

УДК 004.855.5

doi: 10.17586/2226-1494-2020-20-4-525-531

МЕТОД АВТОМАТИЧЕСКОГО ВИДЕОАНАЛИЗА ДВИЖЕНИЙ РУК И РАСПОЗНАВАНИЯ ЖЕСТОВ В ЧЕЛОВЕКО-МАШИННЫХ ИНТЕРФЕЙСАХ

Д. Рюмин

Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация
Адрес для переписки: dl_03.03.1991@mail.ru

Информация о статье

Поступила в редакцию 20.05.20, принята к печати 25.06.20

Язык статьи — русский

Ссылка для цитирования: Рюмин Д. Метод автоматического видеоанализа движений рук и распознавания жестов в человеко-машинных интерфейсах // Научно-технический вестник информационных технологий, механики и оптики. 2020. Т. 20. № 4. С. 525–531. doi: 10.17586/2226-1494-2020-20-4-525-531

Аннотация

Предмет исследования. Предложено одно из решений автоматического анализа и распознавания жестов рук человека. Распознавание элементов жестовых языков является актуальной задачей в современном информационном мире. Проблема эффективности распознавания жестов не решена из-за наличия культурных различий жестовых языков мира, различий в условиях показа жестов. Проблема усложняется небольшим размером пальцев рук. **Метод.** Представленный метод основан на анализе последовательностей кадров видеопотока, получаемого с помощью оптической камеры. Для обработки полученных видеопоследовательностей предложено использовать карту глубины и комбинации современных классификаторов на основе архитектур глубоких нейронных сетей Single Shot MultiBox Detector с уменьшенной моделью сети ResNet-10, NASNetMobile и LSTM. **Основные результаты.** Эксперименты по автоматическому видеоанализу движений рук и распознаванию жестов в режиме реального времени показали большие возможности предложенного метода для задач человеко-машинного взаимодействия. Точность распознавания 48 однорукных жестов на базе данных TheRuSLan составила 79 %. Это является лучшим результатом по сравнению с другими подходами к решению данной проблемы. **Практическая значимость.** Результаты работы могут использоваться в автоматических системах распознавания жестовых языков, а также в тех ситуациях, когда необходимо бесконтактное взаимодействие различных групп пользователей, включая людей с ограниченными возможностями по слуху и зрению, с мобильными информационными роботами посредством автоматического распознавания жестовой информации.

Ключевые слова

видеоанализ движений рук, карта глубины, распознавание жестов, детектирование лиц, глубокие нейронные сети

doi: 10.17586/2226-1494-2020-20-4-525-531

AUTOMATED HAND DETECTION METHOD FOR TASKS OF GESTURE RECOGNITION IN HUMAN-MACHINE INTERFACES

D. Ryumin

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation
Corresponding author: dl_03.03.1991@mail.ru

Article info

Received 20.05.20, accepted 25.06.20

Article in Russian

For citation: Ryumin D. Automated hand detection method for tasks of gesture recognition in human-machine interfaces. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, 2020, vol. 20, no. 4, pp. 525–531 (in Russian). doi: 10.17586/2226-1494-2020-20-4-525-531

Abstract

Subject of Research. The paper presents a solution for automatic analysis and recognition of human hand gestures. Recognition of the elements of sign languages is a topical task in the modern information world. The problem of gesture recognition efficiency has not been resolved due to the presence of cultural diversities in the world sign languages,

the differences in the conditions for showing gestures. The problem becomes more complicated by the small size of fingers. **Method.** The presented method is based on the analysis of frame sequences of a video stream obtained using an optical camera. For processing of the obtained video sequences, it is proposed to use a depth map and a combination of modern classifiers based on Single Shot MultiBox Detector deep neural network architectures with a reduced network model of ResNet-10, NASNetMobile and LSTM type. **Main Results.** Experiments on automatic video analysis of hand movements and gesture recognition in real time show great potential of the proposed method for human-machine interaction tasks. The recognition accuracy of 48 one-handed gestures based on TheRuSLan database is 79 %. This result is better as compared to the other approaches to solving this problem. **Practical Relevance.** The results can be used in automatic systems for recognition of sign languages, as well as in the situations where contactless interaction of various user groups is necessary, for example, people with hearing and vision impairments, mobile information robots through automatic recognition of sign information.

