МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

‹‹КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ››

(ФГБОУ ‹‹КубГУ››)

**Факультет** **компьютерных технологий и прикладной математики**

**Кафедра информационных технологий**

КУРСОВАЯ РАБОТА

МОДЕЛИРОВАНИЕ АНАТОМИЧЕСКИХ ОБЛАСТЕЙ КИСТИ МЕТОДОМ ВИЗУАЛЬНЫХ ПРИЗНАКОВ

Работу выполнил \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Д.А. Тельнов

(подпись)

Направление подготовки 01.02.03 Прикладная математика и информатика курс 3

Направленность Программирование и информационные технологии

Научный руководитель

канд. пед. наук, доц. \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Н.Ю. Добровольская

(подпись, дата)

Нормоконтролер

канд. пед. наук, доц.\_\_\_\_­­­­­­­\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ А.В. Харченко

(подпись, дата)

Краснодар

2022

СОДЕРЖАНИЕ

Введение………………………………………………………………......... 3

1. Теоретические основы распознавания жестов….................................. 5

1.1 Средства распознавания жестов...………………………………… 5

1.2 Идентификация жестов методом визуальных признаков............... 10

1.3 Отслеживание жестов......................................................................... 12

1.4 Классификация жестов...…………………………………………... 17

2. Технологии распознавания жестов......................................................... 24

2.1 Обзор существующих решений для распознавания жестов........... 24

3. Описание работы сервиса…..………………………………………….. 28

4. Эксперимент…………………………………………………………….. 31

Заключение………………………………………………………………... 37

Список использованных источников…………………........................................................................ 38

ВВЕДЕНИЕ

В истории развития персонального компьютера наблюдалась эволюция пользовательского интерфейса в человеко-машинном взаимодействии (ЧМВ) от текстового командного интерфейса до графического интерфейса, от простой клавиатуры до мыши, джойстика, электронной ручки, видео камеры, и т.д. Устройства ЧМВ стали более удобными и естественными для пользователя. В настоящее время, с введением новых понятий, таких как «виртуальная среда - ВС», «человеко-машинная интеллектуальная интеракция -ЧМИИ», «перцепционный пользовательский интерфейс - ППИ» и т.д. требуется разработка более мощных и удобных способов взаимодействия человека с компьютерной системой.

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. По сравнению с традиционными устройствами ЧМВ, жесты являются менее навязчивым, простым, более удобными и естественным способом взаимодействия для пользователей. Тем не менее, выразительность жестов все еще остается недостаточно изученной для решения проблемы человеко-машинного взаимодействия.

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 ЧМИИ и начало привлекать множество исследователей. Кроме того, разработанные в ЧМИИ технологии также находят применение в других областях, таких как управление роботами, телеконференции, перевод языка жестов (сурдоперевод), управление компьютерными играми, и т.д.

1 Теоретические основы распознавания жестов

1.1 Средства распознавания жестов

Перед процессом распознавания жестов рабочие данные должны быть собраны датчиками. Как показано на рисунке 1, существуют две основные категории сбора данных: основанные на изображениях и не основанные на изображениях.



Рисунок 1 – Различные типы датчиков распознавания жестов

1. Метод на основе получения изображений.

Маркер.

В маркерном подходе датчиком является оптическая камера. В большинстве решений на основе маркеров пользователям необходимо носить видимые маркеры [7]. Сегодня мы получаем гораздо более быструю графическую обработку по сравнению с двадцатью годами ранее. В результате на рынке доступно множество датчиков распознавания жестов.

Камера.

В начале 90-х годов исследователи начали анализировать жесты при помощи одиночной камеры. Недостатком однокамерного подхода является ограничение угла обзора, что влияет на надежность системы. Однако в недавних исследованиях применялся однокамерный подход к высокоскоростному распознаванию жестов. Система использует датчик скорости и специально разработанный процессор визуальных вычислений для достижения высокоскоростного распознавания жестов.

Стереокамера.

Чтобы добиться надежного распознавания жестов, исследователи предложили стерео-подход к созданию 3D-зрения. Cтереокамерные подходы - приложения, которые используют две оптические камеры для создания информации о глубине 3D. Хотя системы стереокамер улучшили устойчивость во внешней среде, они по-прежнему страдают от таких проблем, как сложность вычислений и трудности с калибровкой.

Сенсор глубины.

В последнее время технологии глубокого зондирования быстро развиваются. Датчики глубины обладают рядом преимуществ по сравнению с традиционными стереокамерами. Например, можно предотвратить проблемы настройки калибровки и условий освещения. Кроме того, выходной информацией датчика глубины является информацией о глубине 3D. По сравнению с информацией о цвете информация о глубине 3D упрощает проблему идентификации жестов [7]. Существует два типа общих не стерео-датчиков глубины: камера с «временем пролета» (ToF) и Microsoft Kinect (или аналогичные ИК-датчики).

Основным принципом камер ToF является определение времени прохождения света. В различных публикациях были введены примеры распознавания жестов на основе ToF-камер. Преимуществом камер ToF является более высокая частота кадров. Ограничение камеры ToF заключается в том, что разрешение камеры сильно зависит от ее светочувствительности и рефлексии.

Microsoft Kinect предоставляет дешевое и простое решение для распознавания жестов. Kinect — инфракрасный датчик глубины. Аналогичными датчиками являются ASUS Xtion Pro и Apple PrimeSense. Kinect имеет ИК-излучатель, ИК-датчик и датчик цвета. Он широко используется в сфере развлечений, образования и исследований с большим сообществом разработчиков. Из-за ограниченного разрешения, в настоящее время Kinect может использоваться для распознавания жестов тела и распознавания жестов рук на небольшом расстоянии. Для распознавания жестов рук и кистей рук на расстоянии более 2 метров лучше использовать другие подходы.

1. Методы, основанные не на изображениях.

В распознавании жестов в течение долгого времени доминировали датчики на основе изображений. Недавние разработки в MEMS и сенсорных технологиях значительно улучшили технологии распознавания жестов, основанные не на изображениях.

Перчатки.

Жестовые интерфейсы на основе перчаток также используются для распознавания жестов. Обычно методы на основе перчаток требуют проводного подключения акселерометров и гироскопов. Однако громоздкая перчатка с проводами может вызвать проблемы в HRC. Подходы на основе перчаток также имеют сложности в процедурах калибровки и настройки.

