



УДК 004.822:514

СИСТЕМА РАСПОЗНАВАНИЯ ЖЕСТОВ РУССКОГО ЯЗЫКА ГЛУХИХ

Дорофеев Н.С., Розалиев В.Л., Заболева-Зотова А.В.

*Волгоградский Государственный Технический Университет,
г. Волгоград, Россия*

nikita.dorofei@gmail.com

vladimir.rozaliev@gmail.com

zabzot@gmail.com

Данная статья посвящена исследованиям в области автоматизации распознавания жестов русского языка глухих. Авторами предложена информационная модель распознавания дактильных жестов с использованием камеры с сенсором глубины. Разработана удобная модель описания кисти руки, использование которой значительно повышает точность распознавания. Также разработан улучшенный метод детектирования рук на изображении.

Ключевые слова: язык жестов; распознавание жестов; камера с сенсором глубины

ВВЕДЕНИЕ

Адаптация слабослышащих и глухих людей к жизни современного общества является крайне важной задачей. Её достижению могут помочь новые технологии в сфере распознавания жестов. Традиционные визуальные методы распознавания жестов всё ещё недостаточно точные для использования в реальных приложениях. Одной из причин этому являются ограничения оптических сенсоров чувствительных к условиям освещения и шумам фона. Для достижения высокой надежности распознавания жестов также используются информационные перчатки. Такие сенсоры гораздо надежнее, но имеют существенные недостатки: пользователю приходится носить перчатку, которая иногда требует калибровки, что очень неудобно и делает движения не натуральными. Также перчатки зачастую очень дорогие, хотя в последнее время и ведутся работы по уменьшению стоимости таких устройств. В результате, этот метод распознавания жестов не очень популярен. Благодаря разработке компании Microsoft камеры с сенсором глубины Kinect, появились новые возможности для распознавания жестов. Сенсоры глубины существовали уже относительно давно, но Kinect обладает перед ними рядом значительных преимуществ: большое распространение, относительно небольшая стоимость и наличие RGB камеры. Получаемая с Kinect карта глубины инвариантна к условиям освещения и фону,

поскольку она основана на инфракрасном излучении. Помехой для данного типа записывающих устройств может быть только сильный туман и некоторые другие погодные условия.

Несмотря на успешные применения Kinect к распознаванию лица и отслеживанию тела человека, проблема использования этого сенсора к распознаванию небольших жестов рук до сих пор является нерешенной. Основная причина - низкое разрешение карты глубины сенсора [Kean et al., 2011].

В языках жестов передача информации во время общения происходит по нескольким каналам: непосредственно через жесты руками, выражение лица, форму губ, положение тела и головы. Сами жесты руками описываются через положение рук, направление движения, форму и направление кистей рук. Таким образом, возможность определения формы и положения кистей рук является очень важной задачей в контексте распознавания жестового языка, которая еще не была полностью решена.

В данном исследовании была поставлена цель расширения возможностей коммуникации глухонемых с окружающим миром за счёт распознавания дактильных букв русского языка жестов в режиме реального времени. В качестве устройства ввода информации о жесте использовался Microsoft Kinect.

1. Информационная модель распознавания статических жестов

Весь процесс распознавания дактильных жестов можно представить в виде информационной модели:

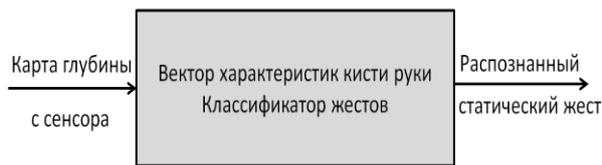


Рисунок 1 - Информационная модель распознавания статических жестов

$$IM = \langle FV, C, I, O \rangle, \quad (1)$$

где IM – информационная модель;

FV – вектор характеристик кисти руки. Первоначальная информация об изображении (цвет и расстояние) анализируется: определяется положение кистей рук и вектор их характерных черт;

C – классификатор жестов. По найденному вектору характерных черт кистей, определяется, какой жест был показан;

I – множество входных значений – карта глубины – расстояние до каждой точки на изображении;

O – множество выходных значений – распознанный на изображении статический жест.