Keywords

hand movement video analysis, depth map, gesture recognition, face detection, deep neural networks

Введение

Прогресс цифровых технологий в последние годы достиг такого уровня, что различные группы пользователей, включая людей с ограниченными возможностями по слуху и зрению могут бесконтактно взаимодействовать с мобильными информационными роботами [1–3] посредством многомодальных человеко-машинных интерфейсов [4–6]. Это обусловлено тем, что процесс взаимодействия человека с автоматизированными системами включает в себя множество областей исследований [7–12]. Такие системы, как правило, позволяют автоматически распознавать и синтезировать речевую [13] и жестовую [14] информации. Так, с помощью жестов можно взаимодействовать с роботом на некотором расстоянии через простые команды, которые будут нести однозначный смысл и эффективны в шумных условиях, когда речь малоэффективна [15, 16]. Кроме того, жесты являются наиболее основными и выразительными формами человеческого общения. Слышащие люди используют жесты как вспомогательные средства при межличностной речевой коммуникации, а для людей, лишенных слуха, они являются основополагающим средством общения. Согласно Всемирной организации здравоохранения, на 2019 год по статистике официально в мире порядка 466 млн человек страдают глухотой или испытывают проблемы со слухом, из которых 432 млн взрослых людей и 34 млн детей. Кроме того, каждый третий человек в возрасте старше 65 лет страдает от потери слуха и, согласно оценкам, к 2050 году более 900 млн человек будут страдать глухотой или испытывать проблемы со слухом¹. Также, по статистике Министерства здравоохранения Российской Федерации на 2018 год в стране насчитывалось около 13 млн людей с нарушением слуха, из которых более 1 млн детей в возрасте до 18 лет².

Таким образом, цель данной работы заключалась в получении автоматической технологии видеоанализа

движений и распознавании жестов в режиме реального времени с помощью оптической камеры и карты глубины для задач человеко-машинного взаимодействия.

Метод видеоанализа движений рук и распознавания жестов

Автоматическое распознавание жестов направлено на машинный анализ и интерпретацию (понимание) значимых движений артикуляторов (рук, головы, туловища) и мимики лица человека. До сих пор проблема эффективности распознавания жестов не решена из-за различий жестовых языков мира, различных условий показов жестов, относительно небольших размеров пальцев рук.

В большинстве случаев распознавание жестов сводится к обработке видеопоследовательности, которая предоставляет информацию о какой-либо части человеческого тела, например, кисти руки или обеих рук во времени и в пространстве [17]. Исключением являются статические жесты, при показе которых не возникает необходимости в непрерывном движении рук, так как положение кисти и пальцев рук неизменно в пространстве по истечении времени [4]. Также присутствие на кадрах видеопотока сложных сцен приводит к довольно серьезным проблемам распознавания ввиду сложности извлечения пространственных особенностей жестов относительно небольших размеров рук и кистей человека по сравнению со всей сценой. Следовательно, процесс распознавания жестов разумно строить на пространственно-временной оценке движения рук человека.

Функциональная схема метода видеоанализа движений рук для распознавания жестов жестового языка представлена на рис. 1. Входными видеоданными являются цветной (RGB) видеопоток и карта глубины, получаемые от сенсора Kinect v2, на котором демонстратор воспроизводит жесты русского жестового языка (РЖЯ). В случае недоступности сенсора Kinect v2 метод автоматически прерывается, иначе осуществляется циклическая, синхронная обработка кадров, на каждой итерации которой происходит проверка на получение определенного кадра. На данном этапе остановка возможна в следующих случаях: возникновение ошибки при получении как RGB-видеокадров, так и карты глубины; завершение одного из описанных видеопотоков.

¹ Глухота и потеря слуха [Электронный ресурс]. <https://www.who.int/ru/news-room/fact-sheets/detail/deafness-and-hearing-loss>, свободный. Яз. рус. (дата обращения: 13.05.2020).

² Статистический сборник Министерства здравоохранения Российской Федерации за 2018 год [Электронный ресурс]. <https://www.rosminzdrav.ru/ministry/61/22/stranitsa-979/statisticheskie-i-informatsionnye-materialy/statisticheskiy-sbornik-2018-god>, свободный. Яз. рус. (дата обращения: 13.05.2020).

