Научная статья
Original article
УДК 004.65

РАСПОЗНАВАНИЕ ЖЕСТОВ НА ВИДЕО

VIDEO GESTURE RECOGNITION

Коломыцев Александр Сергеевич, студент Балтийский федеральный университет Им. И. Канта (236041 Россия, г. Калиниград, ул. Александра невского, д. 14), тел. 8(963)268-55-54, tzhz@mail.ru

Вердиев Орхан Ровшанович, студент Балтийский федеральный университет Им. И. Канта (236041 Россия, г. Калиниград, ул. Александра невского, д. 14), тел. 8(913)039-86-98, orhan-rv@mail.ru

Kolomytsev Alexander Sergeevich, student of the Baltic Federal University. I. Kant (236041 Russia, Kaliningrad, Alexander Nevsky St., 14), tel. 8(963)268-55-54, tzhz@mail.ru

Verdiev Orkhan Rovshanovich, student of the Baltic Federal University. I. Kant (236041 Russia, Kaliningrad, Alexander Nevsky St., 14), tel. 8(913)039-86-98, orhan-rv@mail.ru

Аннотация. В современном мире, в век развития информационных технологий, где искусственный интеллект уже не кажется научной фантастикой, мы являемся свидетелями грандиозных событий, большинство из которых еще не наступили, но уже прорастают сквозь асфальт настоящего и благодаря которому появилось понятие компьютерного зрения. В статье мы

разберем методы распознавания жестов, рассмотрим технологии и проведем эксперимент.

Abstract. In the modern world, in the age of information technology development, where artificial intelligence no longer seems like science fiction, we are witnessing grandiose events, most of which have not yet come, but are already growing through the asphalt of the present and thanks to which the concept of computer vision has appeared. In the article, we will analyze the methods of gesture recognition, consider technologies and conduct an experiment.

Ключевые слова. Жесты, компьютерное зрение, нейросети, алгоритмы обучения, метод Виолы-Джонса, волновой алгоритм, дискрипторы фурье, опорные вектора, глубокое обучение.

Keywords. Gestures, computer vision, neural networks, learning algorithms, Viola-Jones method, wave algorithm, Fourier descriptors, support vectors, deep learning.

ВВЕДЕНИЕ

Жесты – один из методов передачи информации. Все, включая мимику, жесты рук и позы тела являются эффективными передатчиками в человеко-человеческом общении. Выделяют три типа жестов:

- жесты тела, включающих действия или движения всего тела;
- жесты рук и кистей, а также позы рук;
- головные и лицевые жесты, в состав которых входят кивания или вращения головы, подмигивая, губы.

Распознавание жестов есть математическая интерпретация человеческих движений вычислительным устройством. Чтобы иметь возможность взаимодействия с человеком, роботы должны корректно распознавать человеческие жесты и действовать в соответствии с жестом в достаточной степени точности.

Компьютерное зрение — это теория и технология создания машин, которые могут производить обнаружение, отслеживание и классификацию объектов. Для того, чтобы взаимодействие машины и человека проходило наиболее комфортно и удобно для последнего, необходимо создать технологию, способную воспринимать сигналы, посылаемые ей человеком, будь то голосовые команды, определённые невербальные жесты или сигналы мимики лица (проявление эмоций и чувств). С каждым годом люди все чаще и чаще начинают общаться не словами, а жестами. Многие домашние приборы, машины и телефоны используют данную технологию с целью начала работы или передачи информации на другие устройства. Технология распознавания жестов является одной из наиболее сложных и актуальных в области обработки изображений.

Возможность распознавать форму и движение рук может стать большим шагом к улучшению пользовательского опыта для большинства современных приложений. Одной из самых сложных задач является разработка устойчивой легковесной модели для распознавания рук в реальном времени.

В июне 2019 года исследователями из Google был предложен подход, который использует нейросетевые модели для предсказания расположения 21 3D точки руки на основе одного кадра.

Ниже приведена диаграмма, на которой показанно в каких журналах и на каких конференциях, публикуются наиболее важные документы, связанные с распознаванием жестов за последний год.

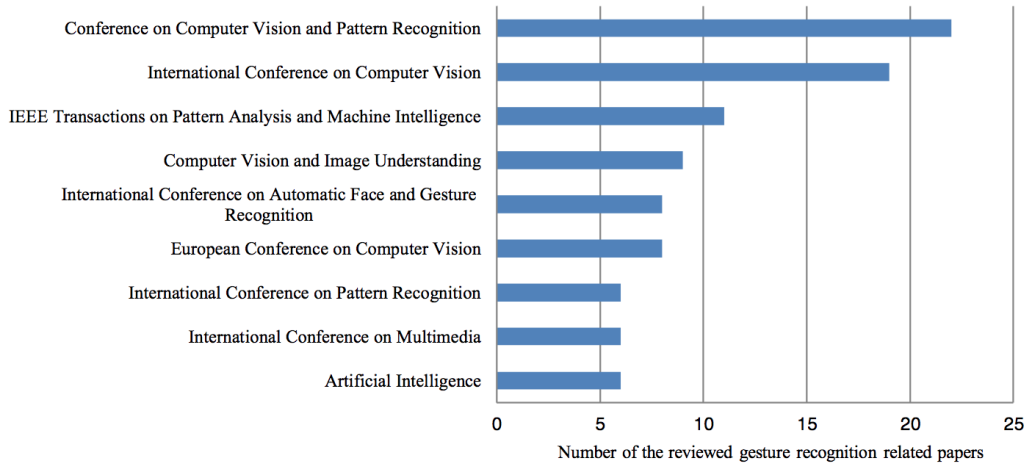

Рисунок 1.

Анализ существующих алгоритмов распознавания жестов.

Прочитав и изучив большое количество материала по данной теме, можно сделать вывод, что все алгоритмы для распознавания жестов просты в реализации, но большинство из данных алгоритмов обладают низкой эффективностью и очень зависит от внешних факторов, таких как:

- освещение;
- качество камеры;
- фон;
- размер объекта.

Чтобы повысить эффективность применения данных алгоритмов их нужно рассматривать в совокупности в то время как многие приложения в современном мире этого не делают.

За последние пять лет алгоритмы распознавания жестов на видео уже имеют более высокое преимущество, чем раньше, но до сих пор их самым главным недостатком является высокая вычислительная сложность и необходимость соблюдения определенных условий. Все современные алгоритмы должны сочетаться с методами обнаружения, для отслеживания корректного отслеживания объектов на видеопотоке. Для всех методов

распознавания существуют предпочтительные качественные характеристики, такие как:

- точность распознавания;
- скорость работы.

