углом 45°), вперед-вниз (под углом 45°). Для правой руки добавляются направления: вправо, вправо-вверх (под углом 45°), вправо-вниз (под углом 45°). Для левой руки аналогично: влево, влево-вверх (под углом 45°), влево-вниз (под углом 45°). Каждый жест представлен в виде двух планарных углов, из которых могут быть получены эталонные направления рук.

Перед началом эксперимента случайным образом генерируются пары из 8 жестов для левой и правой рук. Испытания включает следующие этапы:

1) для очередной пары предопределенных жестов приложение предлагает оператору, стоящему перед камерой, повторить жест;
2) по методике, описанной ранее, определяется направление рук, показанное оператором;
3) производится расчет пространственных векторов по координатам эталонной и экспериментально полученных поз;
4) выполняется расчет угла между полученными векторами;
5) полученный угол сохраняется вместе с векторами;
6) если еще остались необработанные пары жестов, производится возврат к этапу 1;
7) по всем сохраненным углам рассчитываются средние значения разницы углов — чем ближе к 0 полученный результат, тем лучше работает метод.

Производительность FPS рассчитывается при помощи системных часов на текущей аппаратной платформе. На этапе 2 определяется, сколько кадров за секунду удалось обработать. Эта информация сохраняется вместе с другими результатами на каждом этапе. В конце эксперимента выполняется расчет среднего арифметического значения по всем показателям производительности и выводится итоговое значение. Чем оно выше, тем лучше производительность.

Для выполнения экспериментов предложено тестовое приложение, реализующее описанный метод и алгоритм тестирования. В качестве аппаратного обеспечения для проведения эксперимента выбран одноплатный компьютер Raspberry Pi 4B, который имеет следующие преимущества: ориентированность на использование в среде интернета вещей вследствие наличия большого количества периферии и плат расширения; сбалансированность по цене и производительности; потенциально эффективный GPGPU с поддержкой Vulkan API, который может быть использован для ускорения вычислений.

К одноплатному компьютеру подключена глубинная камера Intel RealSense D435i для получения RGB-изображений и соответствующих им карт глубин.

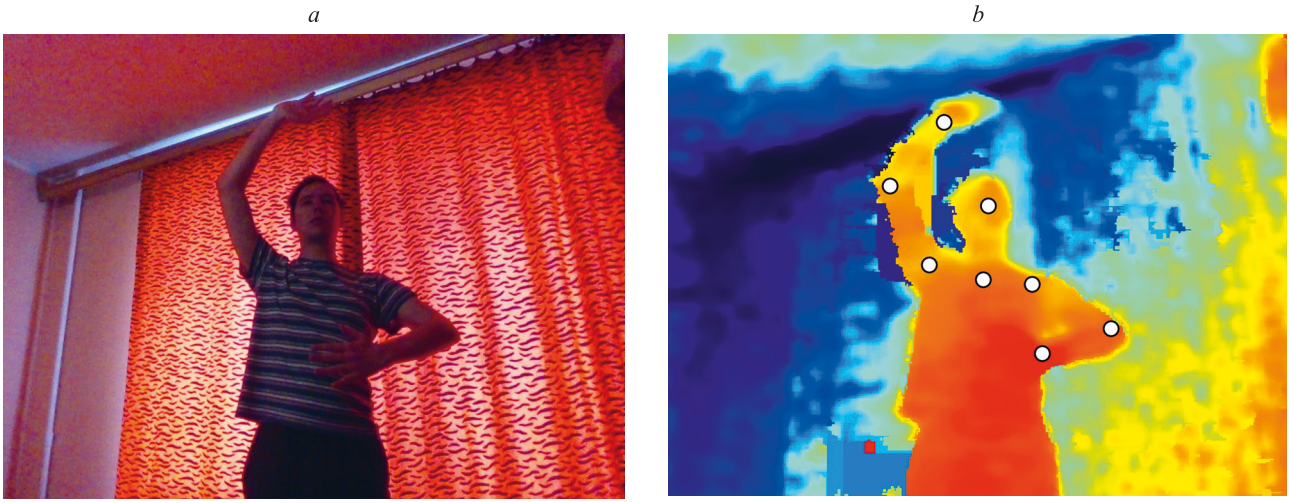

Рис. 1. Исходные данные разработанного метода: RGB-изображение (*a*), карта глубин с отмеченными ключевыми точками (*b*)

Fig. 1. Input data of the developed method: RGB image (*a*), depth heat map with marked key points (*b*)

