Министерство образования Республики Беларусь
Учреждение образования
Белорусский государственный университет информатики и радиоэлектроники

УДК 004.93:004.032.26

Арутюнова
Таисия Арсеновна

Метод Виолы-Джонса для распознавания жестов глухонемых

АВТОРЕФЕРАТ
на соискание степени магистра технических наук

по специальности 1-40 80 02 – Системный анализ, управление и обработка информации

Научный руководитель

Муха Владимир Степанович
доктор технических наук, профессор

Минск 2020

ВВЕДЕНИЕ

В настоящее время в сфере информационных технологий успешно развивается направление, связанное с интеллектуализацией методов обработки и анализа данных. Задача распознавания лиц и жестов послужила стимулом для развития теории распознавания объектов. С появлением новых понятий, таких как «виртуальная среда», «перцепционный пользовательский интерфейс» и т.д. потребовалась разработка более мощных и удобных способов взаимодействия человека с компьютерной системой [1].

В качестве одного из способов обеспечения комфортного взаимодействия с компьютером, человеческая рука может быть использована в качестве интерфейса ввода. Жесты являются мощным каналом связи, который формирует основную часть передачи информации в нашей повседневной жизни.

В последние годы появилась и начала быстро развиваться тенденция использования жестов, особенно жестов руки, как способа взаимодействия с компьютерной системой. Распознавание жестов, таким образом, стало важнейшей частью в человеко-машинной интеллектуальной интеракции и начало привлекать множество исследователей.

Жесты являются неотъемлемой частью человеческого общения, причём для глухонемых и слабослышащих они служат единственным методом оперативного обмена информацией и обучения. Например, для распознавания жестов ручной азбуки *ASL* используется трёхмерная камера *Microsoft Kinect*. Для использования человеческой руки применяются перчатки данных, такие как Киберперчатка (*CyberGlove*), окрашенные перчатки. Однако перчатка данных и прилагаемые к ней провода являются неудобными для практического применения пользователями. Стоимость устройств часто слишком дорога для регулярных пользователей.

Видеокамера представляет собой недорогое и удобное устройство ввода информации, которое может служить эффективным каналом связи при реализации человеко-машинного взаимодействия.

В данной работе рассматривается задача распознавания жестов на цифровых изображениях и видеопоследовательности в режиме реального времени. Эта задача выдвигается в связи актуальностью приобщения к информационным технологиям особой части пользователей с нарушениями слухового и речевого аппаратов. Решение данной задачи находится в области применения метода Виолы-Джонса для создания алгоритма, позволяющего на цифровых изображениях и видеопоследовательности в режиме реального времени распознать необходимые объекты, в первую очередь – жесты, являющиеся единицами языковой системы для данной целевой аудитории.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Цель и задачи исследования

Целью данной работы являются создание нового алгоритма, основанного на применении вейвлет-преобразования и метода главных компонент для распознавания жестов на цифровых изображениях, и разработка оригинального комплексного алгоритма, основанного на применении метода Виолы – Джонса, алгоритма *CAMShift*, вейвлет-преобразования и метода главных компонент для распознавания жестов на видеопоследовательности в режиме реального времени.

Для достижения поставленной цели необходимо решить следующие задачи:

1. Разработать алгоритм распознавания поз руки, способный функционировать в режиме реального времени и инвариантный к преобразованиям и изменению освещения.
2. Разработать алгоритм распознавания движения руки в видеопотоке, обеспечивающий возможность распознавания сложных и деформированных траекторий.
3. Разработать алгоритм распознавания жестов руки на основе предложенных алгоритмов распознавания поз и движения руки, позволяющий распознавать автономные и интерактивные жесты на видеопоследовательностях в режиме реального времени.

Связь работы с приоритетными направлениями научных исследований и запросами реального сектора экономики

Научную новизну полученных в диссертации результатов определяют следующие положения.

1. Предложен алгоритм распознавания поз руки на основе SURF-дескрипторов, алгоритма k-средних и многослойной нейронной сети, предназначенный для распознавания статической компоненты жестов и отличающийся от других способностью функционировать в режиме реального времени, устойчивостью к различным аффинным преобразованиям, изменению освещения, и, частично, к шумам, при обеспечении точности распознавания в пределах 90-98%.
2. Предложен алгоритм распознавания движения руки в видеопотоке на основе нейронной сети, предназначенный для распознавания динамической компоненты жестов в режиме реального времени. В основе алгоритма лежит

идея упрощения и передискретизации траектории, полученной после трекинга, что обеспечивает возможность распознавания сложных деформированных траекторий с точностью выше 96% в реальных условиях применения.