Рассмотрим алгоритм распознавания жестов в котором используются методы машинного обучения [21]. Потому что все данные получаемые с камер имеют высокий уровень шумов и все другие алгоритмы будут менее эффективными, если не использовать в совокупности с ними особых условий, методов или оборудования.

Изучив все компоненты и стадии прохождения видео, была создана схема алгоритма по распознаванию жестов на видеопотоке. Данная схема продемонстрирована на Рисунке 2.

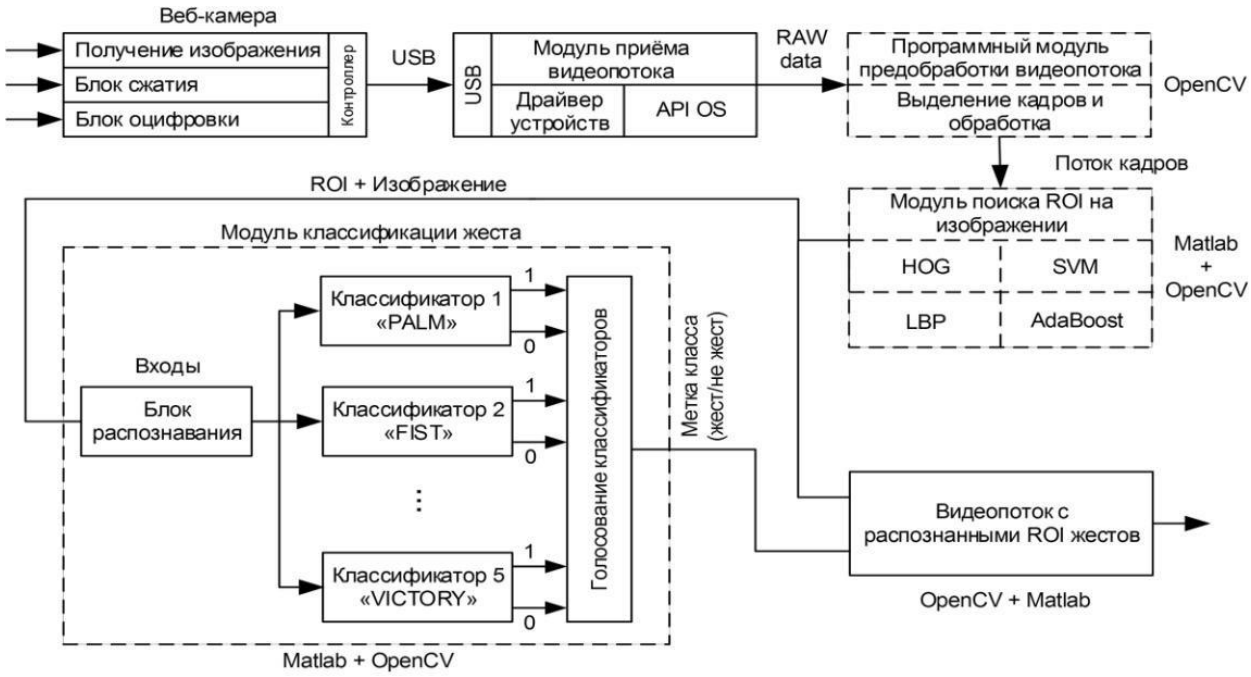

Рисунок 2.

На данной схеме мы видим какие стадии проходит изображение, сделанное на веб-камеру и из каких компонентов, состоит алгоритм, который

использует методы машинного обучения, для более эффективного и качественного распознавания жеста с видеопотока.

Технологии распознавания жестов

Данные для распознавания жестов собираются через датчики, которые могут быть разделены на несколько групп в зависимости от технологий захвата данных. Как показано на Рисунке 3, основными являются две категории сбора данных: основанные на изображениях и не основанные на изображениях.

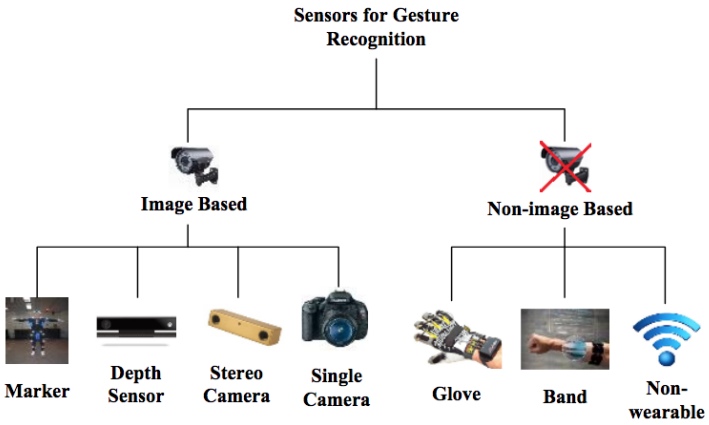

Рисунок 3.

1. Метод на основе получения изображений

Большинство современных технологий были вдохновлены природой. Каждый день мы своим зрение видим жесты и распознаем их. Поэтому и технологии распознавания жестов на основе изображений, которые видят камеры использовать разумней всего.

Маркер. В данном подходе оптическая камера является датчиком. Но чтобы данный метод сработал пользователю необходимо на себе иметь видимые маркеры [6]. Со временем данная технология будет имеет более быструю графическую обработку, чем раньше.

Камера. Анализ жестов используя камеры разработчики начали еще в начале 90-х годов. Система использует датчик скорости и специально

разработанный процессор, для визуальных вычислений, с помощью которого достигается высокоскоростное распознавание жестов.

Стереокамера. Уже в начале 2000-х годов, инженеры разработали стереоподход к созданию 3D-зрения. В данном подходе применяются уже две оптические камеры для создания информации о глубине 3D [16]. Данный метод является более эффективным, но у него есть один самый большой минус - сложность вычислений и трудности с калибровкой.

Сенсор глубины. Последние 10 лет в быстром темпе развивается технология глубокого зондирования. Мы определяем глубинный датчик как моно-датчик глубины, который имеет ряд преимущественных характеристик по сравнению с традиционными стереокамерами [6]. Разработано два типа датчиков глубины: камера с «временем пролета» и Microsoft Kinect [10].