Браслет (носимая электроника).

Другая бесконтактная технология использует сенсоры на браслетах. Эти сенсоры позволяют использовать беспроводные технологии и датчики электромиограммы, что позволяет избежать подключения кабелей. Сенсоры должны контактировать с запястьем; руки и пальцы пользователя могут быть свободны. Примером браслетного датчика является устройство Myo.

Бесконтактные устройства.

В третьем типе технологий, не связанных с изображениями, используются датчики, не предназначенные для ношения. Бесконтактные датчики могут обнаруживать жесты без контакта с человеческим телом. Google представил Project Soli, систему радиолокационного контроля и распознавания жестов на радиочастотном спектре (радар). Устройство способно распознавать разные жесты рук на небольшом расстоянии. В течение многих лет MIT является ведущим новатором в области распознавания жестов. Недавно Adib из MIT представил систему WiTrack и RF-Capture, которая отслеживает движение пользователя по радиочастотным сигналам, отраженным от человеческого тела. Система способна захватывать человеческие жесты даже из другой комнаты через стену с точностью до 20 см. Таким образом, технологии, не пригодные для ношения, являются перспективными и быстрорастущими сенсорными технологиями для распознавания жестов.

1. Сравнение сенсорных технологий.

Таблица 1 содержит сравнение различных сенсорных технологий. Показаны преимущества и недостатки различных подходов.

Таблица 1. Достоинства и недостатки различных средств распознавания

|  |  |  |
| --- | --- | --- |
|  | Достоинства | Недостатки |
| Маркеры | Низкая вычислительная нагрузка | Маркеры на теле пользователя |
| Камера | Простота установки | Низкий уровень надежности |
| Стереокамера | Надежность | Сложность вычислений, трудная калибровка |
| ToF камера | Высокая частота кадров | Разрешение зависит от света и отражения |
| Microsoft Kinect | Поддержка программного обеспечения для распознавания жестов тела | Максимальное расстояние распознавания жестов не более 2 метров |
| Перчатка | Скорость отклика, точность трекинга | Громоздкое устройство с проводами |
| Браслет | Скорость отклика, область действия | Браслет должен контактировать с человеческим телом |
| Бесконтактные устройства | Не требуется контакта с телом | Низкое разрешение, технология требует доработки |

Сенсоры для использования внутри помещений: датчики глубины являются наиболее перспективными технологиями на основе изображений. Они обладают преимуществами простоты калибровки, установки и скорости обработки данных.

Бесконтактные сенсоры являются наиболее перспективной технологией среди подходов, основанных не на изображении. Они не требуют прямого контакта с пользователями. Эта категория сенсоров также является быстрорастущей отраслью на рынке технологий.

1.2 Идентификация жестов

Идентификация жеста — первый шаг в процессе распознавания жестов после получения необработанных данных, полученных с датчиков. Идентификация жеста означает обнаружение жестовой информации и сегментацию соответствующей жестовой информации из необработанных данных.

Человеческие руки и тело обладают уникальными визуальными особенностями. Поэтому использование таких визуальных признаков в идентификации жестов вполне обоснованно.

Цвет.

Цвет – это простая визуальная функция для идентификации жестов из фоновой информации. Однако на системы распознавания жестов на основе цветов сильно влияют освещение и тени в сложной среде HRC (HumanRobotCollaboration) [5]. Еще одна распространенная проблема в обнаружении цвета заключается в том, что цвет кожи человека сильно различается среди человеческих рас. Из-за вышеперечисленных проблем, в современных подходах, цвет кожи рассматривается только как один из многих параметров при идентификации жестов.

Локальные признаки.

В распознавании жестов на основе изображения условия освещения сильно влияют на качество идентификации жестов. Поэтому многие исследователи используют метод локальных признаков, который не чувствителен к условиям освещения. Локальный подход к объектам – это детализированный подход на основе текстур. Он раскладывает изображение на более мелкие области, которые не соответствуют частям тела. Как показано на рисунке 2, одной из наиболее важных локальных функций является преобразование признаков инвариантных объектов (SIFT). Метод SIFT является вращательным, трансляционным, масштабируемым и частично осветляющим инвариантом. Существует несколько подобных методов локальных признаков, например, SURF и ORBare, предложенные в более поздние годы. Как правило, подходы к локальным особенностям также рассматриваются только как один из множества параметров при идентификации жестов. Несколько методов идентификации, таких как методы формы и контура, методы движения и методы обучения, основаны на локальных признаках.

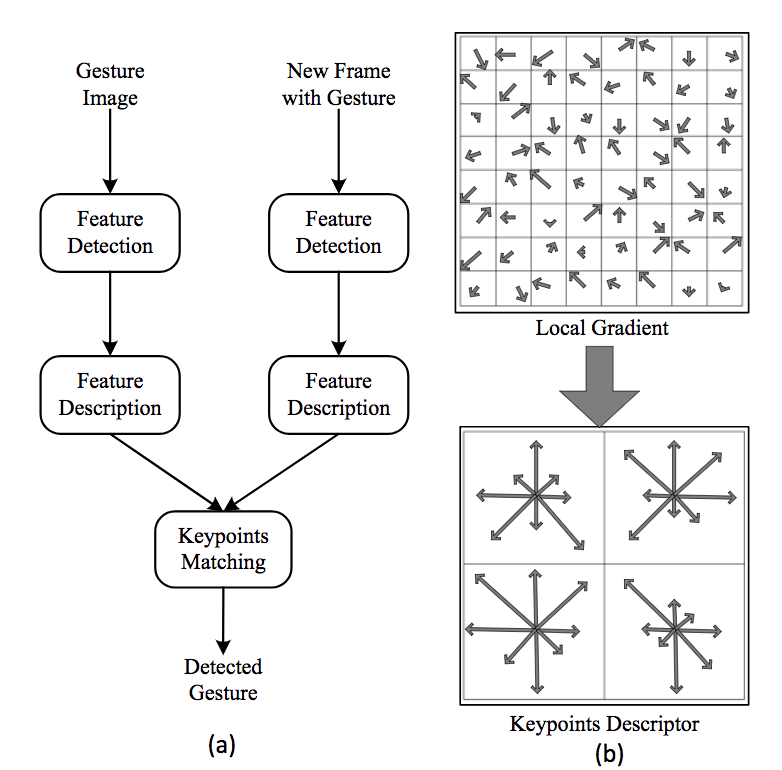


Рисунок 2 – SIFT: (a) алгоритм SIFT для идентификации жестов; (b) Пример дескриптора функции SIFT

Форма и контур.

