

УДК 004.853

АВТОМАТИЗИРОВАННОЕ ВЫДЕЛЕНИЕ КИСТЕЙ РУК ЧЕЛОВЕКА ДЛЯ РАСПОЗНАВАНИЯ ЖЕСТОВОЙ РЕЧИ

Розалиев В.Л., Агафонов Г.В., Кириченко М.И.

Волгоградский государственный технический университет, г. Волгоград, Россия

vladimir.rozaliev@gmail.com

Данная статья посвящена проблеме детектирования и анализа кисти с помощью камеры глубины. Особое внимание уделено анализу и трекингу кисти. В работе скомбинированы алгоритмы поиска, трекинга и анализа кистей рук человека. Рассмотрены существующие системы распознавания жестов.

Ключевые слова: язык жестов; детектирование; Kinect; depth изображение.

Введение

Важной проблемой является создание современных информационных технологий, в том числе методов, с помощью которых можно создавать новые компьютерные системы обучения и коммуникации для людей с нарушениями слуха. В мире насчитывается более 1 процента людей с заболеваниями, проявляющимися нарушением звуковосприятия, по России же их количество превышает 1000000 человек. При создании таких систем необходимо разрабатывать методы распознавания жестов, их интерпретации и перевода в текст на естественном языке.

1. Жестовая речь

Речь жестовая (от франц. *gestes* – деяния) – способ межличностного общения людей, лишенных слуха, при помощи программы жестов, характеризующейся своеобразными лексическими и грамматическими закономерностями. Закономерности жестовой речи обусловлены выраженным своеобразием ее основной семантической единицы – жеста, а также его функциональным назначением (использованием в сфере непринужденного общения). Жестовая речь используется как вспомогательное средство (наряду с основным – словесной речью) в процессе обучения и воспитания детей с недостатками слуха. В сфере официального общения (собрания, перевод лекций) применяется калькирующая жестовая речь, когда жесты последовательно используются для воспроизведения слов. В современном обществе полезно знание интернациональных языков, а жестовый язык является одним из самых понятных во всем мире.

Жесты различают статические и динамические жестов. Под статическим жестом подразумевается положение руки в пространстве в отсутствие каких-либо движений. Напротив, динамический жест характеризуется последовательным движением руки в пространстве из начальной точки в конечную за фиксированный промежуток времени.

В наши дни существует специальная наука – семиотика, занимающаяся изучением знаков и знаковых систем (главным образом в приложении к естественным и искусственным языкам). Семиотика разделяет знаки на три группы: иконические, индексные и символические. К первой группе относят знаки, похожие на обозначаемый предмет (фотографии, географические карты). Во вторую, индексную, группу включают знаки, либо обозначающие часть предмета, либо такие, где изображение выступает как знак предмета в целом. Наконец, к третьей, символической, группе относят изображения, не имеющие ничего или почти ничего общего с обозначаемым предметом.

Приведем примеры некоторых распространенных жестов и определим их положение в соответствующих семиотических группах. Слово «дом» изображают соприкасающимися концами пальцев обеих рук, представляющих собой как бы двускатную крышу. Такое изображение следует отнести ко второй, индексной, группе. Жест статический. Также статическими являются жесты «женщина» — вертикальная ладонь у одной из щек и «мужчина» — согнутая кисть руки у лба. Первый из этих жестов чистый символ (третья группа), второй иногда относят к индексной группе, если понимать его, как прикосновение пальцев к козырьку фуражки или кепки. Соответственно жестам «женщина» и

«мужчина» динамические жесты «мама» и «папа» производят в первом случае движением вертикальной ладони от одной щеки к другой, а во втором — аналогичным перемещением горизонтальной ладони от лба к подбородку (оба жеста, разумеется, чистые символы). Интернациональный жест «мы вас (тебя) любим» представляется условным символом: поднятые вверх большой, указательный пальцы и мизинец одной из рук. Здесь одно движение заменяет три слова.