2. Методы, основанные не на изображениях

Вторая технология по распознаванию жестов основана не на изображениях. Последние исследования в MEMS, микроэлектромеханические системы, и сенсорных технологиях далеко ушли в перед разрабатывая второй тип технологий [5].

Перчатки. Самый большой минус данного метода состоит в том, что он требует проводное подключение акселерометров и гироскопов [6]. Подходы на основе перчаток имеют сложности в процедурах калибровки и настройки.

Браслет. Эта технология использует сенсоры на браслетах [25]. Браслеты используют беспроводные технологии и датчики электромиограммы, благодаря чему можно не использовать подключение кабелей, в отличие от перчаток. Главное условие – контакт сенсоров с запястьем; руки и пальцы пользователя остаются свободны.

Бесконтактные устройства. Третий тип, о котором нужно упомянуть это технологии, не связанные с изображениями, они используют датчики, не

предназначенные для ношения. Данные датчики обнаруживают жесты без возможности контакта с человеческим телом [25].

Методы распознавания жестов

1. Метод Виолы – Джонса

В 2001 году Пол Виола и Майкл Джонсон, компьютерные исследователи, изобрели метод, который является основополагающим для поиска объектов на изображении в реальном времени. Метод использует технологию скользящего окна, который заключается в том, что рамка, размер которой меньше чем размер изображения, двигается с некоторым шагом и с помощью каскада детекторов характеристик определяется есть ли в рассматриваемом окне необходимый жест. Метод состоит из двух под алгоритмов:

- алгоритм обучения, скорость работы, которого не важна;
- алгоритм распознавания, скорость которого очень важна [1].

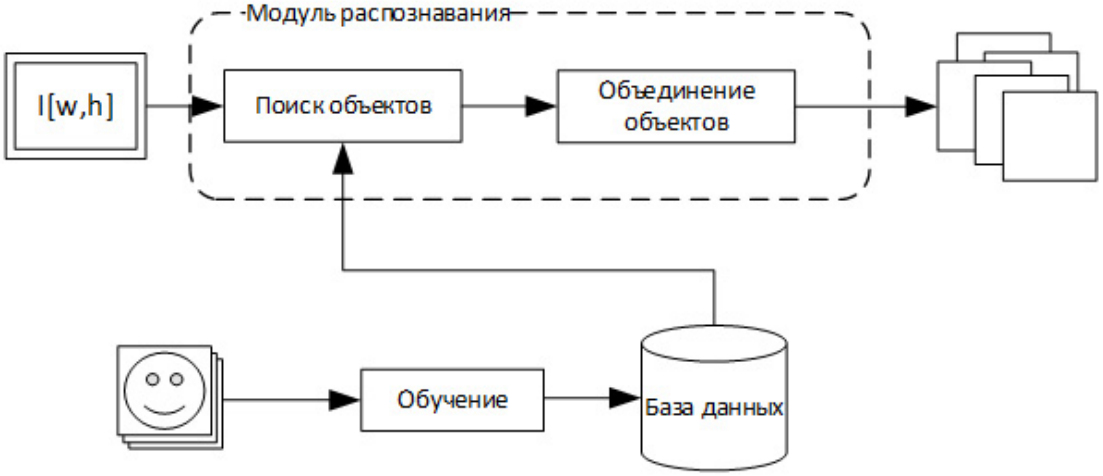

Рисунок 4.

По сравнению с другими методами распознавания жестов, метод Виолы-Джонсона имеет ряд преимуществ:

- возможно обнаружение более одного лица на изображении;

•использование простых классификаторов показывает хорошую скорость и позволяет использовать этот метод в видеопотоке.

Минусом данного метода является сложность его обучения так как для обучения требуется большое количество тестовых данных и предполагает большее время обучения, на которое могут уйти дни и месяцы [21].

2. Методы скелетизации

Скелетом в компьютерной графике называется множество точек, равноудалённых от границ фигуры. Скелет подчёркивает геометрические и топологические свойства фигуры. Фактически скелет является представлением формы фигуры, упрощает её дальнейший анализ. В качестве наиболее популярных алгоритмов скелетизации бинарного изображения можно назвать волновой алгоритм областей, метод Щепина и шаблонный метод [17].

Волновой алгоритм областей. Его задачей является векторное представление изображения в виде нагруженного графа – то есть определение концевых точек, точек пересечения, а также линий и дуг, составляющих фигуры. Метод заключается в анализе пути прохождения сферической волны по изображению [17]. На каждом этапе анализируется смещение центра масс точек, образующих новый шаг волны, относительно его предыдущих положений. После завершения построения скелета с помощью сферической волны, полученный результат оптимизируется и анализируется, отыскиваются особые точки фигуры.

Метод Щепина. Метод разработанный академиком Российского образования Михаила Щепина. Для каждого внешнего и внутреннего контура изображения находятся исходные верхние левые точки. Для очередной точки контура рассматривается конфигурация восьми ее соседей. Точка удаляется, если она не является концевой, и если после ее удаления ее соседи по-

прежнему будут образовывать связное множество [8]. После анализа точки и ее соседей, и возможного удаления точки осуществляется переход к следующей точке контура таким образом, чтобы остаться на границе изображения. Далее шаг за шагом удаляется один слой точек. Слои удаляются до тех пор, пока не останутся только неудаляемые точки.

Шаблонный метод. Метод скелетизации изображения по шаблонам первоначально был предложен Азриэлем Розенфельдом. Для получения скелетного изображения используются шаблоны, предназначенные для удаления лишних пикселей, где знаком «X» отмечены пиксели любого цвета. В любой области, соответствующей одному из шаблонов, удаляется черный центральный пиксель [7]. Осуществляется несколько проходов по изображению, пока не останется пикселей, подлежащих удалению.

3. Вейвлет-преобразование

Для начала разберемся, что такое вейвлет. *Вейвлет* - математическая функция, позволяющая анализировать различные частотные компоненты данных. Вейвлет-преобразование является эффективным для решения многих задач, связанных с обработкой изображения. Коэффициенты вейвлетов содержат информацию об анализируемом процессе, поэтому нужно выбирать правильный вейвлет, который определяет какую информацию нужно извлекать из процесса. Чаще всего используется один из первых вейвлет - Вейвлет Хаа́ра [20]. Он был основан в 1990 году венгерским математиком Альфредом Хааром. Данный вейвлет основан на ортогональной системе функций, обладающие компактным носителем и хорошо локализованы в пространстве. Идея ортогональных вейвлетов в последствии была доработана английским математиков Ингрид Добеши, предложив использовать функции, вычисляемые итерационным путём, названные вейвлетами Добеши.