Другой интуитивный и простой способ идентификации жестов — использование уникальной форму и контура человеческого тела в среде HRC. Существенный вклад в определение формы и соответствия был внесен Belongie (соавторство) [4]. Они ввели метод дескриптора контекста формы. Дескриптор контекста формы используется для обнаружения похожих фигур на разных изображениях. Разработка датчика глубины дает возможность точно измерять форму поверхности. 3D-модели, созданные на основе технологий, позволяют очень детально представлять форму человеческого тела.

Движение.

В конкретной среде HRC человек является единственным движущимся объектом в массиве необработанных данных. Поэтому движение является полезной функцией для обнаружения человеческих жестов. Оптический поток является ключевой технологией идентификации жестов на основе движения. Он не нуждается в вычитании фона, что является преимуществом по сравнению с подходами на основе формы и контура. Среди подходов, основанных не на изображении, распознавание жестов на основе движения является популярным методом.

1.3 Отслеживание жестов

Отслеживания жестов определяется, как процесс поиска временных соответствий между кадрами. В частности, это фокусировка на проблеме отслеживания жестов, которая связывает идентифицированный жест в предыдущих кадрах с текущим фреймом. Что касается статических жестов, которые могут быть представлены одним кадром, отслеживание жестов не требуется. Пример отслеживания жестов показан на рисунке 4.

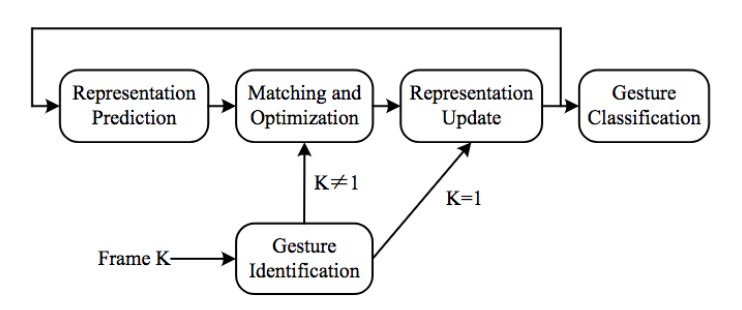


Рисунок 4 – Пример отслеживания жестов

1. Отслеживание единичной гипотезы

Отслеживание единичной гипотезы относится к оценке наилучшего соответствия при согласовании с минимальной ошибкой. Поэтому при использовании единичной гипотезы отслеживания, жест представлен только одной гипотезой. Большинство усовершенствованных алгоритмов отслеживания основаны на технологиях единичной гипотезы.

Средний сдвиг.

Отслеживание методом среднего сдвига – это базовая технология отслеживания. Метод выполняет сопоставление с гистограммами цвета RGB. Для каждого нового кадра средний сдвигающий трекер сравнивает расстояние Бхаттачарьи между гистограммами целевого окна нового кадра и старыми кадрами.

Фильтр Калмана.

Фильтр Калмана (KF) представляет собой рекурсивный алгоритм реального времени, используемый для оптимальной оценки, лежащих в основе состояний ряда шумовых и неточных результатов измерений, наблюдаемых с течением времени. Технологический поток KF показан на рисунке 5 [3]. В настоящее время KF активно развивается и применяется в различных областях, таких как аэрокосмическая промышленность, робототехника и экономика.

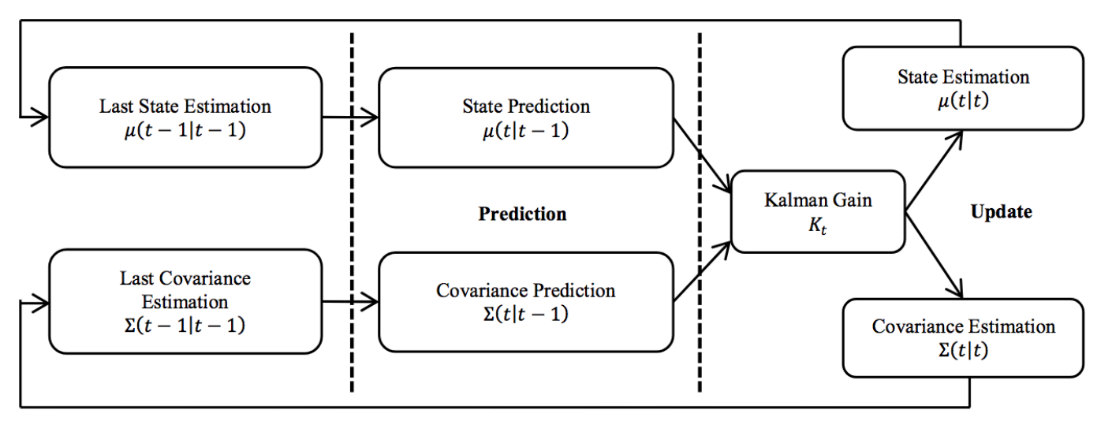


Рисунок 5 – Процесс работы фильтра Калмана

Расширенные фильтры Калмана.

KF предполагает, что вектор состояния является линейной моделью. Расширеный фильтр Калмана (Extend Kalman Filter – EKF) является функциональным алгоритмом отслеживания, даже если модель нелинейна. Другим алгоритмом, который решает ту же проблему под другим углом, является сигма-точечный фильтр Калмана (Unscented Kalman Filter – UKF). UKF решает проблему, применяя детерминированный подход с взвешенной выборкой.

1. Отслеживание множественных гипотез

Во многих сценариях HRC одновременно работают несколько рабочих-людей на одной рабочей станции. Чтобы отслеживать жесты нескольких работников одновременно, следует применять технологии отслеживания множественных гипотез.

Многочастичный фильтр.