2. Существующие системы распознавания жестов

Fingual - это система, способная преобразовывать жесты языка глухонемых в символы на компьютере. Она состоит из компьютера со специализированным программным обеспечением и особой информационной перчатки с постоянными магнитами, закрепленными на кончике каждого пальца. Изменения магнитного поля, возникающие при жестикуляции, регистрируются магнитными датчиками и отправляются на обработку в компьютер. Полученная информация затем анализируется и преобразовывается в символы. Данные о характеристиках магнитного поля отдельных жестов Fingual использует в качестве шаблонов. Это позволяет ускорить процесс распознавания. Основным минус данной системы в том, что набор распознаваемых жестов сильно ограничен, так как перчатка воспринимает только статические жесты. То есть данная разработка не позволяет распознавать язык жестов в целом, а только отдельные символы – дактильный алфавит. Также, несмотря на то, что компоненты перчатки достаточно дешёвы – вряд ли приходится ожидать, что уже в ближайшее время она поступит в производство и станет общедоступной.

Исследователи из китайского отделения Microsoft Research применили технологию распознавания жестов Kinect для перевода жестового языка в текст и живую речь. Их разработка поддерживает, как американский, так и китайский варианты жестового языка для глухих и может переводить и жесты в слова, и слова в жесты, которые демонстрирует на экране виртуальный персонаж. С февраля 2012 года Microsoft Research работает над этим проектом совместно с Академией наук Китая и Пекинским объединенным университетом. Сейчас система умеет распознавать 370 наиболее часто употребляемых слов китайского и американского жестовых языков.

3. Перевод русского жестового языка в текст

На основе рассмотренных систем и методов распознавания жестов предлагается реализовать систему способную переводить русский жестовый язык в текст. Для отслеживания движений (skeleton

tracking) был выбран Microsoft Kinect [Соболев, 2012]. Kinect – бесконтактный сенсорный игровой контроллер, первоначально созданный для консоли Xbox 360, и значительно позднее для персональных компьютеров под управлением ОС Windows. Одним из важных аспектов камеры – диапазон рабочих расстояний, на которых сенсор способен различать объекты.

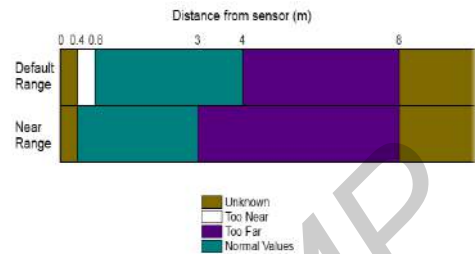


Рисунок 1 – Диапазон рабочих расстояний Kinect

На рисунке 1 представлены два режима работы. Лучшие результаты будут получены при нахождении объектов в рамках второго варианта (Near Range). По skeleton-данным можно получить информацию о координатах (положении) руки человека.

Проектируемая система должна включать в себя следующие подсистемы и предусматривать принцип многомодульности [Заболеева-Зотова и др., 2013а]:

- подсистема работы с пользователем;
- подсистема обработки изображения и вычисления параметров векторной модели;
- подсистема распознавания статического жеста.

Назначение подсистемы работы с пользователем – предоставление пользователю удобного интерфейса работы с системой. Это включает в себя: вывод обработанного изображения на экран, возможность изменять настройки программы, формирование текста из распознанных жестов, вывод сформированного текста на экран, доступ к справочной информации. Также подсистема должна предоставлять возможность работы двум видам пользователей.

Назначение подсистемы обработки изображения и вычисления параметров векторной модели – сегментирование карты глубины с камеры Kinect, обработка сегментированного изображения и вычисление параметров векторной модели по обработанному изображению. 30 раз в секунду на вход подсистеме поступает информация с камеры Kinect. С такой же частотой подсистема обрабатывает входные данные и вычисляет параметры векторной модели кисти руки.

Назначение подсистемы распознавания жеста – отнести данные, получаемые из подсистемы обработки изображения и вычисления параметров векторной модели к одной из дактильных букв или знаку препинания, заложенных в системе. Далее она передаёт значение распознанного жеста в модуль формирования текста подсистемы работы с пользователем (рисунок 2).