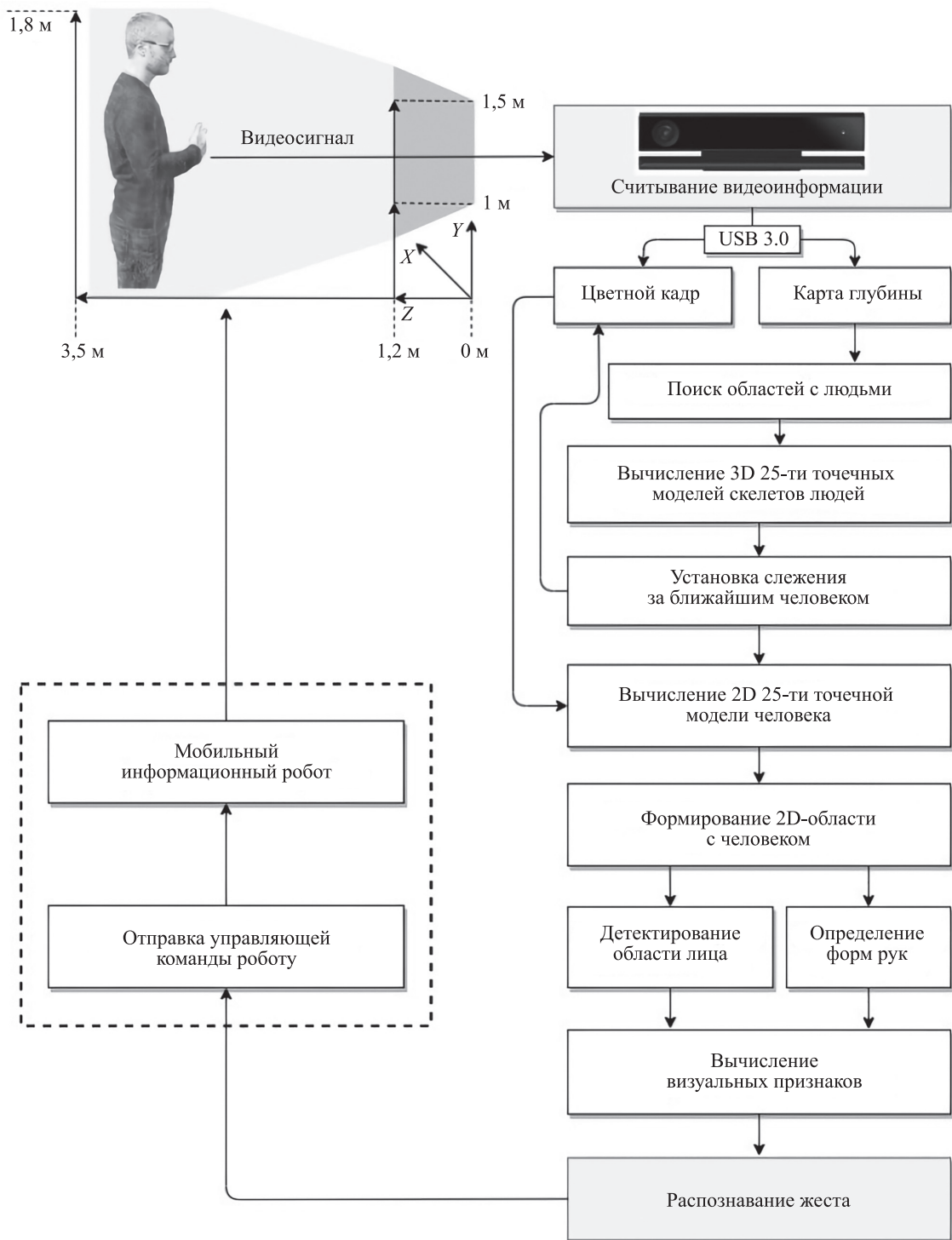


Рис. 1. Функциональная схема метода видеоанализа движений рук и распознавания жестов жестового языка

На каждом 3D-кадре карты глубины с помощью набора средств разработки¹, который поставляется вместе с сенсором Kinect v2, происходит поиск людей (максимум до 6 человек) на расстоянии от 1,2 до 3,5 м и вычисление 3D 25-ти точечных моделей всех найденных скелетов людей. Затем по оси Z трехмер-

ного пространства определяется ближайшая скелетная модель и устанавливается слежение за ней. Далее 3D-координаты преобразуются в 2D при помощи все того же SDK Kinect v2 (рис. 2, а).

Следующий этап позволяет формировать прямоугольные области с ближайшим человеком на 2D-кадре цветного видеопотока на основе 2D 25-ти точечной модели скелета, как можно заметить на рис. 2, б.