Многочастичный фильтр (Фильтр частиц, Particle Filter – PF) является популярной технологией в задачах робототехники. В отличие от KF, PF не делает предположения о поздней модели. Представление PF является непараметрическим приближением, которое может представлять собой более широкое пространство распределения. Следовательно, PF удовлетворяет требованиям по отслеживанию гипотез. Пример PF показан на рисунке 6.

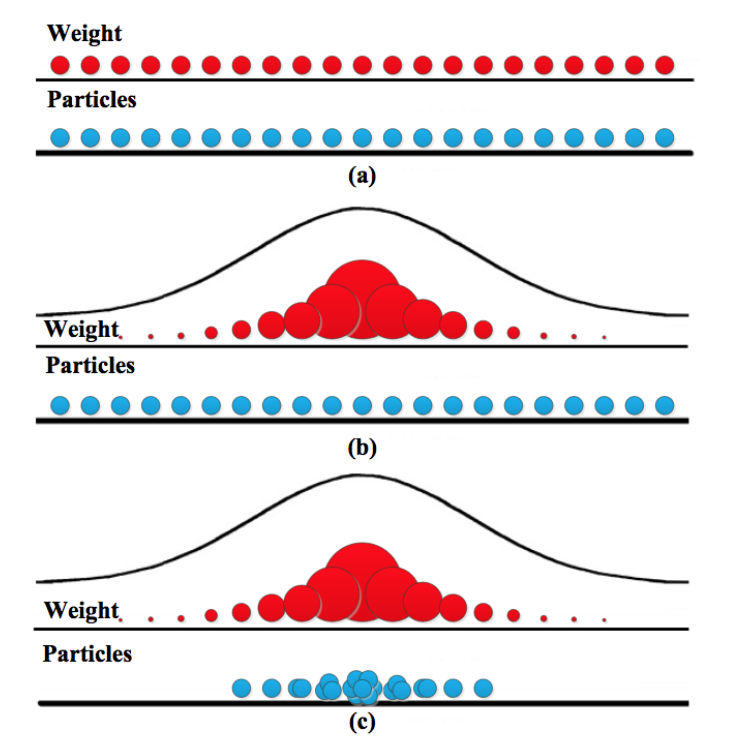


Рисунок 6 – Частицы и весовые коэффициенты: (а) после инициализации частиц; (b) после расчета весового коэффициента; (с) после повторной дискретизации

1. Расширенные методы отслеживания

За последнее время появилось много продвинутых методов отслеживания. Некоторые из этих передовых методов использовали часть алгоритмов отслеживания, упомянутых ранее. Другие методы улучшили отслеживаемость с помощью алгоритмов обнаружения или обучения.

Расширенное отслеживание модели.

Для долгосрочных проблем отслеживания многие алгоритмы терпят неудачу, поскольку цель поддерживает фиксированные модели. Расширенное отслеживание модели сохраняет целевое поведение или внешний вид из последних кадров изображения. Поэтому для целевой оценки зарезервировано больше целевой информации. Инкрементный визуальный трекер использует расширенную модель для сохранения большего объема информации для процесса отслеживания. Квон представил выборочное отслеживание. Расширенная модель сохраняется в процессе выборки. Трекер выбирает из многих трекеров и, соответственно, выбирается наиболее подходящий.

Отслеживание путем обнаружения.

Еще один вид алгоритмов отслеживания строится на базе алгоритмов обучения идентификации жестов. Для них в кадрах изображений применяется классификатор или детектор, чтобы идентифицировать жест из справочной информации. Одним из характерных подходов является трекер отслеживания, обучения и обнаружения. Этот подход объединяет результаты детектора объекта с устройством оптического отслеживания потока. Еще одна типичная технология отслеживания по обнаружению – применять множественное обучение экземпляров [2]. Алгоритм обучения может повысить надежность трекера и уменьшить количество параметров.

1. Сравнение различных подходов отслеживания жестов

Смеулдерз представил результат теста различных алгоритмов отслеживания жестов. Итоговый результат — нормализованный F-балл. Чем выше F-оценка, тем лучше качество отслеживания. На рисунке 7 представлены результаты испытаний в разных условиях видео.

