**УДК 004.852**

# Метод обнаружения и распознавания 3D одноручных жестов рук для человеко-машинного взаимодействия

Д. Рюмин
(Университет ИТМО, г. Санкт-Петербург)
Научный руководитель – д.т.н., доцент А.А. Карпов
(Университет ИТМО, г. Санкт-Петербург)

**Введение.** Задача повышения уровня автоматизации и роботизации всех сфер деятельности человека является одной из ключевых в современном информационном обществе [1]. В связи с этим ученые и руководства развитых, а также развивающихся стран в сотрудничестве с мировыми научными центрами и компаниями уделяют внимание технологиям для эффективного, естественного и универсального взаимодействия человека с компьютерами и роботами [2].

В настоящее время интерактивные информационные системы получают применение в сферах социального обслуживания, медицине, образования, робототехнике, военной индустрии, центрах обслуживания населения, а также для взаимодействия с людьми в различных чрезвычайных ситуациях [3]. Кроме того, все более широкое распространение находят роботы-ассистенты, которые направлены на взаимодействие с людьми для выполнения определенных задач. В этом случае, многих классических интерфейсов недостаточно. Вместо этого необходимы более интуитивные и естественные для человека интерфейсы (речевой, жестовый, многомодальный [4] и т.п.). Так, например, жесты могут передавать простые команды роботу, которые будут нести однозначный смысл и эффективны на некотором расстоянии от робота и в шумных условиях, когда речь малоэффективна.

Также известно, что инвалиды по слуху ограничены в возможностях при общении со слышащими, а при обращении в различные учреждения к ним прикрепляются сурдопереводчики, которых оказывается недостаточно. Поэтому необходимы технологии распознавания жестовых языков глухих людей, при помощи которых будет осуществлять управление и взаимодействие с ассистивными мобильными информационными роботами.

**Цель работы** заключается в разработке метода автоматического обнаружения и распознавания 3D одноручных жестов рук для человеко-машинного взаимодействия.

**Базовые положения исследования.** На вход разработанного метода подаются видеоданные в двумерном (RGB камера) и трехмерном формате (карта глубины) или непосредственно от сенсора Microsoft Kinect v2. Разрешение цветных видеокадров составляет 1920x1080 пикселей, а карты глубины 512x424 пикселей с частотой 30 кадров в секунду. Качество цветопередачи для двумерных данных составляет 8 бит, а для трехмерных 16 бит. Осуществляется потоковая обработка видеоданных. На каждом кадре, используя карту глубины и набор инструментов разработки поставляемых с сенсором Kinect v2, осуществляется поиск людей на расстоянии от 1,2 до 3,5 метров от камеры. Затем для каждого найденного человека вычисляется 3D модель скелета с 25 реперными точками. Трехмерные координаты преобразуются в 2D, на основе которых формируются прямоугольные области, которые обрамляют найденных людей. Кроме того, две модели глубокой сверточной сети используются для определения формы лица и руки. Детектор лица основан на структуре Single Shot MultiBox Detector (SSD) с уменьшенной сетевой моделью ResNet-10. Этот детектор включен в модуль Deep Neural Networks (dnn) библиотеки компьютерного зрения Open Source Computer Vision (OpenCV). Детектор формы руки основан на структуре SSD с моделью сети MobileNetV2. Этот детектор обучен на

мультимедийной базе данных трехмерных жестов русского языка жестов. Для аннотирования базы данных использовался инструмент LabelImg. Аннотированные данные сохраняются в виде файлов расширяемого языка разметки (XML). Данный формат используется в ImageNet. Кроме того, LabelImg также поддерживает YOLO формат. Для обучения детектора использовались первые 4 повторения жеста. Эти данные считаются шаблоном, а остальные использовались в качестве тестовых данных.

Завершающим этапом метода является распознавание одноручных 3D жестов с использованием современных алгоритмов машинного обучения, которые включены в такие инструменты, как scikit-learn, Tensorflow object detection, Keras. Данные инструменты упрощают создание, обучение и развертывание как статических, так и динамических объектов.

**Основной результат.** Таким образом, предложенный в работе метод позволяет обнаруживать и распознавать одноручные трехмерные жесты в реальном времени. Благодаря своей универсальности данный метод может быть использован в задачах биометрии, компьютерном зрении, машинном обучении, автоматических системах распознавания лиц и языка жестов.

**Литература**

1. Ryumin D., Karpov A. Towards Automatic Recognition of Sign Language Gestures using Kinect 2.0 // In Proc. 19th International Conference on Human-Computer Interaction, HCII, Vancouver, Canada, Springer LNCS. – 2017. – V. 10278. – P. 89-104.
2. Ivanko D., Karpov A., Fedotov D., Kipyatkova I., Ryumin D., Ivanko Dm., Minker W., Zelezny M. Multimodal Speech Recognition: Increasing Accuracy using High Speed Video Data // Journal on Multimodal User Interfaces, Springer. – V. 12. – I. 4. – P. 319-328, https://doi.org/10.1007/s12193-018-0267-1.
3. Toyota Global Site. – 2018. – Partner Robot, http://www.toyota-global.com/innovation/partner_robot.
4. Ivanko D., Ryumin D., Axyonov A., Železný M. Designing Advanced Geometric Features for Automatic Russian Visual Speech Recognition // In Proc. 20th International Conference on Speech and Computer SPECOM-2018, Leipzig, Germany, Springer, LNAI. – V. 11096. – P. 245-254.