3. Разработан алгоритм распознавания жестов на основе детектора Джонса-Виолы, трекера *CAM-Shift*, предложенных алгоритмов распознавания поз и движения руки, позволяющий распознавать жесты на видеопоследовательностях в режиме реального времени. Особенностью предложенного алгоритма является сочетание возможности распознавания интерактивных и автономных жестов благодаря разбиению жестов на статическую компоненту (позу) и динамическую компоненту (движение руки).

Научную ценность работы представляет вклад в развитие области распознавания объектов и человеко-машинного взаимодействия, заключающийся в предложенном алгоритме распознавания статических поз руки, позволяющем распознавать формы руки с высокой точностью в реальном времени при обеспечении устойчивости к разным типам искажения внешнего вида входного объекта, и частично, к шумам; в оригинальном алгоритме распознавания движения руки с использованием нейронной сети, реализация которого, вместе с алгоритмом распознавания формы руки, дает полноценное описание жестов руки человека для цели управления компьютером; в оригинальном комплексном алгоритме распознавания жестов, с помощью которого построена программная система для управления компьютером с использованием жестов.

Личный вклад соискателя

Результаты, приведенные в диссертации, получены соискателем лично. Вклад научного руководителя В. С. Мухи, заключается в формулировке целей и задач исследования.

Апробация результатов диссертации

Основные положения диссертационной работы докладывались и обсуждались на 56-ой научно-технической конференции аспирантов, магистрантов и студентов БГУИР (Минск, Беларусь, 2000)

Опубликованность результатов диссертации

По теме диссертации опубликована 1 печатная работа по материалам доклада на 56-ой научно-технической конференции аспирантов, магистрантов и студентов БГУИР.

Структура и объем диссертации

Диссертация состоит из введения, общей характеристики работы, трех глав, заключения, списка использованных источников, списка публикаций автора. В первой главе представлен анализ научной литературы, выявлены основные существующие проблемы в рамках тематики исследования. Вторая глава посвящена описанию использованных методов. В третьей главе предложены методы алгоритм распознавания поз руки, алгоритм распознавания движения руки в видеопотоке, алгоритм распознавания жестов руки на основе предложенных алгоритмов распознавания поз и движения руки, позволяющий распознавать автономные и интерактивные жесты на видеопоследовательностях в режиме реального времени.

Общий объем работы составляет 70 страниц, из которых основного текста – 45 страниц, 27 рисунков на 13 страницах, список использованных источников из 33 наименований на 3 страницах.

1 АНАЛИЗ НАУЧНОЙ ЛИТЕРАТУРЫ

За последние годы распознавание различных по своей природе объектов находит применение во многих сферах человеческой деятельности. Так, например, на современном этапе развития информационно-коммуникационных технологий широко используется распознавание лиц для идентификации личности.

Среди различных подходов к решению задачи распознавания объектов, распознавание жестов на основе компьютерного зрения оказывается доминантной тенденцией благодаря новым достижениям в области компьютерного зрения, повышенной производительности компьютеров, и также популярности и высокого качества недорогих видеокамер. При этом важным является тот факт, что системы распознавания жестов на основе компьютерного зрения обеспечивают более интуитивный и натуральный канал взаимодействия человека с компьютером. Перспективность данного направления подтверждается результатами исследований авторов Kolsch M., Turk M., Lienhard R., Maydt J., Rittscher J., Blake A., Bradski G., Viola P., Jones M., IsardM., Davis J., Bobick A., Comaniciu D.

Самое общее понятие жеста предложил *Kurtenback* и *Hulteen* [9]: жестами являются действия руки и/или части тела, которые несут информацию. Несмотря на различия жестов в разных языковых культурах, в зависимости от своей функции, жесты могут быть сгруппированы в категории:

- Семиотические жесты (*semiotic*): жесты для передачи значимой информации;

- Эрготические жесты (*ergotic*): жесты для манипуляции физическими объектами и создания артефактов;

- Эпистемологические жесты (*epistemic*): жесты для изучения с помощью тактильного обследования. В исследованиях в области человеко-машинного взаимодействия, особое внимание уделяется семиотическим жестам. Данная группа жестов делится на следующие подгруппы [16]:

- Символические жесты (*symbolic*): жесты, которые имеют единственное определенное значение.