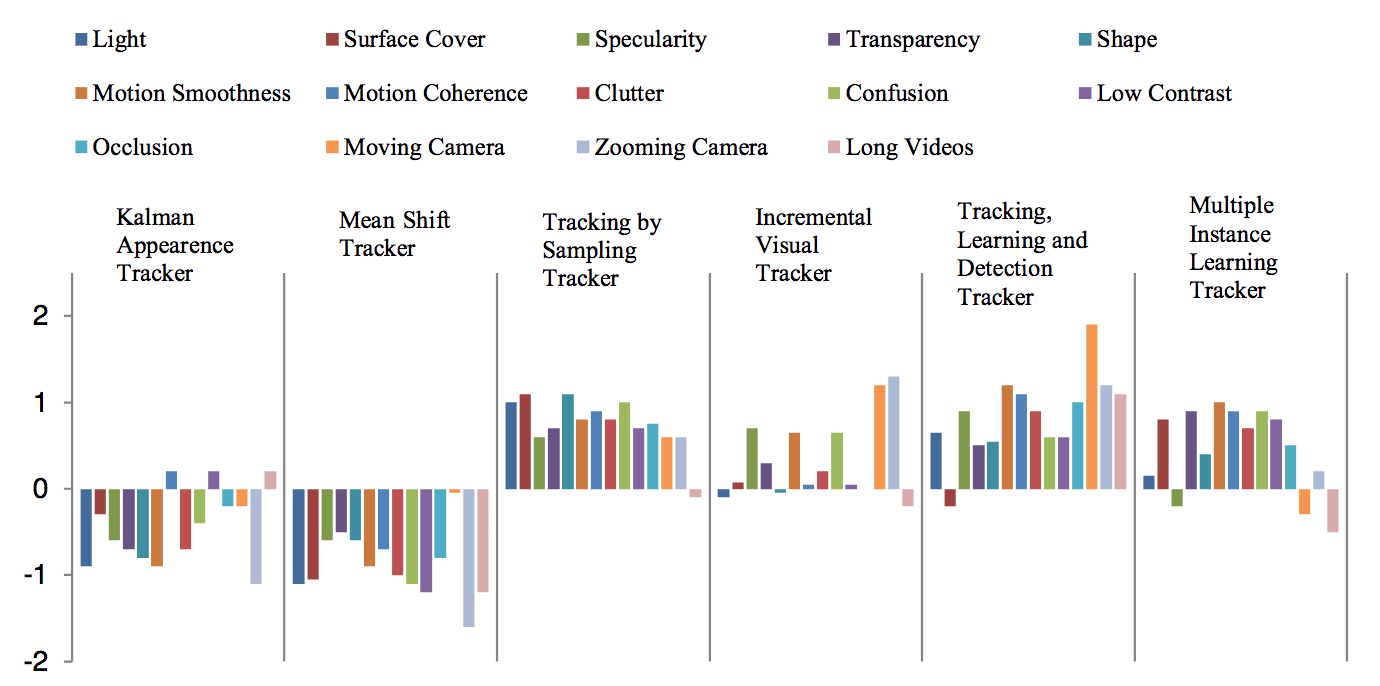


Рисунок 7 – Результат теста алгоритмов отслеживания в разных видео условиях

Легко заметить, что трекеры одиночной гипотезы менее производительны, чем остальные.

1.4 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 должна выполняться после или вместе с отслеживанием жестов.

1. Метод K-ближайших соседей (K-Nearest Neighbours)

Алгоритм K–Nearest Neighbors (KNN) – это фундаментальный и базовый алгоритм классификации жестов, который классифицирует входные данные в соответствии с ближайшими примерами обучения [8].

1. Скрытая Марковская модель

Скрытая Марковская модель (HMM) – популярный алгоритм классификации жестов. HMM – это комбинация ненаблюдаемой цепи Маркова и стохастического процесса. Пример HMM показан на рисунке 8, ненаблюдаемая цепь Маркова состоит из состояний X и вероятностей перехода состояний a. Стохастический процесс состоит из возможных наблюдений O и возможных выводов b. Распознавание жеста – это проблема, которая задает последовательность наблюдений O, идентифицирует наиболее вероятную последовательность состояний X. Для решения проблемы применяется алгоритм максимизации ожидания (EM) [9]

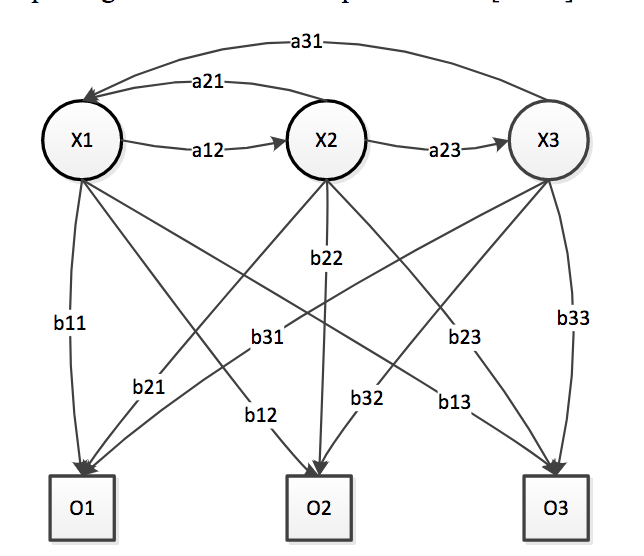


Рисунок 8 – Пример скрытой Марковской модели

1. Метод опорных векторов (Support Vector Machine)

Как показано на рисунке 9, метод опорных векторов (SVM) является дискриминационным классификатором, определяемым разделительной гиперплоскостью [11]. 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.

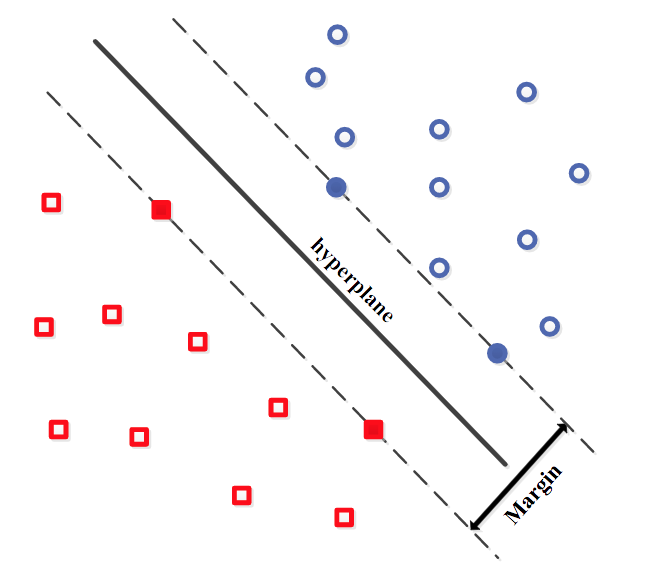


Рисунок 9 – Пример линейного метода опорных векторов

Общей проблемой для SVM является то, что число векторов поддержки растет линейно с размером обучающего набора.

1. 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 методов ансамбля является бустинг. Алгоритм бустинга начинается с нескольких слабых классификаторов. Слабые классификаторы применяются многократно. В обучающей итерации часть обучающих образцов используется в качестве входных данных. После итерации обучения создается новая граница классификации. После всех итераций алгоритм бустинга объединяет эти границы и сливается в одну конечную границу предсказания. Как показано на рисунке 10, другим известным ансамблевым методом является алгоритм Адабуст (Adaboost). Существенным преимуществом алгоритма Адабуст является то, что он не нуждается в большом количестве обучающих данных. [6].

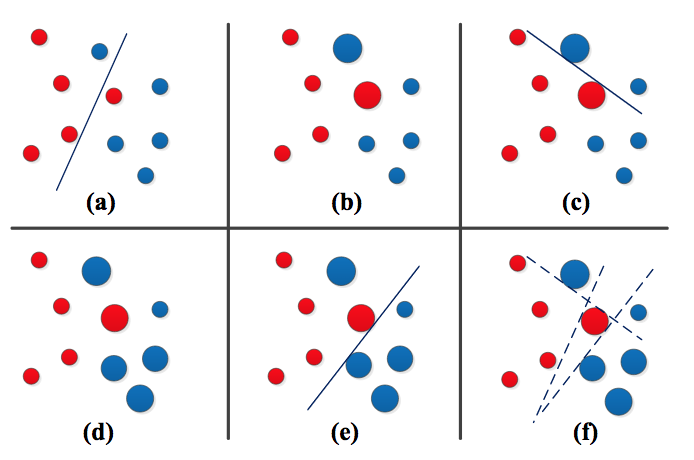


Рисунок 10 – Пример Адабуст: (a) слабый классификатор 1, (b) увеличение веса, (с) слабый классификатор 2, (d) увеличение веса, (e) слабый классификатор 3, (f) граница окончательного решения.