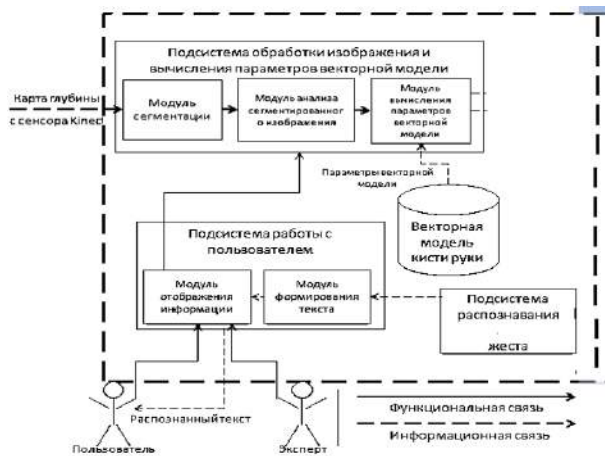


Рисунок 2 – Архитектура системы распознавания жестов

Для реализации распознавания жестов в системе используется визуальные методы распознавания [William, 2008]. Визуальные методы распознавания жестов являются наиболее активной и перспективной областью исследований. По сути, отслеживание голых рук основано на методах, работающих исключительно с визуальной информацией и не предъявляющих специальных требований к оснащению пользователя дополнительным оборудованием. Выделения контуров и силуэтов являются наиболее распространенными характеристиками, используемыми для идентификации позы руки. Пока эти сигналы являются общими и устойчивы к различным условиям освещения и, исходя из этого, требуют больших вычислительно сложных определяющих алгоритмов, которые ищут многомерные позиции пределов руки. В свою очередь визуальные методы распознавания жестов можно разделить на три большие категории. К первой относятся методы, которые основаны на восстановлении полной модели кисти с 27 степенями свободы по входному изображению. В данном подходе, рука представляется как соединение составляющих частей, и каждое состояние называется "степенью свободы". Степени свободы определяют позиции и ориентации реального объекта. Моделирование руки с учетом всех ее степеней свободы представляет собой сложную задачу [Christopher, 2013]. Поэтому используется упрощенная модель руки с 27 степенями свободы, ставшая неофициальным стандартом (рисунок 3).

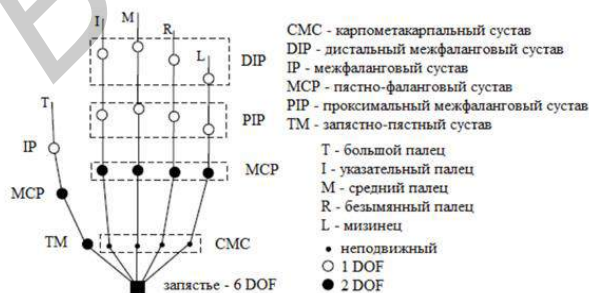


Рисунок 3 – Модель руки с 27 степенями свободы

Ко второй категории относятся статистические методы, которые вместо восстановления полной модели руки предлагают построение признакового описания входного изображения и дальнейшую классификацию жестов именно по этому описанию. Решающую роль в методах подобного рода играет выбор подходящих характеристических признаков. Так для статических поз руки, можно использовать геометрические признаки, такие как кончики пальцев, направление пальцев, контур руки, а также негеометрические признаки (цвет кожи, форма, текстура и др.). Однако эти признаки не всегда доступны или адекватны для распознавания из-за взаимного перекрытия элементов руки и/или из-за условий освещения. Важным моментом в этих подходах является классификатор, который предоставляет возможность разделить (по категориям) различные признаки или информацию, выделенные из "грубых данных", например, изображения или видеопотока.

К третьей категории относятся метрические методы распознавания жестов. Подобные методы предполагают построение некоторой метрики на множестве входных изображений и выполнение классификации за счет сравнения входного изображения с набором эталонов. Так, например, предлагается метрика, характеризующая степень сходства скелетов силуэтов ладони, и выполняется классификация жестов с помощью метода ближайшего соседа [Куракин, 2012].

4. Распознавание кисти

Чтобы реализовать распознавание кисти необходимо решить три подзадачи: детектирование, трекинг и анализ кисти.

На первом этапе детектирования рук на изображении необходимо сегментировать изображение, чтобы найти область интереса - кисть. Для этого используется алгоритм Виолы-Джонса. Обучение классификаторов идет медленно, но сам процесс поиска кисти проходит быстро, именно поэтому был выбран данный метод распознавания кистей рук на изображении. Виола-Джонс является одним из лучших по соотношению показателей эффективность распознавания/скорость работы. Алгоритм даже хорошо работает и распознает кисти под небольшим углом, примерно до 30 градусов. При угле наклона больше 30 градусов процент обнаружений резко падает.