2. Метод детектирования рук на изображении

Первым этапом распознавания является сегментация полученного изображения с сенсора Kinect с целью нахождения на нем кисти руки или обеих рук. Разработка метода для нахождения кисти руки на изображении является одной из самых сложных проблем в процессе создания системы распознавания жестов. Существует несколько признаков, по которым можно детектировать объект на изображении: внешность, форма, цвет, расстояние до объекта и контекст. В таких случаях, как детектирование лица на изображении, хорошим признаком является внешность, так как глаза, нос и рот всегда будут находиться примерно в одинаковых пропорциях. Поэтому основанный на характеристиках внешности объекта метод каскадов Хаара отлично применяется для распознавания лица. В случае распознавания рук дело обстоит сложнее: надежный метод распознавания может быть реализован основываясь, в основном, на цветовых характеристиках. Так как цвет рук может меняться в зависимости от человека и контекста, представляется разумным сначала найти лицо человека на изображении и получать информацию о цвете рук исходя из цвета лица. Введенное ограничение наличия лица человека на изображении

в любом случае является обязательным, поскольку распознавание жестового языка без распознавания лица будет ненадежным.

Имея информацию о цвете объекта, необходимо детектировать его на изображении. Осуществить данную задачу можно с помощью алгоритма Camshift, надежность которого доказана в работах [Hai et al., 2011]. Модель этого алгоритма основана на гистограммах и является обучающейся в процессе распознавания. Естественно, данный алгоритм найдет все объекты данного цвета на изображении. Чтобы этого не допустить, используется информация о расстоянии до объектов на изображении, т.е. карта глубины с сенсора Kinect.

Итак, после нахождения позиции (x,y) и размеров (w,h) лица на изображении с помощью метода каскадов Хаара, можно найти усредненное расстояние до лица, используя карту глубины D:

$$d_f = \frac{1}{wh} \sum_{i=x}^{x+w} \sum_{j=y}^{y+h} D(i, j). \quad (2)$$

Все объекты, находящиеся ближе к камере, чем само лицо человека, могут быть найдены с использованием порогового значения:

$$D(i, j) > d_f + t_h, \quad (3)$$

где t_h – параметр, определяющий, насколько близко должны быть поднесены руки к камере, чтобы показанные жесты распознавались системой.

Таким образом, в системе делается допущение, что руки находятся ближе к камере, чем остальные части тела человека, что вполне обосновано в контексте системы распознавания жестового языка.

3. Исследование контура кисти руки

На втором этапе распознавания на сегментированном изображении находятся контуры оставшихся объектов, т.е. рук человека. Экспериментально было выяснено, что самым эффективным и информативным методом нахождения контуров на изображении в контексте данного исследования, является детектор границ Кенни. После его применения, полученные контуры представляются в виде кривой на графике (для каждой руки своя кривая), как показано на рисунке 2.

График показывает относительное расстояние каждой точки контура к центральной точке. Последняя определяется, как центр вписанной в контур окружности с максимальным радиусом. Начальная точка определяется согласно линии, найденной алгоритмом RANSAC по черному ремешку на запястье человека. В данном представлении, на горизонтальной оси находится угол между каждой точкой контура и начальной точкой относительно центральной точки, нормализованный на 360°. На вертикальной оси

находится относительное Евклидово расстояние между каждой точкой контура и центральной точкой, нормализованное по радиусу максимальной вписанной окружности. Такое представление кисти руки хорошо описывает её топологические свойства – расположение пальцев. Положение последних можно определить, задав некоторое пороговое значение t . Хотя найти оптимальное значение t является непростой задачей. [3]