На следующем шаге в пределах сформированной прямоугольной области с человеком происходит

¹ Kinect for Windows SDK 2.0 [Электронный ресурс]. <https://developer.microsoft.com/en-us/windows/kinect/>, свободный. Яз. англ. (дата обращения: 13.05.2020).

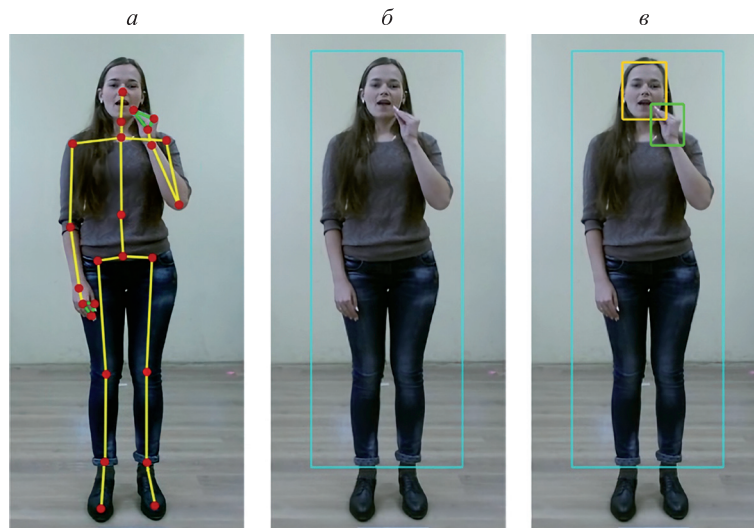


Рис. 2. 25-ти точечная модель скелета человека (а); прямоугольная область с ближайшим человеком (б); определение графических областей лица и форм рук (в)

определение графической области лица и формы руки (рис. 2, в).

Экспериментальное исследование метода

Для выявления оптимального результата по детектированию лиц демонстраторов были протестированы детекторы лиц с помощью базы данных TheRuSLan [18] на основе:

- усовершенствованного метода Виолы–Джонса;
- архитектуры Single Shot MultiBox Detector (SSD) [19] с уменьшенной моделью сети ResNet-10 [20];
- гистограммы направленных градиентов (HOG) и метода опорных векторов (SVM);
- метода Max-Margin Object Detection (MMOD) [21].

В таблице представлен сравнительный анализ используемых детекторов лиц. Для оценки качества работы детекторов использовались такие метрики¹, как среднее от средней точности (mAP), AP₅₀, AP₇₅, AP_{Small (S)}, AP_{Medium (M)}, AP_{Large (L)}, приблизительная скорость обработки кадра (FPS).

Эксперименты по выявлению оптимального детектора лиц производились на вычислительной машине Intel NUC в следующей комплектации: процессор Intel Core i7–8559U с частотой 2,7 ГГц; объем оперативной памяти равный 16 ГБ; твердотельный накопитель формата SSD; графический ускоритель Intel Iris Plus Graphics 655.

Таким образом, в ходе экспериментов выявлено, что для поставленной задачи по определению графической области лица оптимальным детектором является детектор лиц на основе глубокого обучения, который реализован в библиотеке OpenCV². При сравнении с

другими детекторами можно отметить, что он работает при разных ориентациях лица, устойчив к окклюзиям, а также работает в режиме реального времени как на графическом процессоре (GPU), так и на центральном процессоре (CPU).

В случае с определением графической области с формой руки человека использовалась глубокая сверточная нейронная сеть с архитектурой NASNetMobile. Данная архитектура включена в модуль распознавания объектов библиотеки с открытым исходным кодом Keras³.

Обучение глубокой сверточной нейронной сети производилось с помощью размеченных данных с формами рук из базы данных TheRuSLan (48 одноручных жестов) [18] (рис. 3).

Весь процесс аннотирования осуществлялся при помощи инструмента LabelImg⁴. Аннотированные области сохраняются в специальном формате PASCAL VOC [22] в виде текстовых файлов XML. Данный формат широко используется, например, в нейросети ImageNet [23, 24].

Набор данных, состоящий из жестов, был разбит на обучающую и тестовую выборки в примерном соотношении 10:3 демонстраторов (80:20 %).