В общем виде, задача обнаружения кисти человека на цифровом изображении выглядит так, имеется изображение, на котором есть искомые объекты [Гонсалес, 2005]. Оно представлено двумерной матрицей пикселей размером $w \times h$, в которой каждый пиксель имеет значение:

- от 0 до 255, если это черно-белое изображение;
- от 0 до 255^3 , если это цветное изображение (компоненты R, G, B).

В результате своей работы, алгоритм должен определить кисть и пометить ее – поиск осуществляется в активной области изображения прямоугольными признаками, с помощью которых и описывается найденная кисть: $rectangle_i = \{x, y, w, h, a\}$, где x, y – координаты центра i -го прямоугольника, w – ширина, h – высота, a – угол наклона прямоугольника к вертикальной оси изображения.

Иными словами, применительно к рисункам и фотографиям используется подход на основе сканирующего окна (scanning window): сканируется изображение окном поиска (так называемое, окно сканирования), а затем применяется классификатор к каждому положению. Система обучения и выбора наиболее значимых признаков полностью автоматизирована и не требует вмешательства человека, поэтому данный подход работает быстро.

Для решения проблемы проведения сложного обучения существует технология бустинга. Бустинг – комплекс методов, способствующих повышению точности аналитических моделей. Бустинг представляет собой жадный алгоритм построения композиции алгоритмов (greedy algorithm) – это алгоритм, который на каждом шагу делает локально наилучший выбор в надежде, что итоговое решение будет оптимальным. Т.е. при бустинге наряду с множествами X и Y вводится вспомогательное множество R , называемое пространством оценок. Далее рассматриваются алгоритмы, имеющие вид суперпозиции $a(x) = C(b(x))$, где функция $b: X \rightarrow R$ называется алгоритмическим оператором, функция $C: R \rightarrow Y$ – решающим правилом. В алгоритме Виолы-Джонса используется пороговое решающее правило, где, как правило, сначала строится оператор при нулевом значении, а затем подбирается значение оптимальное. Процесс последовательного обучения базовых алгоритмов применяется, пожалуй, чаще всего при построении композиций. Для критерия остановки используется критерий точности на обучающей выборке.

Алгоритм бустинга для детектирования кисти:

- 1) определение слабых классификаторов по прямоугольным признакам;
- 2) для каждого перемещения сканирующего окна вычисляется прямоугольный признак на каждом примере;
- 3) выбирается наиболее подходящий порог для каждого признака;
- 4) отбираются лучшие признаки и лучший подходящий порог;
- 5) перевзвешивается выборка.

Каскадная модель сильных классификаторов – это по сути то же дерево принятия решений, где каждый узел дерева построен таким образом, чтобы детектировать почти все интересующие образы и отклонять регионы, не являющиеся образами. Помимо этого, узлы дерева размещены таким образом, что чем ближе узел находится к корню дерева, тем из меньшего количества примитивов он

состоит и тем самым требует меньше времени на принятие решения. Данный вид каскадной модели хорошо подходит для обработки изображений, на которых общее количество детектируемых образов мало.

В этом случае метод может быстрее принять решение о том, что данный регион не содержит образ, и перейти к следующему. Пример каскадной модели сильных классификаторов:

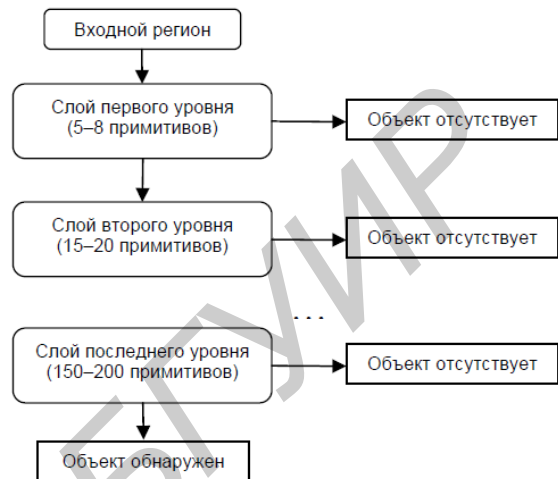


Рисунок 4 – Алгоритм работы каскада классификаторов

Сложность обучения таких каскадов равна $O(xyz)$, где применяется x этапов, y примеров и z признаков.