1. 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. В последнее время существует множество приложений DTW в распознавании жестов [1].

1. 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. Каждый союз (нейрон) получает входные данные, обрабатывает входные данные и дает выходные данные. ANN может использоваться для оценки функций, зависящих от большого количества входных данных. В последнее время существует много исследований, в которых используются ANN для распознавания жестов [12].

1. Глубокое обучение (Deep Learning)

Глубокое обучение – это быстро развивающаяся отрасль машинного обучения. Она 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 котором содержится большое количество данных. В настоящее время наиболее популярны две революционные нейронной сети: свёрточные (CNN) и рекуррентные (RNN) для глубоких обучающих архитектур.

1. Сравнение подходов классификации жестов

Таблица 2 содержит преимущества и недостатки подходов к классификации жестов.

Таблица 2. Преимущества и недостатки подходов к классификации жестов

|  |  |  |
| --- | --- | --- |
|  | 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. |

Одной из тенденций является подход глубокого обучения. Основными ограничениями глубокого обучения являются ограниченные вычислительные мощности. Однако экспоненциально возрастающая вычислительная мощность может легко решить эту проблему. Количество приложений классификации жестов, основанных на глубоком обучении, быстро растет. Другая тенденция заключается в объединении различных алгоритмов классификации. Каждый алгоритм классификации имеет свои преимущества и недостатки. Чтобы использовать это, различные классификаторы могут быть объединены для достижения лучшей производительности.

2 Технологии распознавания жестов

2.1 Обзор существующих решений для распознавания жестов

В настоящее время ведется довольно много исследований по созданию методов распознавания образов, позволяющих бесконтактно взаимодействовать с компьютером посредством жестов рук. В результате мы видим множество новых прототипов, готовых к массовому внедрению продуктов и патентов, в основе которых лежит технология бесконтактного жестового взаимодействия.

1. Высокоточная система жестового управления

Начинающая компания 3Gear из Сан-Франциско (США) создаёт новую систему жестового управления, способную отслеживать малейшие изменения положения кистей рук и пальцев [13].

Аппаратная составляющая комплекса включает две 3D-камеры, подвешенные на специальной рамке над поверхностью рабочего стола (рисунок 11).

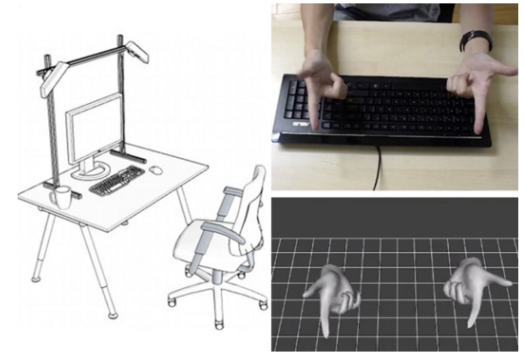


Рисунок 11 – Высокоточная система жестового управления

Изображения, снятые со скоростью 30 кадров в секунду, обрабатываются специализированным программным обеспечением. Этот процесс предполагает поиск совпадения в обширной базе данных, содержащей информацию о 30 тысячах возможных положений пальцев и рук.

Разработчики подчеркивают, что система способна реагировать на малейшие изменения — когда пальцы двигаются всего на один миллиметр. В этом комплексе высокая скорость реакции и распознавание жестов занимает в среднем 33 миллисекунды.

Возможные сферы применения: игровые и развлекательные комплексы, CAD-системы и др.

1. Volkswagen Golf R Touch Gesture Control

Компания Volkswagen на примере электрического автомобиля e-Golf Touch продемонстрировала некоторые особенности нового бортового информационно - развлекательного комплекса. Презентация состоялась на выставке CES 2016 в Лас-Вегасе (Невада, США) [14]. Бортовой центр поддерживает управление при помощи жестов (рисунок 12) и голосовых команд.



Рисунок 12 – Жестовое управление мультимедийной системой Volkswagen

Такой подход должен минимально отвлекать водителя от управления транспортным средством, делая взаимодействие с бортовой электроникой интуитивно понятным и максимально удобным.

1. Патент US8634980B1: Driving pattern recognition and safety control

Google предлагает отслеживать движения рук водителя и преобразовывать определённые жесты в команды для контроля бортовой электроники (рисунок 13) [10].

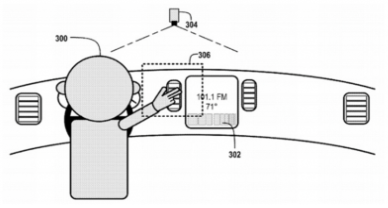


Рисунок 13 – Схема работы системы, описанной в патенте US8634980B1

К примеру, взмахнув ладонью рядом с окном, автомобилист сможет поднять или опустить стекло. А аналогичный жест рядом с мультимедийной системой позволит увеличить или уменьшить громкость.

Для отслеживания положения рук водителя предлагается использовать специальную камеру на потолке транспортного средства, позволяющую формировать карту глубины пространства внутри салона. Кроме того, данные будут собираться при помощи лазерного сканера.

3 Особенности реализации сервиса

Для реализации проекта используется язык Python с подключенной библиотекой компьютерного зрения и обработки изображения OpenCV, а также фреймворком, представленным Google, который помогает создавать мультимодальные конвейеры машинного обучения Mediapipe.

Вначале необходимо получить изображение с камеры. Для этого создаем объект VideoCapture() библиотеки OpenCV, принимающий на вход целочисленный аргумент, обозначающий номер камеры.