Распознавание формы руки осуществлялось при следующих условиях:

- обученная модель глубокой сверточной нейронной сети определяет форму руки;
- центральная координата руки, полученная от сенсора Kinect v2, находится в пределах распознанной области с формой руки.

¹ Detection Evaluation [Электронный ресурс]. <http://cocodataset.org/#detection-eval>, свободный. Яз. англ. (дата обращения: 25.05.2020).

² Deep Neural Network module [Электронный ресурс]. https://docs.opencv.org/4.3.0/d6/d0f/group__dnn.html, свободный. Яз. англ. (дата обращения: 13.05.2020).

³ Instantiates a Mobile NASNet model in ImageNet mode [Электронный ресурс]. <https://keras.io/api/applications/nasnet/#nasnetmobile-function>, свободный. Яз. англ. (дата обращения: 13.05.2020).

⁴ LabelImg is a graphical image annotation tool [Электронный ресурс]. <https://github.com/tzutalin/labelImg>, свободный. Яз. англ. (дата обращения: 13.05.2020).

Таблица. Сравнительный анализ детекторов лиц

Детекторы лиц	Реализация	Метрики, отн. ед.						Время распознавания, мс
		mAP	AP ₅₀	AP ₇₅	AP _S	AP _M	AP _L	FPS
Метод Виолы–Джонса	OpenCV	0,15	0,56	0,02	0,04	0,10	0,21	19
SSD + ResNet-10		0,41	0,86	0,25	0,06	0,37	0,46	62
HOG и SVM	Dlib	0,12	0,65	0,01	0,03	0,09	0,17	15
MMOD		0,08	0,44	0,01	0,02	0,07	0,13	9



Рис. 3. Примеры форм рук жестов русского жестового языка

Следующие этапы направлены на вычисление координат области лица и рук с их последующей нормализацией. Затем подсчитывается 2D-расстояние между самой верхней левой координатой области лица и такими же координатами областей рук. Кроме того, подсчитывается площадь пересечения областей лица и рук.

Таким образом, отличительными характеристиками жеста в определенный момент времени являются:

- нормализованные 2D-расстояния от лица до рук (зона артикуляции жеста);
- нормализованные 2D-площади пересечения лица и рук;
- формы рук (представляются числовым значением);
- результат детектирования области рта (представлен числовым значением).

Заключительный этап направлен на распознавание жестов РЖЯ с помощью глубокой нейронной сети с длинной кратковременной памятью (LSTM).

На вход данной нейронной сети подавались функциональные ядра жестов, которые состоят из контекстно-независимых движений рук по отношению к другим жестам. В более расширенном понимании LSTM нейронная сеть принимает последовательность

N кадров на 8 значений из характеристик жеста, в частности:

- нормализованные 2D-расстояния от лица до рук и 2D-площади пересечения лица и рук, представляющие собой числа с плавающей точкой;
- формы рук – целое число;
- результат детектирования области рта представлен числами 0 (область не найдена) и 1 (область найдена).

Процесс обучения производился с помощью библиотеки: глубокого машинного обучения Keras и с открытым исходным кодом TensorFlow v2.

Средняя точность распознавания 48 одноручных жестов из базы данных TheRuSLan составила 78,94 %, что превышает ранее полученные результаты [17]. Наиболее низкую точность распознавания показали жесты, при демонстрации которых формы рук схожи и область артикуляции находится в районе лица.

Заключение

Распознавание элементов жестовых языков — очень актуальная задача в современном информационном мире. Кроме того, до сих пор проблема эффективности

распознавания жестов не решена из-за культурных различий жестовых языков мира, различных условий показов жестов, относительно небольших размеров пальцев рук.

Предложен метод, который объединяет в себе современные алгоритмы компьютерного зрения и машинного обучения, что позволяет осуществлять автоматический видеоанализ движений рук и распознавания жестов в режиме реального времени с помощью оптической камеры и карты глубины для задач человеко-машинного взаимодействия. Данный подход может использоваться в автоматических системах распознавания жестовых языков, а также в тех ситуациях, когда необходимо бесконтактное взаимодействие различных групп пользователей, включая людей с ограниченными возмож-

ностями по слуху и зрению, с мобильными информационными роботами посредством автоматического распознавания жестовой информации.