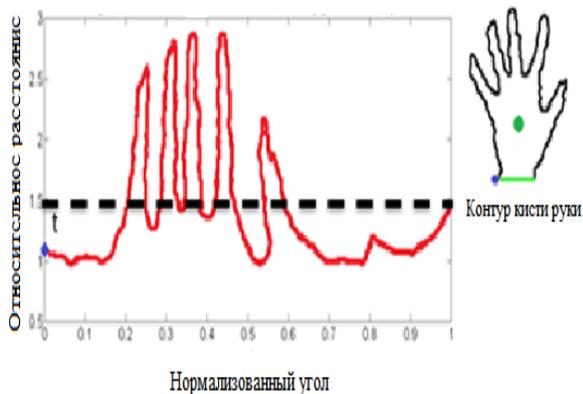


Рисунок 2 - Графическое представление контура кисти руки

4. Классификация полученных данных

На третьем этапе необходимо выбрать метод классификации полученных данных. В ходе исследования было протестировано более 10 различных методов классификации. Самым точным методом показал себя многослойный персептрон (98% в некоторых случаях). Надежность нейросетевого подхода подтверждается и другими исследованиями [Розалиев и др., 2010], [Заболеева-Зотова и др., 2011a]. Далее на основе обучающей выборки, состоящей из данных, описывающих жест в виде, представленном на рисунке 2, была обучена нейронная сеть с найденными в ходе экспериментов оптимальными параметрами. Для тестирования был разработан прототип системы распознавания дактильных букв. Нейронная сеть была обучена 13 жестам – буквам русского жестового языка. Тесты показали достаточно высокую точность распознавания – 88%.

ЗАКЛЮЧЕНИЕ

В ходе данного исследования была разработана система распознавания дактильных жестов русского жестового языка, которая была обучена 13 жестам.

Как было отмечено, жестовый язык состоит из целого ряда каналов передачи информации, поэтому распознавание небольших жестов рук не решает проблему распознавания русского жестового языка полностью, но является важным элементом будущей полноценной системы.

Предметом дальнейшего исследования является поиск более точного, чем пороговое значение,

метода определения характерных черт кисти руки на графике, поскольку при некоторых положениях пальцев могут происходить коллизии. Из-за этого возникают затруднения в обучении системы полной азбуке дактильных жестов.

Полученные результаты планируется использовать в работах по определению эмоциональных реакций человека [Заболеева-Зотова и др., 2011b], [Розалиев и др., 2010], [Заболеева-Зотова и др., 2012]. Используемые в данном исследовании методы распознавания мелких движений рук человека, включая движения пальцами, безусловно, найдут применение и в этой области.

Работа частично поддержана Российским фондом фундаментальных исследований (проекты 12-07-00266, 12-07-00270).

Библиографический список

[Kean et al., 2011] Sean Kean, Jonathan Hall, Phoenix Perry, Meet the Kinect: An Introduction to Programming Natural User Interfaces (Technology in Action). Apress, 2011. – 220 с.

[Hai et al., 2011] H. Hai, L. Bin, H. BenXiong and C. Yi, "Interaction System of Treadmill Games based on depth maps and CAM-Shift", IEEE 3rd International Communication Software and Networks (2011), pp.219-222

[Заболеева-Зотова и др., 2011a] Применение нечётких темпоральных высказываний для описания движений при эмоциональных реакциях / А.В. Заболеева-Зотова, Ю.А. Орлова, В.Л. Розалиев, А.С. Бобков // Изв. ВолгГТУ. Серия "Актуальные проблемы управления, вычислительной техники и информатики в технических системах". Вып. 10 : межвуз. сб. науч. ст. / ВолгГТУ. - Волгоград, 2011. - № 3. - С. 60-64.

[Розалиев и др., 2010] Розалиев, В.Л. Применение нейронных сетей и грануляции при построении автоматизированной системы определения эмоциональных реакций человека / В.Л. Розалиев, А.С. Бобков, О.С. Федоров // Изв. ВолгГТУ. Серия «Актуальные проблемы управления, вычислительной техники и информатики в технических системах»: межвуз. сб. науч. ст. / ВолгГТУ. - Волгоград, 2010. - Вып. 9, № 11. - С. 63-68.

[Заболеева-Зотова и др., 2011b] Развитие системы автоматизированного определения эмоций и возможные сферы применения / А.В. Заболеева-Зотова, Ю.А. Орлова, В.Л. Розалиев, А.С. Бобков // Открытое образование. - 2011. - № 2. - С. 59-62.

[Розалиев и др., 2010] Розалиев, В.Л. Моделирование эмоционального состояния человека на основе гибридных методов / В.Л. Розалиев, А.В. Заболеева-Зотова // Программные продукты и системы: международный науч.-практ. журнал. – Тверь, 2010 – Вып.2 (90). – С.141-146.