Далее создаем экземпляр класса Hands библиотеки MediaPipe. Он позволяет отслеживать кисти на изображении, используя конвейер машинного обучения, состоящий из нескольких моделей, работающих вместе: модель обнаружения кисти, которая работает с полным изображением и возвращает ориентированную ограничивающую рамку руки. Модель ориентира кисти, которая работает с обрезанной областью изображения, определяемой детектором ладони, и возвращает высокоточные трехмерные ключевые точки, расположенные на кисти и ее анатомических облостях. Передаем экземпляру класса настройки обнаружения: static\_image\_mode = False – изображение будет не статичным, max\_num\_hands = 2, – максимальное количество рук, min\_tracking\_confidence = 0.5, min\_detection\_confidence = 0.5 – минимальные значения достоверности, при котором обнаружение считается успешным.

Создаем экземпляр класса drawing\_utils библиотеки MediaPipe, чтобы отрисовать распознанные кисти на изображении.

После этого нам необходимо считывать каждый кадр изображения с камеры. Для этого запускаем бесконечный цикл и получаем изображение с помощью метода read() класса VideoCapture.

Для того, чтобы запустить процесс распознавания кисти в каждом кадре, передаем полученное изображение в метод process() класса Hends. Благодаря функционалу этого методу мы сможем получить координаты в пространстве отдельных анатомических областей с помощью метода multi\_hands\_landmarks[i].lendmark. Так как MediaPipe присваивает каждой области свою точку (рисунок 14).

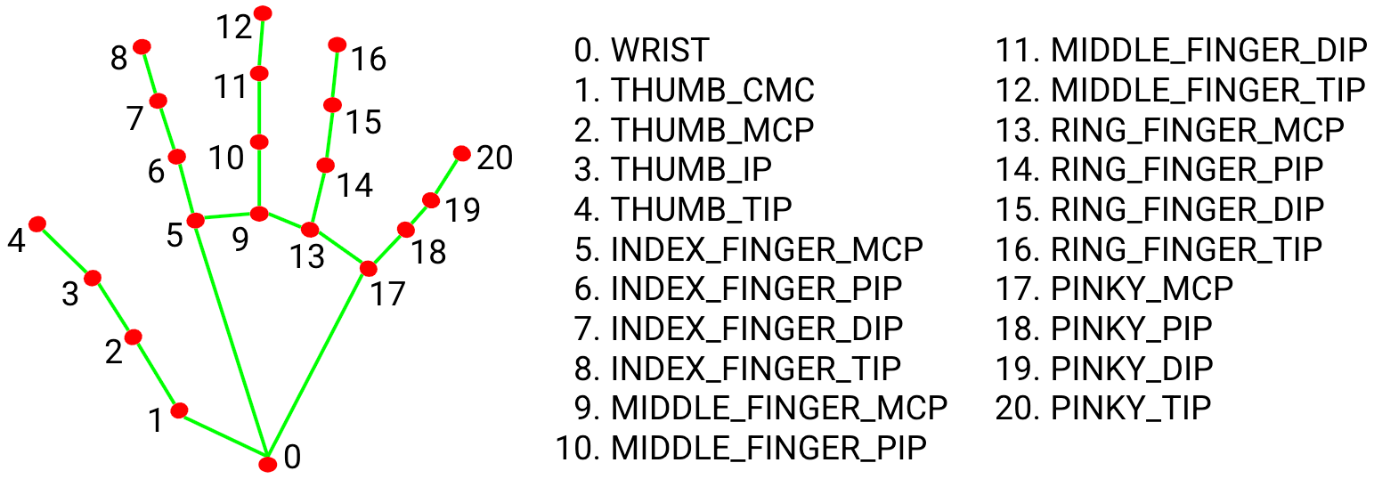


Рисунок 14 – Разбиение руки по точкам

Когда мы получили координаты расположения в пространстве каждой части кисти, нам остается обработать эти координаты для идентификации количества показанных пальцев.

После этого, выводим полученное изображение методом imshow() библиотеки OpenCV с отрисованным на анатомических областях точками в соответствии с рисунком 14 благодаря методу draw\_landmarks() класса drawing\_utils. Также выводим количество показанных пальцев.

Весь процесс получения изображения, обнаружения количества показанных пальцев и вывода можно представить в виде диаграммы (рисунок 15).

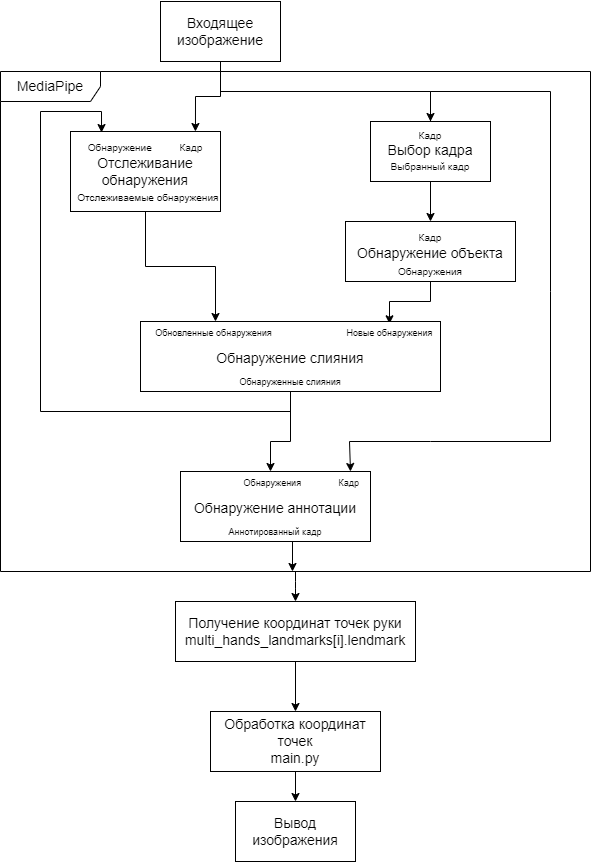


Рисунок 15 – Схема работы программы

4 Экспериментальная работа

Выделим критерии оценки качества распознавания анатомических областей кисти.

– Освещение

– Угол поворота

– Количество кистей

Опыт 1.

На рисунке 16 приведен кадр видеопотока с низкой освещенностью.