4. Классический алгоритм роя частиц

Данный метод был разработан в 1995 году двумя инженерами, Джеймсом Кеннеди и Расселом Эберхартом. Он оптимизирует непрерывные нелинейные функции, которые и называются алгоритмом роя частиц. Алгоритм роя частиц широко применяется в задачах машинного обучения (в частности, для обучения нейросетей и распознавания изображений [9]), параметрической и структурной оптимизации в области проектирования, в областях биохимии и биомеханики [15]. По эффективности он может соперничать с другими методами глобальной оптимизации, а низкая алгоритмическая сложность способствует простоте его реализации. Алгоритм роя частиц появился относительно недавно, однако различными исследователями уже был предложен целый ряд его модификаций. Можно выделить несколько путей улучшения классического алгоритма, реализованных в большинстве из них, например:

- соединение алгоритма с другими алгоритмами оптимизации;
- уменьшение вероятности преждевременной сходимости путем изменения характеристик движения частиц;
- динамическое изменение параметров алгоритма во время оптимизации.

5. Инвариантные дескрипторы Фурье

Понятие дескрипторов Фурье было введено Косгрифом в 1960 году применительно к использованию преобразований Фурье в задачах описания геометрических фигур. Основная идея данного метода — представление контура (замкнутой кривой) конечным набором коэффициентов разложения Фурье. Дескрипторы Фурье широко используются в задачах распознавания образов, при обработке медицинских, картографических [11]. По данному методу рассматриваемый контур состоит из пронумерованных пикселей. Каждый пиксель имеет свои координаты, соединив которые получаем описанный контур изображения.

Классификация жестов

Классификация жестов – это последний и самый важный шаг в распознавании жестов. Большинство человеческих жестов – это динамические жесты. Один динамический жест всегда состоит из нескольких кадров. Чтобы классифицировать динамические жесты, классификация жестов должна выполняться после или вместе с отслеживанием жестов.

1. Метод K-ближайших соседей

Данный алгоритм является фундаментальным и базовым. Он классифицирует входные данные в соответствии с ближайшими примерами обучения.

2. Скрытая Марковская модель

Скрытая Марковская модель (HMM) – популярный алгоритм классификации жестов. HMM – это комбинация ненаблюдаемой цепи Маркова и стохастического процесса. Распознавание жеста – это проблема, которая задает последовательность наблюдений O, идентифицирует наиболее вероятную последовательность состояний X [13]. Для решения проблемы применяется алгоритм максимизации ожидания (EM).

3. Метод опорных векторов (Support Vector Machine)

Метод опорных векторов (SVM) является дискриминационным классификатором, определяемым разделительной гиперплоскостью. Границы решения классификации определяются путем максимизации расстояния от границы. Оптимальная разделительная гиперплоскость максимизирует запас обучающих данных. Примеры обучения, наиболее близкие к оптимальной гиперплоскости, называются вспомогательными векторами. На Рисунке 4 представлен пример метода опорных векторов.

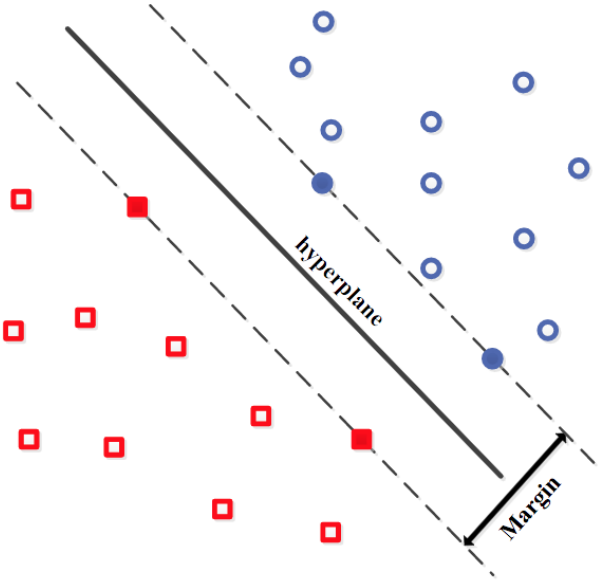

Рисунок 5.

4. Метод ансамбля

Метод ансамбля – это еще один широко используемый алгоритм классификации жестов. Основное предположение ансамблевого метода состоит в том, что ансамбли более точны, чем слабые отдельные классификаторы. Одним из известных ансамблевых методов является алгоритм Адабуст [4]. Существенным преимуществом данного алгоритма является то, что он не нуждается в большом количестве обучающих данных. На рисунке 5 представлен пример алгоритма Адабуст.

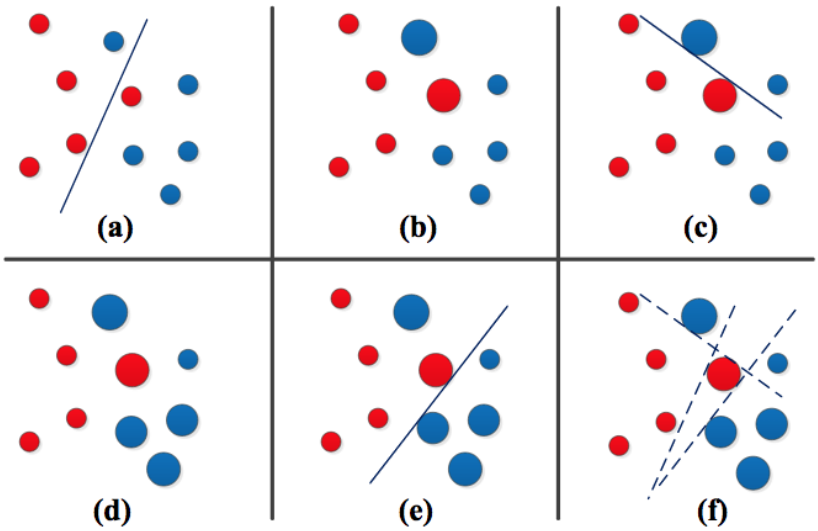

Рисунок 6.

5. Динамическое деформирование временных рядов

Динамическое деформирование временных рядов (DTW) является оптимальным алгоритмом выравнивания для двух последовательностей. DTW генерирует кумулятивную матрицу расстояний, которая искажает последовательности нелинейным способом. Первоначально DTW использовался для распознавания речи, но в последнее время существует множество приложений DTW в распознавании жестов.