- Дейктические жесты (*deictic*): это тип жестов указания руки, которые чаще всего встречаются в человеко-машинном взаимодействии.

- Иконические жесты (*iconic*): жесты для передачи информации о размерах, форме, ориентации и т.д. объекта.

- Пантомимические жесты (*pantomimic*): жесты, используемые для демонстрации движения объекта.

Компьютерные системы могут «понимать» натуральные человеческие жесты с помощью распознавания жестов – процесса обработки и преобразования данных для описания жестов человека, используя математический аппарат. Аппарат распознавания жестов позволяет создать так называемые «интерфейсы на основе жестов» (*gesture-based interfaces*), в которых взаимодействие человека с компьютером осуществляется с помощью жестов. Устройствами ввода для такого типа интерфейса (и так же для используемого аппарата распознавания жестов) могут быть специальные перчатки или маркеры, инфракрасные сенсоры, трехмерные камеры, стереокамеры, обычные видео камеры. В зависимости от типа устройства ввода, методы, алгоритмы, и способы для решения задачи распознавания жестов принадлежат одному из следующих направлений [17]:

- Методы с использованием устройств, работающих вне спектра видимого света (тепловые сенсоры, инфракрасные камеры и т.д.).

- «Активные методы», которые требуют активного проецирования света.

- «Инвазивные методы», которые требуют модификации или изменения среды (например, ношение специальных перчаток или цветовых маркеров).

- Методы на основе компьютерного зрения (*vision-based*), в которых жесты наблюдаются и записываются с использованием видео камеры.

Распознавание жестов требует описывать не только пространственные, но и временные признаки. Распознавание некоторых жестов возможно на основе использования двумерных данных о положении руки. Одним из фундаментальных признаков данного типа является «блоб». *Wren* и др. [33] применили в своей системе «*Pfinder*» мульти-классовую статистическую модель цветов и форм, чтобы получить двумерное представление руки в широком диапазоне наблюдения.

2 ОПИСАНИЕ ИСПОЛЬЗУЕМЫХ МЕТОДОВ

Метод Виола-Джонса был разработан и представлен в 2001 г. Полом Виолой и Майклом Джонсом и до сих пор эффективен для поиска объектов на цифровых изображениях и видеопоследовательностях в режиме реального времени [1, 2]. Основной его идеей является использование каскада простых классификаторов – детекторов характеристик вместо одного сложного классификатора. На базе этой идеи возможно построение детектора, способного работать в режиме реального времени. Данный метод в общем виде ищет объекты по общему принципу сканирующего окна.

В общем виде, задача обнаружения объекта и особенностей объекта на цифровом изображении выглядит именно так:

- имеется изображение, на котором есть искомые объекты. Оно представлено двумерной матрицей пикселей размером *w*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 объекты и их особенности и пометить их – поиск осуществляется в активной области изображения прямоугольными признаками, с помощью которых и описывается найденное лицо и его черты(формула 2.1):

$$rectangle_i = \{x,y,w,h,a\}, \tag{2.1}$$

где *x, y* – координаты центра *i*-го прямоугольника,

w – ширина,

h – высота,

a – угол наклона прямоугольника к вертикальной оси изображения.

Задача поиска и нахождения объектов на изображении с помощью данного принципа часто бывает очередным шагом на пути к распознаванию характерных черт, к примеру, определение жеста глухонемого человека по распознанному положению руки и особенностей значения положения руки.

Для того чтобы рассчитывать яркость прямоугольного участка изображения, используется интегральное представление [4]. По интегральной матрице можно очень быстро вычислить сумму пикселей произвольного прямоугольника.

Виола и Джонс адаптировали идею использования вейвлетов Хаара и разработали то, что было названо признаками Хаара. Признак Хаара состоит

из смежных прямоугольных областей. Они позиционируются на изображении, далее суммируются интенсивности пикселей в областях, после чего вычисляется разность между суммами. Эта разность и будет значением определенного признака, определенного размера, определенным образом спозиционированного на изображении.

В контексте алгоритма, имеется множество объектов (изображений), разделённых некоторым образом на классы. Требуется построить алгоритм, способный классифицировать произвольный объект из исходного множества.

Постановка классификации выглядит следующим образом:

- есть X – множество, в котором хранится описание объектов, Y – конечное множество номеров, принадлежащих классам;

- между ними есть зависимость – отображение $Y^*: X => Y$. Обучающая выборка представлена $Xm = \{(x1,y1), \ldots, (xm,ym)\}$;

- конструируется функция f от вектора признаков X, которая выдает ответ для любого возможного наблюдения X и способна классифицировать объект $x \in X$.