Далее, каскад применяется к изображению. Работа с «простыми» классификаторами – при этом отбрасывается часть «отрицательных» окон. Положительное значение первого классификатора запускает второй, более приспособленный и так далее. Отрицательное значение классификатора на любом этапе приводит к немедленному переходу к следующему сканирующему окну, старое окно отбрасывается. Цепочка классификаторов становится более сложной, поэтому ошибок становится намного меньше.

Алгоритм трекинга CAMShift был создан Гарри Брадски в 1998 г. и способен отслеживать лица [Местецкий 2002]. Он комбинирует алгоритм отслеживания объекта Mean Shift, основанный на карте вероятности цвета кожи, с адаптивным шагом изменения размера области отслеживания. Вероятность цвета кожи каждого пикселя изображения определяется методом Histogram Backprojection, основанным на цвете, представленном в виде цветового тона (Hue) модели HSV. Так как алгоритм CAMShift способен отслеживать лица на основе вероятности цвета кожи, то он может применяться для отслеживания руки.

Преимуществами данного алгоритма являются: низкие требования к вычислительным ресурсам, гибкие настройки точности позиционирования, возможность работы в различных условиях освещенности. Также дополнительным

преимуществом алгоритма является возможность работы в условиях частичного перекрытия отслеживаемого объекта.

Рассмотрим множество точек (распределение пикселей на гистограммы обратного проецирования). Имеется небольшое окно, которое должно двигаться в сторону области с максимальной вероятностью совпадения пикселей (рисунок 5).

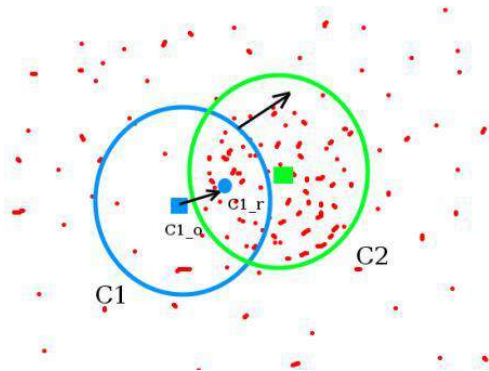


Рисунок 5 – Пример работы алгоритма CAMShift

Алгоритм CAMShift состоит из следующих шагов:

1) Построить гистограмму для канала «Тон» для области изображения с найденной кистью.

2) Вычислить «обратную проекцию» гистограммы. Обратная проекция – это новая гистограмма, в которой показаны вероятности принадлежности пикселя к кисти.

3) Найти новое положение отслеживаемых объектов. Данное действие производится с помощью вероятностей из обратной проекции. Центр смещается к точке с наибольшей концентрацией высоких вероятностей принадлежности пикселя к кисти.

Наибольшая концентрация высоких вероятностей пикселей кисти находится с помощью моментов, которые вычисляются по следующим формулам.

Нулевой момент:

$$Z_{00} = \sum_x \sum_y I(x, y) \quad (1)$$

Моменты первого порядка:

$$\begin{aligned} Z_{10} &= \sum_x \sum_y xI(x, y) \\ Z_{01} &= \sum_x \sum_y yI(x, y) \end{aligned} \quad (2)$$

где $I(x, y)$ - значение вероятности принадлежности пикселя к кисти.

По вычисленным моментам можно найти центр и размер отслеживаемой области:

$$\begin{aligned} x_c &= Z_{10} / Z_{00} \\ y_c &= Z_{01} / Z_{00} \end{aligned} \quad (3)$$

$$s = 2 \sqrt{\frac{Z_{00}}{256}} \quad (4)$$

После нахождения области интереса (кисти), чтобы определить основные характеристики кисти необходимо проанализировать полученные данные. Для анализа кисти необходимо определить контур с помощью детектора границ Кенни и выделить где находятся пальцы с помощью алгоритма k-изгиба.