6. Искусственные нейронные сети

Искусственная нейронная сеть (ANN) – это семейство моделей обработки информации, основанных на биологических нейронных сетях. ANN состоит из множества взаимосвязанных обрабатывающих объединений (нейронов), которые работают параллельно. В последнее время существует много исследований, в которых используются ANN для распознавания жестов [9]. В нескольких статьях также представлена система распознавания жестов, в которой объединены ANN с другими методами классификации.

7. Глубокое обучение

Глубокое обучение – это быстро развивающаяся отрасль машинного обучения. Глубокое обучение позволяет компьютеру моделировать данные с абстракциями высокого уровня, используя множественную нейронную сеть на уровне обработки. Более того, в отличие от традиционных алгоритмов обучения, глубокое обучение не требует ручной подготовки данных, что позволяет использовать преимущества экспоненциально увеличивающихся объемов доступных данных и вычислительных мощностей. В настоящее время глубокое обучение применяется в распознавании изображений, распознавании речи, анализе данных и т. д. В частности, глубокое обучение используется для решения проблемы распознавания человеческих действий в режиме реального времени в режиме видеомониторинга, в котором содержится большое количество данных.

Рассмотрев все имеющиеся методы проведем сравнительную характеристику и составим таблицу с достоинствами и недостатками по каждой отдельной характеристики.

Таблица 1. – обзор подходов в классификации жестов.

Подход	**Преимущества**	**Недостатки**
Метод K-ближайших соседей	Простота	K-параметр следует выбирать осторожно
Скрытая Марковская модель	Гибкость обучения и проверки, прозрачность модели	Необходимо отрегулировать множество свободных параметров
Метод опорных векторов	Могут применяться различные функции ядра	Число опорных векторов растет линейно с размером обучающего набора
Метод ансамбля	Не требуется большого количества данных обучения	Легко переобучить, чувствительность к шумам и выбросам (outliers)
Динамическое деформирование временных рядов	Надежное нелинейное выравнивание между образцами	Сложность времени и по объему данных
Искусственные нейронные сети	Может обнаруживать сложные нелинейные зависимости между переменными	Принцип «черного ящика», не может использоваться при наличии небольшого набора данных для обучения
Глубокое обучение	Не нуждаются в хорошей подготовке признаков, превосходит другие методы машинного обучения	Требуется большое количество обучающих данных и значительные вычислительные мощности.

Эксперимент

Для тестирования работы предложенных алгоритмов и методов использован открытый фреймворке для обработки видео- и аудиоданных в MediaPipe для Android. Программа протестирована на телефоне Sony Experia с процессором Qualcomm Snapdragon 650, 1,8-ГГц, объемом оперативной памяти 3 ГБ, видеокамерой 23 МП 1/2.3 Exmor™ RS, передающей 30 кадров в секунду с разрешением 320x180. Для корректной работы приложения на телефон был установлен инструмент для сборки и тестирования с открытым исходным кодом Bezel 1.1.0 и Python 3.7.

Для проведения нашего эксперимента мы используем технологию на основе получения изображений через камеру. Изучив методы и алгоритмы по распознаванию жестов для проведения экспериментальной части работы был выбран метод вейвлет-преобразований Хаара и Добеши. А в качестве метода классификации жестов будет использован -

Наша программа содержит три процесса для распознавания жеста:

- Программа BlazePalm [18], которая принимает на вход изображение и помогает распознать границы ладони;
- Распознавание точек. Модель для разметки ладони, которая принимает на вход обрезанное по границам изображение ладони и выдает 3D точки ладони. Модель распознает положение 21 точки ладони и выдает их координаты на выходе;
- Распознавание жестов [19]. Детектор жестов, принимает скелет, полученный на прошлом этапе с последовательностью точек ладони по заранее размеченным классам. Поверх распознанному скелету ладони применяется модель для классификации жестов. В программу уже были записаны жесты: кулак, один, два, три, четыре, пять.

Ниже представлена диаграмма визуализации процессов по распознаванию жестов:

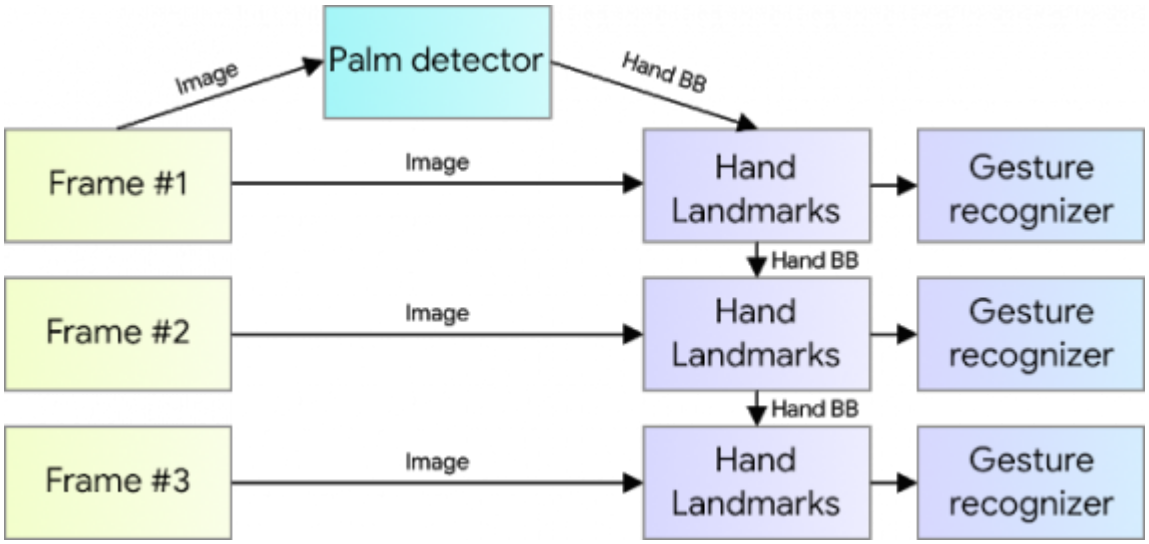

Рисунок 7.

Заранее в базу данных были добавлены тестовые фотографии, на которых были выделены точки с координатами. На каждые 6 жестов было добавлено по пять фотографий различных ракурсов с намеченными координатами. Тестирование программы проводилось на различных фонах (улица, дом) и при разном освещении (сильном, среднем и слабом).

