

РАСПОЗНАВАНИЕ ЖЕСТОВ РУК ДЛЯ ИСПОЛЬЗОВАНИЯ В МОБИЛЬНОМ ПРИЛОЖЕНИИ СУРДОПЕРЕВОДА

Видман В.В., Видман А.Я, Саклаков В.М

к.т.н., доцент С.В.Аксенов

Национальный исследовательский Томский политехнический университет
vitalii_vidman@mail.ru

Введение

Для того, чтобы люди с проблемами слуха могли общаться со здоровыми людьми необходимо приложение, которое сможет не только переводить текст или речь на язык жестов, но и жесты интерпретировать в текст. Данная проблема в настоящее время еще не решена. На данный момент нет ни одного приложения, которое смогло бы осуществлять двусторонний сурдоперевод.

В данной статье будут рассмотрены программно-аппаратные средства, которые могут быть применены, а также будет приведена их сравнительная характеристика.

Основная часть

В данный момент разработан прототип приложения для сурдоперевода. В данном приложении имеется три раздела: алфавит – для изучения дактильной азбуки, словарь – для изучения русских жестов сурдоязыка и переводчик – раздел, в котором можно перевести текст или речь на язык жестов. В настоящее время словарь включает в себя более 200 слов, и в течении года их количество планируется увеличить до 800. Если во время перевода данные слова не находятся в базе, то они переводятся с помощью дактильной азбуки. Все воспроизведенные анимации сопровождаются артикуляцией, цель которой увеличение понятности жестов.

Для ввода функции перевода жестов в текст или речь необходимо наличие в приложении специализированной камеры и функционала, которой сможет интерпретировать жесты как определенные слова.

В качестве камеры для распознавания жестов в приложении рассматривались камеры от трех производителей: RealSense от Intel, Kinect от Microsoft, а также Leap Motion от OcuSpec.

Камера Kinect содержит:

- Камера видимого диапазона — обычная RGB-камера: 640×480 и 30 кадров в секунду.
- Инфракрасный лазерный проектор, который создает в пространстве сетку из точек.
- Камера, снимающая в инфракрасном спектре.
- Stereo микрофон с продвинутой системой шумоподавления — для правильного голосового управления.
- Мотор, регулирующий положение датчика.

– Чип PrimeSensor — производит основную работу, обрабатывая изображение с ИК-камеры, отправляет на выход — 3D-картинку.

Камеры Kinect позволяют распознавать жесты на расстоянии до 4,5 метров, но они не направлены на детальное распознавание пальцев кистей [1].

Минус камеры Kinect в том, что она больше направлен на игровую индустрию. А новые камеры Kinect 2 разрабатываются только для Xbox One.

Еще одним представителем технологий компьютерного зрения является технология Leap Motion. Технология Leap Motion основана на захвате (сканировании) движений пальцев и кистей рук с последующей оцифровкой, что позволяет перемещать курсор одним взмахом руки в воздухе и запускать программу или открывать нужное окно, просто ткнув пальцем в пустоту.

Программно-аппаратный комплекс Leap Motion распознаёт запястье, ладонь и пальцы двух рук отдельно, поэтому действие может быть абсолютно любым, начиная с нажатия клавиш на виртуальной клавиатуре и заканчивая имитацией стрельбы из виртуальной винтовки или сборкой какой-нибудь модели из виртуального конструктора [2]. Но поскольку руки оно может распознавать лишь по отдельности, то данное решение не подходит для использования в разрабатываемом приложении.

Компания Intel выпустила несколько камер использующих технологию RealSense. Все камеры используются для того, чтобы распознавать объекты, но у них имеются немного разные уклоны. Так, к примеру, ZR300 используется для сканирования помещений и создания их трехмерной модели. А камера F200 для создания 3D моделей человека. Рассмотрев все предложенные камеры от Intel на данный момент, выбор был остановлен на SR300, так как специализация этой камеры наиболее точно подходила для разрабатываемого приложения. Данная камера способна распознавать две кисти включая 10 пальцев.

Камера может сделать точный скелет полной руки, сегментирующий руку от фона и не требующий никакой калибровки. Скелет отслеживает позицию и ориентацию 22 соединений в руке (рис. 1).