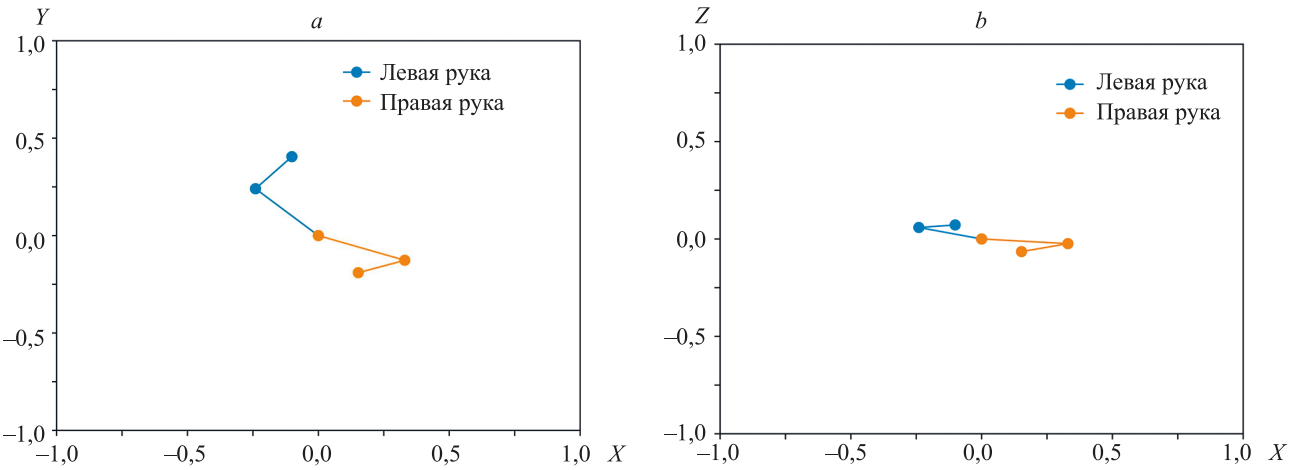

Рис. 2. Результаты работы предложенного метода — проекции расчетного расположения рук в пространстве: плоскостей *OXY* (*a*) и *OXZ* (*b*). Значения координат нормированы к величине половины диагонали кадра (в пикселах)

Fig. 2. Output results of the developed method — projections of estimated spatial arm directions: *OXY* plane (*a*) and *OXZ* plane (*b*)

На тестовом стенде выполнена серия экспериментов согласно предложенной методике.

Примеры входных данных, полученных с помощью глубинной камеры для алгоритмов OpenPose и PoseNet, показаны на рис. 1.

Пример выходных данных разработанного метода показан на рис. 2. Начало координат в обоих случаях находится в ключевой точке, соответствующей груди оператора. Оси ориентированы естественным образом (ось X на обеих проекциях направлена вправо, оси Y и Z — вверх).

Результаты экспериментов представлены в таблице.

В результате выполненной работы продемонстрирована возможность детектирования направления рук человека в реальном времени с использованием одноплатного компьютера в целях решения задач управления окружением интернета вещей. Предложены метрики точности и производительности.

Выполнено тестирование метода с использованием алгоритмов OpenPose и PoseNet. Тестирование произведено на одноплатном компьютере Raspberry Pi 4B с использованием глубинной камеры Intel RealSense D435i.

Таблица. Сравнение алгоритмов решения задач HPE по предложенным метрикам качества и производительности

Table. HPE algorithms comparison with proposed metrics

Метрики	Алгоритмы	
	OpenPose	PoseNet
Средний угол ошибки, рад	0,29	1,30
Средняя производительность, кадр/с	0,18	2,12

Результат тестирования с использованием алгоритма на базе OpenPose показал приемлемую точность, однако скорость обработки оказалась недостаточной для режима реального времени. В то же время алгоритм PoseNet продемонстрировал более высокую производительность, однако качественные характеристики оказались ниже ожидаемого значения.

Литература

1. Шматков В.Н., Бонковски П., Медведев Д.С., Корзухин С.В., Голендухин Д.В., Спыну С.Ф., Муромцев Д.И. Взаимодействие с устройствами интернета вещей с использованием голосового интерфейса // Научно-технический вестник информационных технологий, механики и оптики. 2019. Т. 19. № 4. С. 714–721. https://doi.org/10.17586/2226-1494-2019-19-4-714-721
2. Chen W., Yu C., Tu C., Lyu Z., Tang J., Ou S., Fu Y., Xue Z. A survey on hand pose estimation with wearable sensors and computer-vision-based methods // Sensors. 2020. V. 20. N 4. P. 1074. https://doi.org/10.3390/s20041074
3. Shotton J., Fitzgibbon A., Cook M., Sharp T., Finocchio M., Moore R., Kipman A., Blake A. Real-time human pose recognition in parts from single depth images // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2011. P. 1297–1304. https://doi.org/10.1109/CVPR.2011.5995316
4. Hernández-Vela A., Zlateva N., Marinov A., Reyes M., Radeva P., Dimov D., Escalera S. Graph cuts optimization for multi-limb human segmentation in depth maps // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. 2012. P. 726–732. https://doi.org/10.1109/CVPR.2012.6247742
5. Ye M., Wang X., Yang R., Ren L., Pollefeys M. Accurate 3D pose estimation from a single depth image // Proc. of the IEEE International Conference on Computer Vision. 2011. P. 731–738. https://doi.org/10.1109/ICCV.2011.6126310
6. Shafaei A., Little J.J. Real-time human motion capture with multiple depth cameras // Proc. of the 13th Conference on Computer and Robot Vision (CRV). 2016. P. 24–31. https://doi.org/10.1109/CRV.2016.25
7. Marin-Jimenez M.J., Romero-Ramirez F.J., Muñoz-Salinas R., Medina-Carnicer R. 3D human pose estimation from depth maps using a deep combination of poses // Journal of Visual Communication and Image Representation. 2018. V. 55. P. 627–639. https://doi.org/10.1016/j.jvcir.2018.07.010
8. Cao Z., Hidalgo G., Simon T., Wei S.-E., Sheikh Y. OpenPose: Realtime Multi-person 2D pose estimation using part affinity fields // IEEE Transactions on Pattern Analysis and Machine Intelligence. 2021. V. 43. N 1. P. 8765346. https://doi.org/10.1109/TPAMI.2019.2929257
9. Kendall A., Grimes M., Cipolla R. PoseNet: A convolutional network for real-time 6-dof camera relocalization // Proc. of the 15th IEEE International Conference on Computer Vision (ICCV). 2015. P. 2938–2946. https://doi.org/10.1109/ICCV.2015.336