Стоит заметить, что наиболее существенным недостатком предложенного метода является необходимость наличия специализированной технической составляющей в виде сенсора Kinect v2. Данный недостаток является свойством метода, так как необходимо устройство, которое способно синхронно получать цветные видеоданные и карту глубины в режиме реального времени. Кроме того, решение реализовано таким образом, что в дальнейших исследованиях возможно произвести замену сенсора Kinect v2 на более современное решение в виде Azure Kinect или на аналог.

Литература

1. Карпов А.А., Юсупов Р.М. Многомодальные интерфейсы человеко-машинного взаимодействия // Вестник Российской академии наук. 2018. Т. 88. № 2. С. 146–155. doi: 10.7868/S0869587318020056
2. Kagirow I., Karpov A., Kipyatkova I., Klyuzhev K., Kudryavcev A., Kudryavcev I., Ryumin D. Lower limbs exoskeleton control system based on intelligent human-machine interface // Studies in Computational Intelligence. 2020. V. 868. P. 457–466. doi: 10.1007/978-3-030-32258-8_54
3. Parker L.E., Rus D., Sukhatme G.S. Multiple mobile robot systems // Springer Handbook of Robotics. Springer, Cham, 2016. P. 1335–1384. doi: 10.1007/978-3-319-32552-1_53
4. Ryumin D., Kagirow I., Ivanko D., Axyonov A., Karpov A.A. Automatic detection and recognition of 3D manual gestures for human-machine interaction // International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences — ISPRS Archives. 2019. V. 42. N 2/W12. P. 179–183. doi: 10.5194/isprs-archives-XLII-2-W12-179-2019
5. Mahmud S., Lin X., Kim J.H. Interface for Human Machine Interaction for assistant devices: A Review // Proc. 10th Annual Computing and Communication Workshop and Conference (CCWC). 2020. P. 768–773. doi: 10.1109/CCWC47524.2020.9031244
6. Ivanko D., Ryumin D., Kipyatkova I., Axyonov A., Karpov A. Lip-reading using pixel-based and geometry-based features for multimodal human-robot interfaces // Smart Innovation, Systems and Technologies. 2020. V. 154. P. 477–486. doi: 10.1007/978-981-13-9267-2_39
7. Janssen C.P., Donker S.F., Brumby D.P., Kun A.L. History and future of human-automation interaction // International Journal of Human Computer Studies. 2019. V. 131. P. 99–107. doi: 10.1016/j.ijhcs.2019.05.006
8. Prostejovsky A.M., Brosinsky C., Heussen K., Westermann D., Kreusel J., Marinelli M. The future role of human operators in highly automated electric power systems // Electric Power Systems Research. 2019. V. 175. P. 105883. doi: 10.1016/j.epr.2019.105883
9. Chakraborty B.K., Sarma D., Bhuyan M.K., MacDorman K.F. Review of constraints on vision-based gesture recognition for human-computer interaction // IET Computer Vision. 2018. V. 12. N 1. P. 3–15. doi: 10.1049/iet-cvi.2017.0052
10. Dey D., Habibovic A., Pflöging B., Martens M., Terken J. Color and animation preferences for a light band eHMI in interactions between automated vehicles and pedestrians // Proc. of the 2020 CHI Conference on Human Factors in Computing Systems. 2020. P. 1–13. doi: 10.1145/3313831.3376325
11. Biondi F., Alvarez I., Jeong K.A. Human–Vehicle cooperation in automated driving: A multidisciplinary review and appraisal // International Journal of Human-Computer Interaction. 2019. V. 35. N 11. P. 932–946. doi: 10.1080/10447318.2018.1561792
12. Kennedy J., Lemaignan S., Montassier C., Lavalade P., Irfan B., Papadopoulos F., Senft E., Belpaeme T. Child speech recognition in human–robot interaction: evaluations and recommendations // Proc. 12th ACM/IEEE International Conference on Human-Robot Interaction. 2017. P. 82–90. doi: 10.1145/2909824.3020229
13. Kipyatkova I. LSTM-based language models for very large vocabulary continuous russian speech recognition system // Lecture Notes in Computer Science (including subseries Lecture Notes in