Освещение	**Вычисление средней ошибки при различном освещении**
Слабое освещение	16.1 %
Сильное освещение	25.7 %
Среднее освещение	13.4 %

В ходе экспериментов было выяснено, что распознавание жестов не работает на тыльной стороне ладони. Алгоритм не распознает 21 одну контрольную точку, а значит в базу данных необходимо добавлять фотографии с координатами.

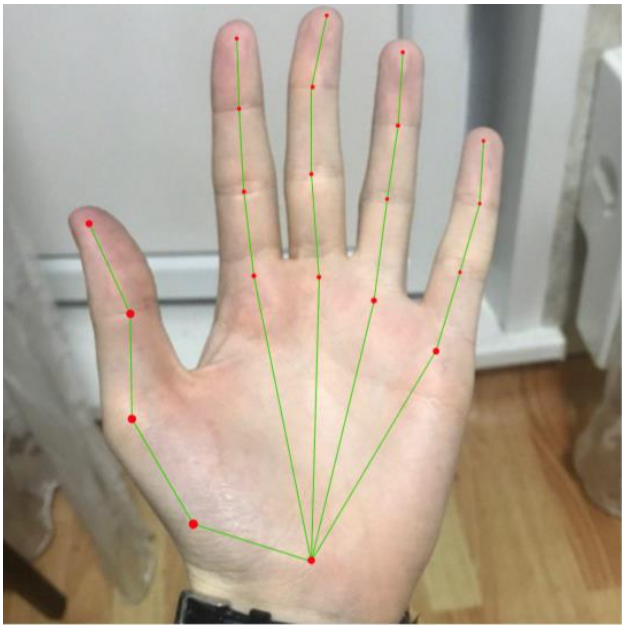

Рисунок 8.

На рисунке 8 мы видим пример идеального распознавания жеста «пять» на видео. Все точки отчетливо распознаны и соединены между собой. Для большей наглядности координатные точки имеют красный цвет, а линии соединяющие их зелены. Данный жест алгоритм легко распознавал при любом типе освещения и любой местности.

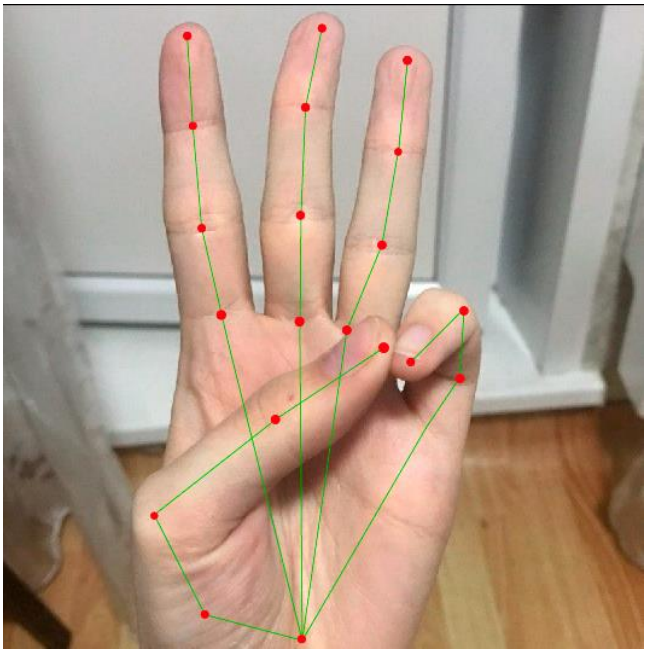

Рисунок 9.

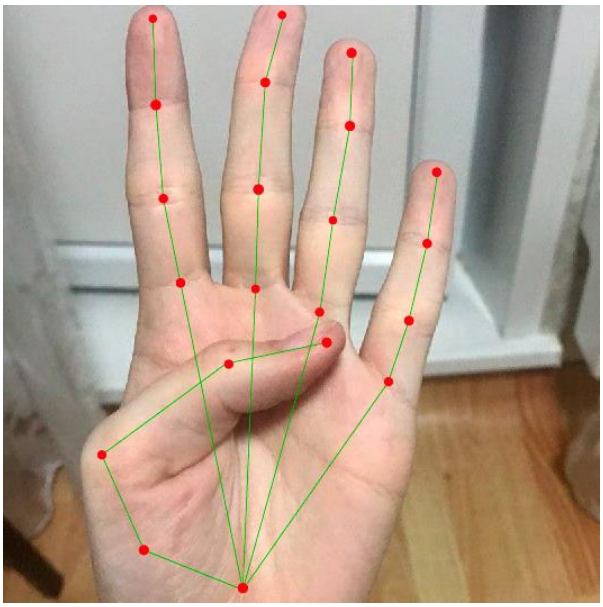

Рисунок 10.

На рисунках 9 - 10 мы видим удачные примера распознавания жестов «три» и «четыре». При тестировании была выявлена проблема данного алгоритма, которая заключается в том, что при перекрытии координатной точки алгоритм не может распознать жест. Поэтому руку приходится смещать на несколько градусов в право или лево, чтобы камера смогла найти необходимую координатную точку. Так же при плохом или слишком ярком освещении алгоритм распознает на 10-15% меньше, чем при среднем освещении. При распознавании более сложных жестов на улице алгоритм может принимать ветку дерева за палец и выдавать не правильный результат распознавания.

Проведя цикл экспериментов можно выявить какие характеристики нужны для фотографий, которые содержаться в базе данных для обучения алгоритма. Так же очень важным фактором является характеристики камеры. В следующем разделе выделены основные характеристики для изображений и камеры, которые подойдут как для исследуемого мной алгоритма, так и для других существующих.

Практические рекомендации для проектировщика

Залогом корректной работы распознавания жестов является правильная фотография в базе данных, а также качественная картинка с камеры. После проведения эксперимента и изучения дополнительной литературы можно сделать вывод какие изображения, должны находиться в базе данных и какие требования стоит предъявлять к видеокамерам.

1. Требования к изображению в базе данных

Для того чтобы сравнивать изображение, полученное с веб-камеры необходимо для начала занести в базу данных фотографии необходимых жестов в различных ракурсах. Фотографии должны быть качественными. Для более точного и быстрого сопоставления рисунка из базы данных и изображения с веб-камеры нужно использовать линии с опорными точками, что позволит многим методам оптимизировать работу.

С использованием линий и опорных точек фон уже не является важным фактором, как и освещение. Но если фотографии в базе данных не содержат линии и опорные точки, то следует соблюдать правила такие, как:

- фон изображения должен быть нейтральным;
- освещение должно быть равномерным.