References

1. Shmatkov V.N., Bąkowski P., Medvedev D.S., Korzukhin S.V., Golendukhin D.V., Spynu S.F., Mouromtsev D.I. Interaction with Internet of Things devices by voice control. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2019, vol. 19, no. 4, pp. 714–721. (in Russian). https://doi.org/10.17586/2226-1494-2019-19-4-714-721
2. Chen W., Yu C., Tu C., Lyu Z., Tang J., Ou S., Fu Y., Xue Z. A survey on hand pose estimation with wearable sensors and computer-vision-based methods. *Sensors*, 2020, vol. 20, no. 4, pp. 1074. https://doi.org/10.3390/s20041074
3. Shotton J., Fitzgibbon A., Cook M., Sharp T., Finocchio M., Moore R., Kipman A., Blake A. Real-time human pose recognition in parts from single depth images. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1297–1304. https://doi.org/10.1109/CVPR.2011.5995316
4. Hernández-Vela A., Zlateva N., Marinov A., Reyes M., Radeva P., Dimov D., Escalera S. Graph cuts optimization for multi-limb human segmentation in depth maps. *Proc. of the IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 726–732. https://doi.org/10.1109/CVPR.2012.6247742
5. Ye M., Wang X., Yang R., Ren L., Pollefeys M. Accurate 3D pose estimation from a single depth image. *Proc. of the IEEE International Conference on Computer Vision*, 2011, pp. 731–738. https://doi.org/10.1109/ICCV.2011.6126310
6. Shafaei A., Little J.J. Real-time human motion capture with multiple depth cameras. *Proc. of the 13th Conference on Computer and Robot Vision (CRV)*, 2016, pp. 24–31. https://doi.org/10.1109/CRV.2016.25
7. Marin-Jimenez M.J., Romero-Ramirez F.J., Muñoz-Salinas R., Medina-Carnicer R. 3D human pose estimation from depth maps using a deep combination of poses. *Journal of Visual Communication and Image Representation*, 2018, vol. 55, pp. 627–639. https://doi.org/10.1016/j.jvcir.2018.07.010
8. Cao Z., Hidalgo G., Simon T., Wei S.-E., Sheikh Y. OpenPose: Realtime Multi-person 2D pose estimation using part affinity fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021, vol. 43, no. 1, pp. 8765346. https://doi.org/10.1109/TPAMI.2019.2929257
9. Kendall A., Grimes M., Cipolla R. PoseNet: A convolutional network for real-time 6-dof camera relocalization. *Proc. of the 15th IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 2938–2946. https://doi.org/10.1109/ICCV.2015.336

Авторы

Медведев Дмитрий Сергеевич — инженер, Университет ИТМО, Санкт-Петербург, 197101, Российская Федерация, https://orcid.org/0000-0001-5596-3988, dsmedvedev@itmo.ru
Игнатов Андрей Дмитриевич — инженер-исследователь, Федеральный исследовательский центр «Информатика и управление» РАН, Москва, 119333, Российская Федерация, SC 57205407925, https://orcid.org/0000-0003-4343-3407, rayignatov@outlook.com

Authors

Dmitrii S. Medvedev — Engineer, ITMO University, Saint Petersburg, 197101, Russian Federation, https://orcid.org/0000-0001-5596-3988, dsmedvedev@itmo.ru
Andrei D. Ignatov — Research Engineer, Federal Reserch Center "Computer Science and Control" of RAS, Moscow, 119333, Russian Federation, SC 57205407925, https://orcid.org/0000-0003-4343-3407, rayignatov@outlook.com

Статья поступила в редакцию 14.01.2022
Одобрена после рецензирования 11.02.2022
Принята к печати 20.03.2022

Received 14.01.2022
Approved after reviewing 11.02.2022
Accepted 20.03.2022