[Заболеева-Зотова и др., 2012] Заболеева-Зотова, А.В. Automated Identification of Emotional States / Заболеева-Зотова А.В., Розалиев В.Л., Бобков А.С. // Advances in Decision Technology and Intelligent Information Systems. Vol. XIII : 24th Int. Conf. of Systems Research, Informatics and Cybernetics (July 30 – August 3, 2012, Baden-Baden, Germany) / The Int. Institute for Advanced Studies in Systems Research and Cybernetics. – [Tecumseh], Canada, 2012. – P. 21-25. – Англ.

SYSTEM FOR RUSSIAN SIGN LANGUAGE RECOGNITION

Dorofeev N.S., Rozaliev V.L., Zaboleeva-Zotova A.V.

*Volgograd State Technical University,
Volgograd, Russia*

zabzot@gmail.com

vladimir.rozaliev@gmail.com

nikita.dorofei@gmail.com

This article is devoted to research in the field of gesture recognition of the Russian Sign Language. The authors propose an information model of dactyl gesture recognition using sensor Microsoft Kinect. A convenient model for hand description was developed, the use of which greatly improves the accuracy of recognition. Also an improved method for hands detection in the picture was elaborated.

INTRODUCTION

Adaptation of hearing impaired people to the life of the modern society is an extremely important problem. New technologies in the field of gesture recognition can help to solve it. Traditional visual methods of gesture recognition are still not accurate enough for using in real applications. One of the reasons for this is that optical sensors have some limitations and they are very sensitive to lighting conditions and background noises.

To achieve high reliability of gesture recognition data gloves are often used. These sensors are much more reliable, but have significant drawbacks: the user has to wear a glove, which sometimes requires calibration, what is very uncomfortable and makes movements not natural. Also gloves are often very expensive, although in recent years situation has become better. As a result, this method of gesture recognition is not very popular. Due to the development of Kinect – Microsoft's camera with depth sensor, new opportunities for gesture recognition appeared. Despite the successful use of Kinect to face recognition and body tracking, the problem of application of this sensor to the recognition of subtle hand gestures is still not solved. The main reason – sensor has low resolution of depth map.

In sign languages information transmission during the communication is performed using several channels: directly through the hands gestures, face expression, mouth shape, location of body and head. Hand gestures can be described through hands and arms location, direction of the movement, shape of palms and their orientation. Thus capability of hands shape and direction recognition is very important task in the context of sign language recognition, which isn't solved completely yet.

This research aims to enlarge opportunities for deaf people to communicate with the outside world by recognizing dactyl letters of the Russian sign language

in real time. As an input device for gestures Microsoft Kinect was used.

MAIN PART

In the first section of this report information model of static gesture recognition was presented.

In the second part method for hand tracking on the image was suggested. It is based on the skin color. Since hands color can change depending on human and context, it is reasonable to find face on the image as a first step and then use its color to find all skin-colored objects, including hands. To perform that task Haar cascades are used for face detection and Camshift algorithm for further segmentation based on previously found skin color. To find hands on the image depth map from the Kinect is used. Average distance to the face is calculated. All objects that are situated closer to the camera than face are considered to be hands (threshold value also can be specified).

In the next section approach for analysis of found hand is described. To extract main features of shown gesture only contours of hands are studied. Values of relative distance from each point on the contour to the contour center are calculated. Angles between starting point and each point on the contour, normalized by 360°, are taken into account too. Starting point is defined as the first point on the black belt, which is put on the wrist. Those values describe hand well from the topological point of view.

In the fourth part of the report method for classification of obtained features vector is discussed. During the research more than 10 classification methods were tried. The best results were shown by neural networks. For testing purposes the system of Russian sign language dactyl letters was developed. It was taught 13 gestures. Average accuracy of recognition is quite high – 88%.

CONCLUSION

The system for Russian sign language dactyl letters was developed in the course of that research. It has shown quite high level of recognition accuracy on 13 gestures - letters of Russian sign language.

Results of that research are going to be used in works on emotional reactions recognition. Methods of subtle hands movement recognition used in this work undoubtedly will have application in that field too.

This work was partially supported by the Russian Foundation for Basic Research (projects 11-07-00398, 12-07-00266, 12-07-00270).