2. Требования к установке камеры и ее характеристикам

Начнем с *плотности пикселов*. Минимальное количество 500 pix/м во всей предполагаемой зоне распознавания. Данное условие определяет разрешение и угол обзора камеры.

Глубина резкости должна покрывать всю зону контроля. Глубина резкости всегда зависит от фокусного расстояния объектива и диафрагмы. То есть рекомендуется заранее просчитывать глубину резкости и выбирать камеру с возможностью управления диафрагмой. Профессиональные камеры позволяют задавать пределы открытия и закрытия диафрагмы объектива.

Получаемый кадр должен быть четким, для этого значение *выдержки* на камере должно составлять не менее 1/100 секунды.

Для хорошей картинки стоит выбирать камеру с *высокой светочувствительностью* (матрица от 1/3“ и больше), и лучше всего чтобы на ней стоял *светосильный объектив* (от F1,4 и лучше).

Освещение в зоне контроля должно быть *хорошее и ровное* (150 и более люкс). Это, наверно, одно из самых сложно выполнимых требований. Тут нужен индивидуальный подход к каждому объекту. Например, контроль входной группы может быть затруднен засветкой камеры светом с улицы.

Технические требования к организации системы распознавания жестов можно свести в таблицу:

Таблица 2 – технические методы системы

ПАРАМЕТР	ТРЕБОВАНИЕ
Плотность пикселей	Не менее 500 пикс/м
Углы отклонения	Не более 15^0 по вертикали и горизонтали
Диафрагма объектива	Регулируемая; с возможностью ручного управления
Светосила объектива	Не менее 1:1.4 т.е., например, объектив 1:1.2 имеет большую светосилу, чем 1.4
Размер матрицы	Не менее 1/3”
Выдержка	Не более 1/100 с т.е. с выдержкой, например, 1/25 с возможно появление смазывания движущихся объектов
Освещенность	Не менее 150 лк; освещение лица должно быть равномерным

ЗАКЛЮЧЕНИЕ

В заключении хочется сказать, в ходе эксперимента было выявлено, что алгоритмы по распознаванию жестов нуждаются в доработке. И хотя приведенные выше разделы предоставляют общую картину распознавания жестов, можно смело сказать, что уже в скором времени нас может ожидать революционный скачок технических решений. Для компаний это повысит конкурентоспособность, а значит заинтересованность различных компаний в развитии систем распознавания жестов пойдет потребителям только на пользу. Проанализировав все данные можно представить некоторые из прогнозируемых будущих тенденций.

Датчик глубины и распознавание жестов на основе модели человека. Несмотря на понимание жестов человеческого тела, датчик глубины вместе с подходом к модели тела контролирует движение человека, что обеспечивает более безопасную среду для различных инструментов. Здесь можно применять различные методы скелитизации.

Система распознавания жестов с несколькими датчиками. Все датчики имеют преимущества и недостатки. Например, датчик-браслет имеет большую зону зондирования; обладает хорошими характеристиками в распознавании жестов тела. Для более эффективного использования потенциала производительности системы, в одной системе можно использовать комбинации различных датчиков распознавания жестов.

Комбинационный алгоритмический подход. Изучив множество алгоритмов классификации жестов видно, что они имеют свои преимущества и недостатки. Как говорилось выше, для улучшения скорости работы можно использовать еще и комбинацию алгоритмов по распознаванию жестов.

Список использованных источников

1. Виола П., Джонс М. Дж. Быстрое обнаружение объектов с использованием усиленного каскада простых признаков // IEEE Conf. по

компьютерному зрению и распознаванию образов. Кауаи, Гавайи, США, 511–518, 2001 г.

2. А. Эрол, Г. Бебис, М. Николеску, Р. Д. Бойл, Х. Твомбли, Оценка позы рук на основе зрения: обзор, Компьютерное зрение и понимание изображений, 52–73, 2007.
3. М. А. Херст, С. Т. Дюме, Э. Осман, Дж. Платт, Б. Шолкопф, Опорные векторные машины, Интеллектуальные системы и их приложения, IEEE, 18-28, 1998.
4. Алгоритм машинного обучения AdaBoost [Электронный ресурс] - Режим доступа: http://www.machinelearning.ru.
5. Н. Р. Хоу, М. Э. Левентон, В. Т. Фриман, Байесовская реконструкция трехмерного движения человека по видео с одной камеры, в: NIPS, 820–826, 1999.
6. С. Митра, Т. Ачарья, Распознавание жестов: обзор, Системы, человек и кибернетика, Часть С: Приложения и обзоры, IEEE Transactions on, 311–324, 2007.
7. Гонсалес П., Вудс Р. «Цифровая визуализация», Техносфера, 1072, 2005.
8. Местецкий Л.М. «Математические методы распознавания образов», МГУ, ВМиК, 42–44, 2002–2004.
9. Преимущества нейронных сетей [Электронный ресурс] - Режим доступа: http://www.aiportal.ru/articles/neuralnetworks/advantages.html.
10. Kinect 3D Hand Tracking [Электронный ресурс]. – Режим доступа: http://cvrlcode.ics.forth.gr/handtracking/.
11. Т. Берг и П. Н. Бельюмер. «Классификаторы Том против Пита и выравнивание с сохранением личности для проверки лица», BMVC, 2012.
12. Xtion PRO [Электронный ресурс] - Режим доступа: https://www.asus.com/3D-Sensor/Xtion_PRO/.
13. Фольц П.В., Кинч В., Ландауэр Т.К. Измерение согласованности текста с помощью латентного семантического анализа, 285–307, 1998.