Бустинг – комплекс методов, способствующих повышению точности аналитических моделей. В основе такой идеи лежит построение цепочки (ансамбля) классификаторов, который называется каскадом, каждый из которых (кроме первого) обучается на ошибках предыдущего.

Развитием данного подхода явилась разработка более совершенного семейства алгоритмов бустинга *AdaBoost* (адаптивное улучшение). Усиление простых классификаторов - подход к решению задачи классификации (распознавания), путём комбинирования примитивных «слабых» классификаторов в один «сильный».

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.

Вейвлет-преобразование широко используется для анализа нестационарных процессов. Оно показало свою эффективность для решения широкого класса задач, связанных с обработкой изображения. Коэффициенты вейвлет-преобразования содержат информацию об анализируемом процессе и используемом вейвлете.

Метод главных компонент (*Principal Component Analysis*, *PCA*) – один из наиболее распространенных методов для уменьшения размерности данных, потери наименьшего количества информации.

2 АРХИТЕКТУРА КОМПЛЕКСНОГО АЛГОРИТМА РАСПОЗНАВАНИЯ ЖЕСТОВ

В диссертационной работе предлагается двухуровневая архитектура для комплексного алгоритма распознавания жестов. Первый уровень включает шаги получения последовательных кадров из видеокамеры, предобработки полученных кадров, и обнаружение руки на видеокадре. На втором уровне выполняется слежения за рукой во времени, распознавание позы, и распознавание глобального движения. На втором уровне также возможен выбор режима распознавания автономных или интерактивных жестов. Общая схема алгоритма приведена на рисунке 3.1.

Первый уровень предназначен для обнаружения присутствия руки в области видимости видеокамеры и для инициализации работы алгоритмов распознавания и трекинга второго уровня. Обычная видеокамера записывает в одну секунду от 15 до 30 кадров размера 640*x*480 пикселей, и в каждый момент времени видеосистема доставляет только один кадр для обработки. Система распознавания жестов в реальном времени должна обрабатывать видеокадры по мере их поступления с видеосистемы. Таким образом, время обработки каждого кадра должно быть не более 66 миллисекунд, чтобы не было задержки в работе системы. Для сокращения времени обработки, размер кадров уменьшается до 320*x*240 пикселей на шаге предобработки. На этом шаге также происходит преобразование полученных цветных кадров в полутоновые изображения, т.к. в дальнейшем алгоритмы обработки работают только с полутоновыми изображениями.

Детектор руки выполняет поиск присутствия руки на каждом поступающем видеокадре. Если на видеокадре не обнаруживается рука, то кадр отменяется и в обработку запускается следующий видеокадр с видеосистемы. Если рука присутствует на кадре, то детектор руки отключается. Целью шага обнаружения руки является инициализация работы алгоритмов трекинга второго уровня путем определения прямоугольной области кадра, в которой находится рука. Данная область кадра будет использоваться как модель объекта для алгоритма трекинга.

Для распознавания автономных жестов, на втором уровне применяется распознавание позы и связанного с ней глобального движения. После того, как положение руки определяется с помощью детектора руки (на первом уровне), прямоугольная область, где находится рука, сохраняется и передается в алгоритм трекинга, а сам детектор отключается.