Детектор границ Кенни состоит из пяти отдельных шагов [Грин, 2014]:

1) Сглаживание. Фильтр Гаусса. Размытие изображения для удаления шума.

2) Поиск градиентов. Границы отмечаются там, где градиент изображения приобретает максимальное значение.

3) Подавление не-максимумов. Только локальные максимумы отмечаются как границы.

4) Двойная пороговая фильтрация. Потенциальные границы определяются порогами.

5) Трассировка области неоднозначности. Итоговые границы определяются путём подавления всех краёв, несвязанных с определенными (сильными) границами.

Перед применением детектора, преобразуем изображение в оттенки серого, чтобы уменьшить вычислительные затраты, иначе быстрого преобразования не получится.

После определения границ необходимо найти пальцы на изображении. Контур кисти является достаточно изменчивой характеристикой, поэтому необходимо выделить более постоянные характеристики, например центр ладони, а также концы пальцев кисти.

Центр ладони находится как центр окружности с максимальным радиусом, вписанной в ладонь. Далее от центра до каждой точки границы ладони считается евклидово расстояние. Такое представление кисти хорошо описывает ее топологические свойства. Для нахождения концов пальцев кисти применим алгоритм k-изгиба [Grauman, 2004].

На первом шаге алгоритма вычисляется минимально выпуклая оболочка (МВО) контура руки. Задача поиска МВО решается с помощью алгоритма Грэхема [Graham, 1972]. Алгоритм решает данную задачу с помощью стека, сформированного из точек кандидатов. Все точки входного множества заносятся в стек, а затем точки, которые не являются вершинами постепенно удаляются из него. По завершении работы алгоритма в стеке остаются только вершины.

В результате определяется кисть человека и ее характеристики (ключевые точки, углы между

векторами). На основе этих данных можно распознавать движения и жесты.

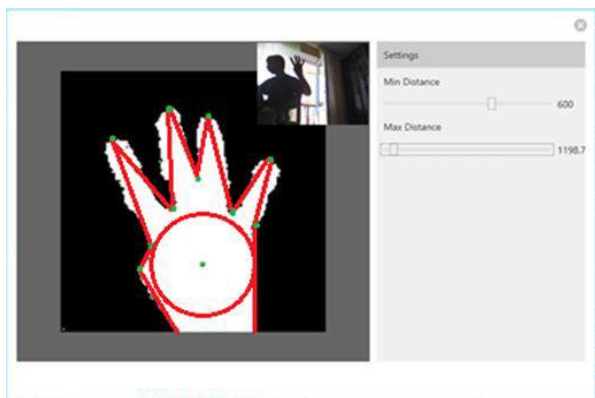


Рисунок 6 – Распознавание четырех пальцев

Заключение

Автоматизированный перевод русского жестового языка в текст играет важную роль в жизни людей с ограниченными возможностями. Создание подобных приложений имеет важное социальное значение и должно помочь адаптации таких людей. Кроме того, разрабатываемые подходы имеют и практическое значение, поскольку жестовые интерфейсы получают все большее распространение и развитие.

В виде отдельного программного приложения реализованы алгоритмы детектирования, трекинга и анализа кисти. Для обучения при детектировании руки используется технология бустинга. На данный момент улучшается производительность и скорость работы разработанных алгоритмов детектирования движений рук [Розалиев и др., 2013] и строится система определения по движениям смысла жеста [Розалиев и др., 2014] и его эмоциональной оценки [Заболеева-Зотова и др., 2013b].

Работа частично поддержана Российским фондом фундаментальных исследований (проекты 13-07-00459, 13-07-97042, 14-07-97016, 15-07-06322).