14. Сабина Хельвиг и Рольф Ванка. Оптимизация роя частиц в многомерных ограниченных пространствах поиска. На симпозиуме Swarm Intelligence Symposium, страницы 198–205. ИИЭР, 2007.
15. И. Ойкономидис, Н. Кириазис и А. Аргирос, «Эффективное трехмерное отслеживание артикуляций рук на основе моделей с использованием Kinect», в BMVC 2011, 2011.
16. Intel Perceptual Computing SDK (Хабр) [Электронный ресурс] — Режим доступа: https://habr.com/ru/company/intel/blog/166529/.
17. Обнаружение ладоней и пальцев на изображении (Хабр) [Электронный ресурс] - Режим доступа: https://habr.com/ru/post/169709/.
18. В. Базаревский, Ю. Картынник, А. Вакунов, К. Равендран, М. Грундманн, «BlazeFace: субмиллисекундное нейронное распознавание лиц на мобильных графических процессорах», Семинар CVPR по компьютерному зрению для дополненной и виртуальной реальности, 4–9 , 2019.
19. Hand Tracking (CPU) на GitHub [Электронный ресурс] — Режим доступа: https://github.com/google/mediapipe/blob/master/mediapipe/docs/hand_tracking_mobile_gpu.md.
20. Буй Тхи Тху Чанг, Фан Нгок Хоанг, Спицын В.Г. «Алгоритмические и программные средства классификации цифровых изображений с использованием вейвлет-преобразования Хаара и нейронных сетей», Вестник Томского политехнического университета. Т. 319. № 5, 103–106, 2011.
21. Алгоритм и методы обнаружения и распознавания жестов рук на видео в режиме реального времени [Электронный ресурс]. URL: http://sj.kubsau.ru/2014/03/20.
22. OpenCV (Библиотека компьютерного зрения с открытым исходным кодом) [Электронный ресурс] — Режим доступа: https://opencv.org/.

23. Блог Google AI [Электронный ресурс] — Режим доступа: https://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html.
24. С. Рой, С. Гош, А. Барат, М. Чаттопадхьяй, Д. Чоудхури, Реализация электромиографии в режиме реального времени для обнаружения жестов рук с использованием микроакселерометра, в: Искусственный интеллект и эволюционные вычисления в инженерных системах, Springer, 357 –364, 2016.
25. Google, Project Soli [Электронный ресурс] - Режим доступа: https://www.google.com/atap/project-soli/.
26. М. Хансард, С. Ли, О. Чой, Р. П. Хоро, Времяпролетные камеры: принципы, методы и приложения, Springer Science & Business Media, 2012.

Reference

1. Viola P., Jones M.J. Rapid object detection using a boosted cascade of simple features // IEEE Conf. on Computer Vision and Pattern Recognition. Kauai, Hawaii, USA, 511–518, 2001.
2. A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, X. Twombly, Vision-based hand pose estimation: A review, Computer Vision and Image Understanding, 52–73, 2007.
3. M. A. Hearst, S. T. Dumais, E. Osman, J. Platt, B. Scholkopf, Support vector machines, Intelligent Systems and their Applications, IEEE, 18-28, 1998.
4. AdaBoost machine learning algorithm [Electronic resource] - Access mode: http://www.machinelearning.ru.
5. N. R. Howe, M. E. Leventon, W. T. Freeman, Bayesian Reconstruction of 3D Human Motion from Single-Camera Video, in: NIPS, 820–826, 1999.
6. S. Mitra, T. Acharya, Gesture recognition: A survey, Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 311–324, 2007.

7. Gonzalez P., Woods R. "Digital Imaging", Technosphere, 1072, 2005.
8. Mestetsky L.M. "Mathematical Methods of Pattern Recognition", Moscow State University, VMiK, 42–44, 2002–2004.
9. Advantages of neural networks [Electronic resource] - Access mode: http://www.aiportal.ru/articles/neuralnetworks/advantages.html.
10. Kinect 3D Hand Tracking [Electronic resource]. – Access mode: http://cvrlcode.ics.forth.gr/handtracking/.
11. T. Berg and P. N. Belhumeur. “Tom-vs-pete classifiers and identitypreserving alignment for face verification”, BMVC, 2012.
12. Xtion PRO [Electronic resource] - Access mode: https://www.asus.com/3D-Sensor/Xtion_PRO/.
13. Foltz P.W., Kintsch W., Landauer T.K. The measurement of textual coherence with Latent Semantic Analysis, 285 – 307, 1998.
14. Sabine Helwig and Rolf Wanka. Particle Swarm Optimization in High-Dimensional Bounded Search Spaces. In Swarm Intelligence Symposium, pages 198–205. IEEE, 2007.
15. I. Oikonomidis, N. Kyriazis, and A. Argyros, “Efficient model-based 3D tracking of hand articulations using Kinect”, in BMVC 2011, 2011.
16. Intel Perceptual Computing SDK (Habr) [Electronic resource] - Access mode: https://habr.com/ru/company/intel/blog/166529/.
17. Detection of palms and fingers in the image (Habr) [Electronic resource] - Access mode: https://habr.com/ru/post/169709/.
18. V. Bazarevsky, Y. Kartynnik, A. Vakunov, K. Raveendran, M. Grundmann, “BlazeFace: Sub-millisecond Neural Face Detection on Mobile GPUs”, CVPR Workshop on Computer Vision for Augmented and Virtual Reality, 4 – 9 , 2019.
19. Hand Tracking (CPU) on GitHub [Electronic resource] - Access mode: Https://github.com/google/mediapipe/blob/master/mediapipe/docs/hand_tracking_mobile_gpu.md.

20. Bui Thi Thu Chang, Phan Ngoc Hoang, Spitsyn V.G. "Algorithmic and software for classifying digital images using the Haar wavelet transform and neural networks", Bulletin of the Tomsk Polytechnic University. T. 319. No. 5, 103–106, 2011.
21. Algorithm and methods for detecting and recognizing hand gestures on video in real time [Electronic resource]. URL: http://sj.kubsau.ru/2014/03/20.
22. OpenCV (Open Source Computer Vision Library) [Electronic resource] – Access mode: https://opencv.org/.
23. Google AI Blog [Electronic resource] - Access mode: https://ai.googleblog.com/2019/08/on-device-real-time-hand-tracking-with.html.
24. S. Roy, S. Ghosh, A. Barat, M. Chattopadhyay, D. Chowdhury, Real-time Implementation of Electromyography for Hand Gesture Detection Using Micro Accelerometer, in: Artificial Intelligence and Evolutionary Computations in Engineering Systems, Springer, 357 –364, 2016.
25. Google, Project Soli [Electronic resource] - Access mode: https://www.google.com/atap/project-soli/.
26. M. Hansard, S. Lee, O. Choi, R. P. Horaud, Time-of-flight cameras: principles, methods and applications, Springer Science & Business Media, 2012.

© Коломыцев А.С., Вердиев О.Р., 2022 Научно-образовательный журнал для студентов и преподавателей «StudNet» №7/2022.

Для цитирования: Коломыцев А.С., Вердиев О.Р. РАСПОЗНАВАНИЕ ЖЕСТОВ НА ВИДЕО // Научно-образовательный журнал для студентов и преподавателей «StudNet» №7/2022.