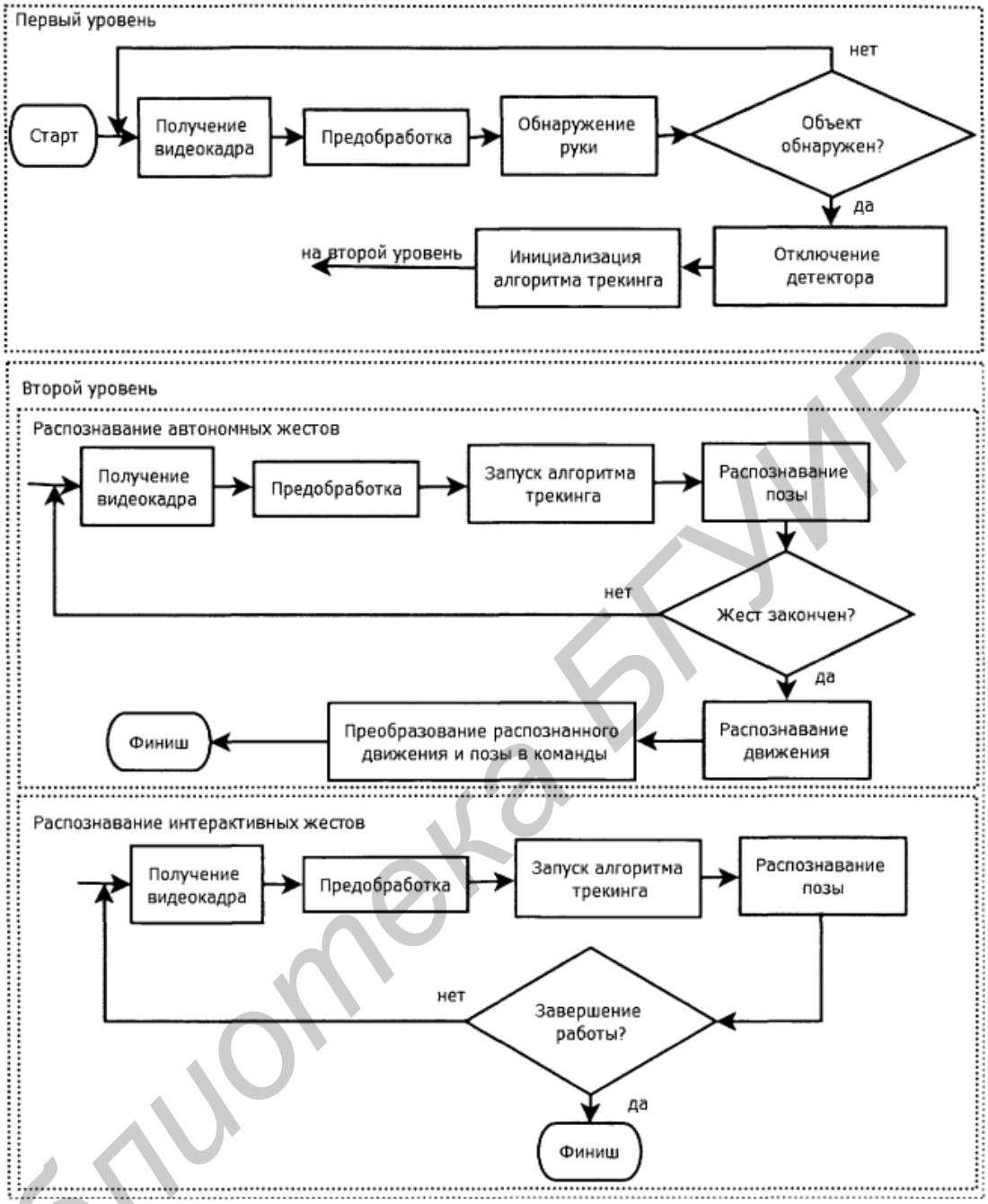

Рис. 3.1 Схема комплексного алгоритма распознавания жестов

Алгоритм трекинга анализирует эту область для создания модели объекта и начинает процесс трекинга. Поступающие видеокадры с видеосистемы затем передаются сразу на второй уровень. После предобработки (уменьшения размера и преобразования кадра в полутоновое изображение), механизм трекинга запускается для поиска местоположения руки на новом видеокадре. Алгоритм распознавания поз затем работает только с областью кадра, где находится рука, а не с целым кадром, и, таким образом, может обеспечить высокую скорость распознавания, независимо от реального размера видеокадра. Если жест не завершен, новый кадр

пропускается на обработку, иначе алгоритм распознавания движения запускается для распознавания полученного жеста.

Условием завершения жеста принимается отсутствие руки на кадре, например, когда рука двигается за пределом области видимости камеры и алгоритм трекинга не может определить местоположения руки на кадре. Распознанный жест затем может преобразовываться в команды для работы с компьютерной системой. Данный режим может использоваться в системе взаимодействия на основе жестов для выполнения команд, функции которых аналогичны горячим клавишам.

Для распознавания интерактивных жестов для прямой манипуляции, на втором уровне применяется распознавание позы и алгоритм трекинга. При этом трекинг также отвечает за наблюдение за положением руки на каждом кадре. Распознавание позы позволяет выполнить команды, такие как щелчок мыши. Данный режим распознавания предназначается для непрерывной работы с компьютерной системой, такой как удаленное управление курсором мыши, рисование с помощью жестов, и т.д. Условием завершения работы является отсутствие руки на кадре, когда рука двигается за пределом области видимости камеры.