References

1. Karpov A.A., Yusupov R.M. Multimodal interfaces of human-computer interaction. *Herald of the Russian Academy of Sciences*, 2018, vol. 88, no. 1, pp. 67–74. doi: 10.1134/S1019331618010094
2. Kagirow I., Karpov A., Kipyatkova I., Klyuzhev K., Kudryavcev A., Kudryavcev I., Ryumin D. Lower limbs exoskeleton control system based on intelligent human-machine interface. *Studies in Computational Intelligence*, 2020, vol. 868, pp. 457–466. doi: 10.1007/978-3-030-32258-8_54
3. Parker L.E., Rus D., Sukhatme G.S. Multiple mobile robot systems. *Springer Handbook of Robotics*. Springer, Cham, 2016, pp. 1335–1384. doi: 10.1007/978-3-319-32552-1_53
4. Ryumin D., Kagirow I., Ivanko D., Axyonov A., Karpov A.A. Automatic detection and recognition of 3D manual gestures for human-machine interaction. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences — ISPRS Archives*, 2019, vol. 42, no. 2/W12, pp. 179–183. doi: 10.5194/isprs-archives-XLII-2-W12-179-2019
5. Mahmud S., Lin X., Kim J.H. Interface for Human Machine Interaction for assistant devices: A Review. *Proc. 10th Annual Computing and Communication Workshop and Conference (CCWC)*, 2020, pp. 768–773. doi: 10.1109/CCWC47524.2020.9031244
6. Ivanko D., Ryumin D., Kipyatkova I., Axyonov A., Karpov A. Lip-reading using pixel-based and geometry-based features for multimodal human-robot interfaces. *Smart Innovation, Systems and Technologies*, 2020, vol. 154, pp. 477–486. doi: 10.1007/978-981-13-9267-2_39
7. Janssen C.P., Donker S.F., Brumby D.P., Kun A.L. History and future of human-automation interaction. *International Journal of Human Computer Studies*, 2019, vol. 131, pp. 99–107. doi: 10.1016/j.ijhcs.2019.05.006
8. Prostejovsky A.M., Brosinsky C., Heussen K., Westermann D., Kreusel J., Marinelli M. The future role of human operators in highly automated electric power systems. *Electric Power Systems Research*, 2019, vol. 175, pp. 105883. doi: 10.1016/j.epr.2019.105883
9. Chakraborty B.K., Sarma D., Bhuyan M.K., MacDorman K.F. Review of constraints on vision-based gesture recognition for human-computer interaction. *IET Computer Vision*, 2018, vol. 12, no. 1, pp. 3–15. doi: 10.1049/iet-cvi.2017.0052
10. Dey D., Habibovic A., Pflöging B., Martens M., Terken J. Color and animation preferences for a light band eHMI in interactions between automated vehicles and pedestrians. *Proc. of the 2020 CHI Conference on Human Factors in Computing Systems*, 2020, pp. 1–13. doi: 10.1145/3313831.3376325
11. Biondi F., Alvarez I., Jeong K.A. Human–Vehicle cooperation in automated driving: A multidisciplinary review and appraisal. *International Journal of Human-Computer Interaction*, 2019, vol. 35, no. 11, pp. 932–946. doi: 10.1080/10447318.2018.1561792
12. Kennedy J., Lemaignan S., Montassier C., Lavalade P., Irfan B., Papadopoulos F., Senft E., Belpaeme T. Child speech recognition in human–robot interaction: evaluations and recommendations. *Proc. 12th ACM/IEEE International Conference on Human-Robot Interaction*, 2017, pp. 82–90. doi: 10.1145/2909824.3020229
13. Kipyatkova I. LSTM-based language models for very large vocabulary continuous russian speech recognition system. *Lecture Notes in Computer Science (including subseries Lecture Notes in*