Приложения имеют доступ к этой 3D модели, которая может надежно обработать недостающую информацию, такую как пальцы, находящиеся вне поля зрения [3].

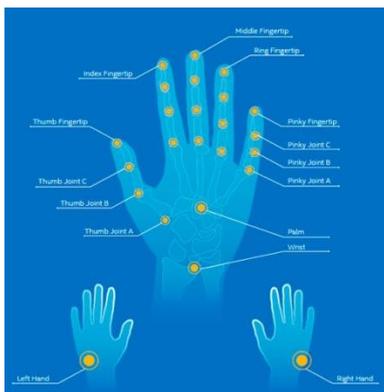


Рис. 1. Распознаваемые камерой точки рук

При создании систем по распознаванию жестов, рук, образов используются методы, в которых вычисляется площадь изображения и далее она сравнивается с имеющимися образцами. Данные методы применяются при создании приложений, где не используется большое количество жестов, возможно, это 10-20 различных жестов. Большой проблемой является то, что количество жестов в жестовом языке велико, и все они имеют разные значения. В языке жестов тактильная азбука представляет собой статические жесты, но со словами дело обстоит по-другому. Большинство слов являются динамическими, поэтому для распознавания необходимо запоминать изменение положения костей. Еще одной проблемой является то, что каждый человек индивидуален, и показываемый жест может отличаться от внесенного в словарь распознаваемых жестов. Для решения данной проблемы необходимо учитывать погрешность в движении. В этом случае из-за допущения определенного уровня погрешностей жест может быть похож сразу на несколько других жестов. Также необходимо будет учитывать скорость показа жестов, это может существенно влиять на результат.

Заключение

Данный этап разработки является переходным этапом. Поскольку подключение специальной камеры возможно только к компьютеру, разработка ведется на данный тип устройств. Главной целью является разработка технологии, с помощью которой будет возможно распознавание моторики человеческих движений, количество которых достаточно большое и многие из них похожи друг на друга.

После изучения трех камер от разных производителей была составлена сравнительная характеристика (таблица 1).

Таблица 1. Сравнение устройств

Устройство	Kinect 1.0	Leap Motion	RealSense SR300
Фокусное расстояние	0.5-4.5	0.07-0.6	0.2-1.5
Горизонтальный угол обзора	57	140	59
Вертикальный угол обзора	43	140	46
микрофон	есть	нету	Есть
Разрешение картинки	640*480	-	1920-1080
Распознавание одновременно двух кистей полностью	нет	нет	да

По таблице приведенной выше можно сделать вывод, что для реализации поставленной цели, а именно создание функционала для сурдоперевода, подходит только камер RealSense. Создав данный функционал, появится возможность использования его в дальнейшем в мобильных приложениях. Уже сейчас имеется мобильное устройство от Intel, но его приобретение пока не доступно (рис. 2). В настоящее время технологии компьютерного зрения получают большую популярность, и готовые мобильные решения появятся на рынке в скором времени.

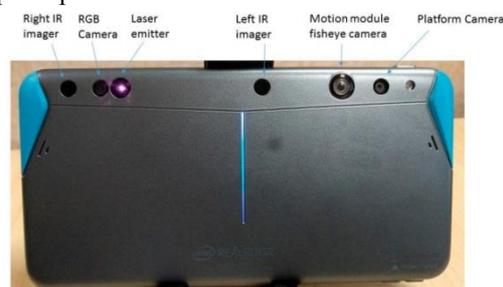


Рис. 2. Мобильное устройство, использующее камеру Intel Realsense ZR300

Список использованных источников

1. Kinect: разбираемся с новым девайсом и учимся писать для него приложения - «Хакер» [Электронный ресурс]. – URL: <https://xaker.ru/2011/09/21/56900/> (дата обращения 05.09.2017).
2. Обзор контроллера Leap Motion: пассы руками / Периферия [Электронный ресурс]. – URL: <https://3dnews.ru/908660> (дата обращения 15.09.2017).
3. Возможности распознавания камеры Intel® RealSense™ [Электронный ресурс]. – URL: <https://www.intel.ru/content/www/ru/ru/support/emerging-technologies/000006099.html> (дата обращения 28.07.2017)