В диссертационной работе, обнаружение руки выполняется с помощью известного метода Джонса-Виолы. Для слежения за положением руки применяется метод *CAM-Shift* на основе использования цветовой информации. Для задачи распознавания позы руки предложен и реализован алгоритм на основе многослойной нейронной сети и *SURF*-дескрипторов. Анализ и распознавание движения руки осуществляется с использованием созданного в работе классификатора на основе нейронной сети.

Начальное положение руки на видеокадре определяется на этапе обнаружения с помощью детектора Джонса-Виолы. В процессе трекинга, траектория движения руки записывается для дальнейшей обработки в алгоритме распознавания движения. Для решения этой задачи был разработан быстрый алгоритм распознавания траектории движения руки на основе нейронной сети. Алгоритм распознавания траектории движения руки состоит из следующих этапов:

1) *Упрощение и сглаживание траектории движения*. В процессе трекинга, траектория движения руки записывается в виде массива точек. Однако, при перемещении руки, человек часто делает случайные движения малой амплитуды, что приводит к не «гладкой» траектории движения руки. Таким образом, в массиве данных, полученном после окончания движения, всегда присутствует какой-то шум. Кроме того, количество точек для каждого движения достаточно большое, что усложняет обработку. Для

решения этой проблемы, применяется алгоритм Рамера-Дугласа-Пекера (*Ramer-Douglas-Peucker*), позволяющий "сгладить" траекторию.

2) *Передискретизация и преобразование траектории*. Количество точек в траектории движения руки не фиксировано, а меняется в зависимости от скорости перемещения руки и от скорости обработки данных программой. Механизм распознавания всегда требует, чтобы число входов было фиксировано. Повторная дискретизация (*resampling*) используется для фиксирования количества точек траектории. На данном этапе также выполняется масштабирование и перемещение траектории для повышения эффективности дескрипторов, вычисленных на следующем шаге.

3) *Вычисление дескриптора*. Массив точек траектории преобразуется в массив векторов наклона и выполняется вычисление синуса или косинуса углов наклона. Массив синусов (или косинусов) в дальнейшем будет служить входом для классификатора.

4) *Обучение и распознавание в нейронной сети*. Для распознавания формы траектории движения руки в качестве классификатора применяется многослойная нейронная сеть с обратным распространением ошибок.

ЗАКЛЮЧЕНИЕ

В диссертационном исследовании для усовершенствования взаимодействия пользователей с расширенными потребностями (с нарушениями слухового и речевого аппаратов) при помощи компьютерных технологий была проведена работа по созданию алгоритма распознавания жестов на цифровых изображениях и в видеопоследовательности.

Представлен комплексный алгоритм распознавания жестов на видеопоследовательностях в реальном времени, который может работать в режиме автономного распознавания и в режиме интерактивного распознавания жестов.

Предложена двухуровневая архитектура для комплексного алгоритма распознавания жестов, содержащая на первом уровне шаги получения последовательных кадров из видеокамеры, предобработки полученных кадров, и обнаружение руки на видеокадре. На втором уровне выполняется слежение за рукой во времени, распознавание позы и распознавание глобального движения.

Предложено применение алгоритма Джонса-Виолы для обнаружения руки в видеопотоке с возможностью функционирования в реальном времени. Алгоритм работает на основе признаков Хаара, интегрального изображения, и каскадного *AdaBoost* классификатора.

Изложен метод *CAM-Shift* для трекинга руки на основе использования цветовой информации кожи.

Из проведённых экспериментов видно, что средняя точность распознавания цифр составила 93,7%, а полнота — 89,6%. Эти же показатели для распознавания букв составили соответственно 94,08% и 88,2%.

Существующие алгоритмы при проведении тестирования распознавания жестов показывают недостаточный объем распознавания, что объясняется существенным различием форм и цвета кожи руки людей, показывающих этот жест.

Решению данной проблемы в значительной степени помогает предложенный в данном исследовании алгоритм, направленный на совершенствование метода распознавания.

В ходе выполнения данной научной работы был создан алгоритм распознавания жестов в видеопотоке на основе детектора Джонса-Виолы, трекера CAM-Shift, разработанных алгоритмов распознавания поз и движения руки

Список публикаций соискателя

Арутюнова, Т.А. Метод Виолы-Джонса для распознавания жестов глухонемых / Т.А Арутюнова // Материалы 56-ой научной конференции аспирантов, магистрантов и студентов. – Минск: БГУИР, 2020. – с. 24-26.