- Artificial Intelligence and Lecture Notes in Bioinformatics). 2019. V. 11658. P. 219–226. doi: 10.1007/978-3-030-26061-3_23
14. Ryumin D., Karpov A.A. Towards automatic recognition of sign language gestures using kinect 2.0 // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017. V. 10278. P. 89–101. doi: 10.1007/978-3-319-58703-5_7
 15. Mazhar O., Ramdani S., Navarro B., Passama R. A Framework for real-time physical Human-Robot Interaction using hand gestures // *Proc. of the 2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, 2018. P. 46–47. doi: 10.1109/ARSO.2018.8625753
 16. Рюмин Д. Метод обнаружения и распознавания 3D одноручных жестов рук для человеко–машинного взаимодействия // Сборник тезисов докладов конгресса молодых ученых. 2019 [Электронное издание]. URL: <https://kmu.itmo.ru/digests/article/1902>, свободный. Яз. рус. (дата обращения: 13.05.2020).
 17. Kagirow I., Ryumin D., Axyonov A. Method for multimodal recognition of one-handed sign language gestures through 3D convolution and LSTM neural networks // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019. V. 11658. P. 191–200. doi: 10.1007/978-3-030-26061-3_20
 18. Кагиrow И.А., Рюмин Д.А., Аксёнов А.А., Карпов А.А. Мультимедийная база данных жестов русского жестового языка в трехмерном формате // *Вопросы языкознания*. 2020. № 1. С. 104–123. doi: 10.31857/S0373658X0008302-1
 19. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C-Y., Berg A. SSD: single shot multibox detector // *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016. V. 9905. P. 21–37. doi: 10.1007/978-3-319-46448-0_2
 20. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition // *Proc. 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. P. 770–778. doi: 10.1109/CVPR.2016.90
 21. King D.E. Max-margin object detection // *arXiv*. preprint arXiv 1502.00046. 2015.
 22. Parkhi O.M., Vedaldi A., Zisserman A. Deep face recognition // *Proc. 26th British Machine Vision Conference (BMVC)*, 2015. P. 41.1–41.12. doi: 10.5244/C.29.41
 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. doi: 10.1145/3065386
 24. Everingham M., Van Gool L., Williams C.K., Winn J., Zisserman A. The pascal visual object classes (VOC) challenge // *International Journal of Computer Vision*, 2010. V. 88. N 2. P. 303–338. doi: 10.1007/s11263-009-0275-4
 - Artificial Intelligence and Lecture Notes in Bioinformatics), 2019, vol. 11658, pp. 219–226. doi: 10.1007/978-3-030-26061-3_23
 14. Ryumin D., Karpov A.A. Towards automatic recognition of sign language gestures using kinect 2.0. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2017, vol. 10278, pp. 89–101. doi: 10.1007/978-3-319-58703-5_7
 15. Mazhar O., Ramdani S., Navarro B., Passama R. A Framework for real-time physical Human-Robot Interaction using hand gestures. *Proc. of the 2018 IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO)*, 2018, pp. 46–47. doi: 10.1109/ARSO.2018.8625753
 16. Riumin D. Detection and recognition method of 3D single-handed gestures for human-machine interaction. *Proc. Conferences of Young Scientists*, 2019. Available at: <https://kmu.itmo.ru/digests/article/1902> (accessed: 13.05.2020). (in Russian)
 17. Kagirow I., Ryumin D., Axyonov A. Method for multimodal recognition of one-handed sign language gestures through 3D convolution and LSTM neural networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11658, pp. 191–200. doi: 10.1007/978-3-030-26061-3_20
 18. Kagirow I.A., Ryumin D.A., Axyonov A.A., Karpov A.A. Multimedia database of Russian sign language items in 3D. *Voprosy Jazykoznanija*, 2020, no. 1, pp. 104–123. (in Russian). doi: 10.31857/S0373658X0008302-1
 19. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C-Y., Berg A. SSD: single shot multibox detector. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2016, vol. 9905, pp. 21–37. doi: 10.1007/978-3-319-46448-0_2
 20. He K., Zhang X., Ren S., Sun J. Deep residual learning for image recognition. *Proc. 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 770–778. doi: 10.1109/CVPR.2016.90
 21. King D.E. Max-margin object detection. *arXiv*, preprint arXiv 1502.00046. 2015.
 22. Parkhi O.M., Vedaldi A., Zisserman A. Deep face recognition. *Proc. 26th British Machine Vision Conference (BMVC)*, 2015, pp. 41.1–41.12. doi: 10.5244/C.29.41
 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. doi: 10.1145/3065386
 24. Everingham M., Van Gool L., Williams C.K., Winn J., Zisserman A. The pascal visual object classes (VOC) challenge. *International Journal of Computer Vision*, 2010, vol. 88, no. 2, pp. 303–338. doi: 10.1007/s11263-009-0275-4

Авторы

Рюмин Дмитрий — научный сотрудник, Санкт-Петербургский институт информатики и автоматизации Российской академии наук (СПИИРАН), Санкт-Петербург, 199178, Российская Федерация, Scopus ID: 57191960214, ORCID ID: 0000-0002-7935-0569, dl_03.03.1991@mail.ru

Authors

Dmitry Ryumin — Researcher, St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint Petersburg, 199178, Russian Federation, Scopus ID: 57191960214, ORCID ID: 0000-0002-7935-0569, dl_03.03.1991@mail.ru