Библиографический список

- [Гонсалес, 2005] Гонсалес, Р. Цифровая обработка изображений: пер. с англ. / Р. Гонсалес, Р. Вудс.-Москва: 2005.-1072 с.
- [Грин, 2014] Грин, Б. Алгоритм выделения контуров Кенни [Электронный ресурс] / Б. Грин. - Режим доступа: http://www.pages.drexel.edu/~weg22/can_tut.html.
- [Заболеева-Зотова и др., 2011] Заболеева-Зотова А. В., Орлова Ю. А., Розалиев В. Л., Бобков А. С. Применение нечетких темпоральных высказываний для описания движений при эмоциональных реакциях. – Открытое образование, 2011.
- [Заболеева-Зотова и др., 2013а] Formalization of initial stage of designing multi-component software / Заболеева-Зотова А.В., Орлова Ю.А., Розалиев В.Л., Фоменков С.А., Петровский А.Б. // Multi Conference on Computer Science and Information Systems 2013 (Prague, Czech Republic, July 23-26, 2013) : Proceedings of the IADIS International Conference Intelligent Systems and Agents 2013 / IADIS (International Association for Development of the Information Society). – [Prague], 2013. – P. 107-111.
- [Заболеева-Зотова и др., 2013б] Automated identification of human emotions based on analysis of body movements / Заболеева-Зотова А.В., Бобков А.С., Орлова Ю.А., Розалиев В.Л., Половинкин А.И. // Multi Conference on Computer Science and

Information Systems 2013 (Prague, Czech Republic, July 23-26, 2013) : Proceedings of the IADIS International Conferences Interfaces and Human Computer Interaction and Game and Entertainment Technologies 2013 / IADIS (International Association for Development of the Information Society). – [Prague], 2013. – P. 299-304.

[Куракин, 2013] Куракин А.В. Распознавание динамических жестов в системе компьютерного зрения на основе медиального представления формы изображений // Московский государственный университет имени М. В. Ломоносова. – 2012.

[Местецкий 2002] Местецкий, Л. М. Математические методы распознавания образов / Л. М. Местецкий.-Москва: МГУ, 2002–2004.- С. 42 – 44

[Розалиев и др., 2013] Автоматизация построения векторной модели тела человека / Розалиев В.Л., Орлова Ю.А., Шпирко А.А., Дорофеев Н.С. // Электротехнические и информационные комплексы и системы. - 2013. - № 2, т. 9. - С. 102-106.

[Розалиев и др., 2014] Розалиев, В.Л. Applying the automated system for the determination of emotions in the education's tasks for people with disabilities / Розалиев В.Л., Орлова Ю.А. // Innovation Information Technologie : mater. of the 3rd Int. scien.-pract. conf. (Prague, April 21-25, 2014). Part 2 / МИЭМ ВШЭ, Рос. центр науки и культуры в Праге. - М., 2014. - С. 446-452.

[Соболев, 2012] Соболев Е.Г. Средство интерактивного взаимодействия реального и виртуального пространств на основе трехмерного сканирования объектов с использованием платформы Microsoft Kinect. – Москва, 2012.

[Christopher, 2013] Christopher Kawatsu. Development of a Fall Detection System with Microsoft Kinect / Christopher Kawatsu, Jiaying Li, C. J. Chung // Robot Intelligence Technology and Applications 2012, 2013.

[Grauman, 2004] K. Grauman and T. Darrell. Fast contour matching using approximate earth mover's distance. In CVRP, 2004.

[Graham, 1972] Graham, R. L. An efficient algorithm for determining the convex hull of a finite planar set / R. L. Graham.- Information Processing Letters, 1972.- 7:175-180.

[William, 2008] William L. Johnson, Michael Feuerstein. An Interpreter's Interpretation: Sign Language Interpreters' View of Musculoskeletal Disorders / William L. Johnson, Michael Feuerstein // Journal of Occupational Rehabilitation, September 2008.

AUTOMATED ALLOCATION OF THE HANDS OF THE PERSON TO RECOGNIZE SIGN LANGUAGE

Rozaliev V.L., Agafonov G.V., Kirichenko M.I.

*Volgograd State Technical University,
Volgograd, Russia*

vladimir.rozaliev@gmail.com

This article is devoted to the problem of detecting and analyzing brushes with the camera depth. Special attention is given to the analysis and tracking of the brush. In the work of the combined algorithms for searching, tracking and analysis of the hands of man. Reviewed existing systems for gesture recognition. Also in the paper highlighted the advantages and disadvantages of visual methods.

The aim of the study was to finding and rapid analysis of the hands of man. To achieve this goal were as follows: conduct a review of existing approaches and software for the detection and analysis of the hands; develop a method for the detection of hands on the depth image; develop a method for analyzing the brush; implement the methods and algorithms in software.