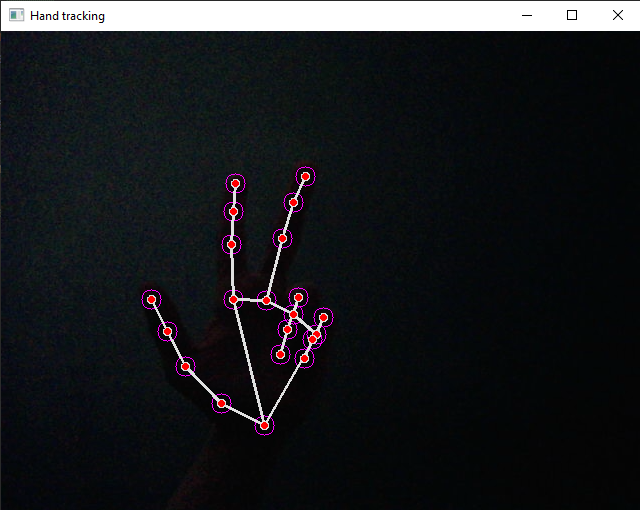


Рисунок 16 – Кадр с низкой освещенностью

Результат: при слабом освещении, сервис способен идентифицировать кисть и ее анатомические области.

Опыт 2.

На рисунке 17 приведен кадр видеопотока, на котором кисть повернута к камере под углом 90 градусов.

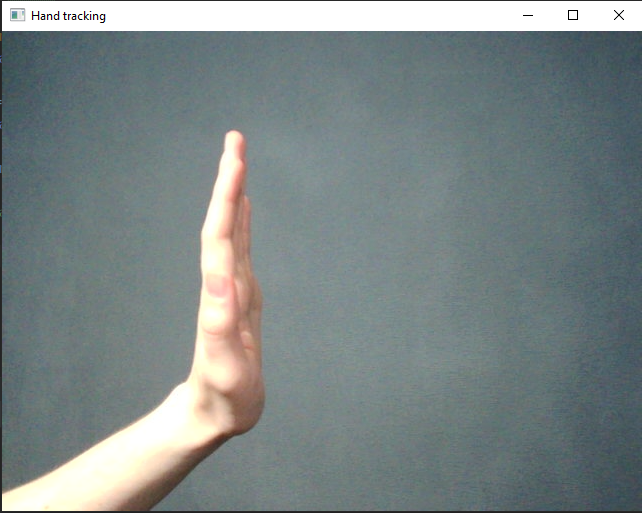
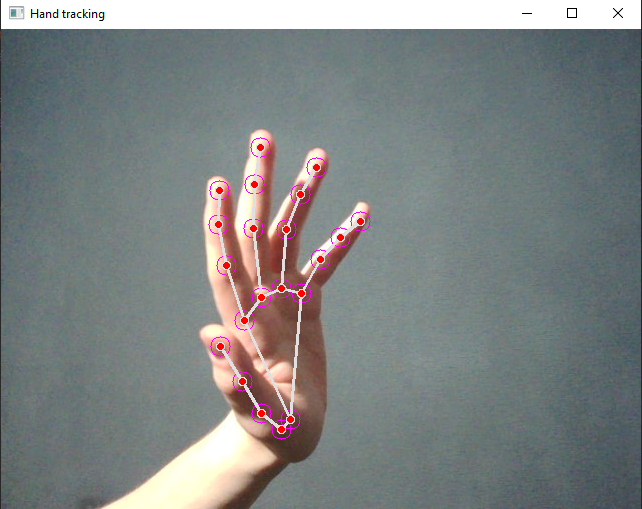


Рисунок 17 – Кадр с кистью, повернутой к камере на 90 градусов

На рисунке 18 приведен кадр видеопотока, на котором кисть повернута к камере менее чем на 90 градусов.



Рисунке 18 – Кадр с кистью, повернутой к камере менее чем на 90 градусов

На рисунке 19 приведен кадр видеопотока, на котором кисть повернута к камере более чем 90 градусов.

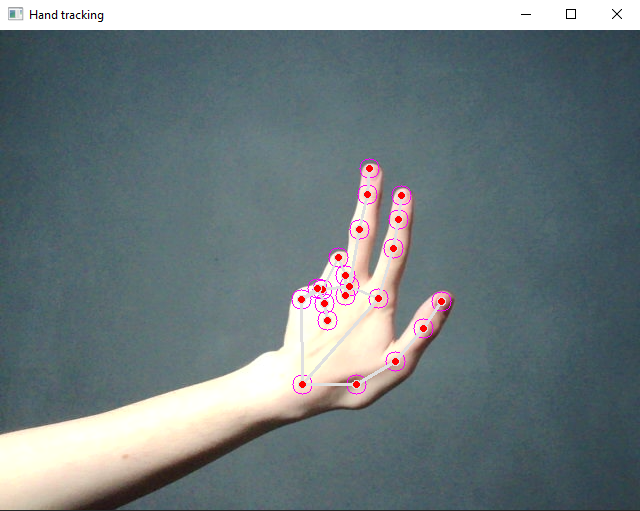


Рисунок 19 – Кадр с кистью, повернутой к камере более чем на 90 градусов

Результат: сервис способен распознать и идентифицировать кисть под любым углом, кроме угла в 90 градусов из-за того, что не может отследить все анатомические области.

Опыт 3.

На рисунке 20 приведен кадр видеопотока, в котором находится одновременно 3 кисти.

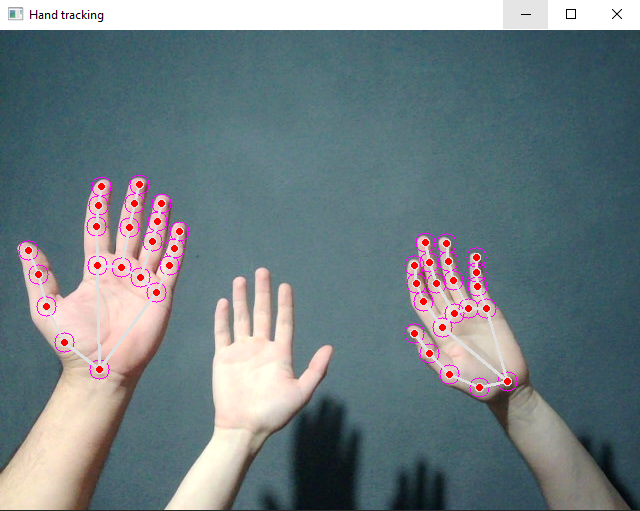


Рисунок 20 – Кадр с тремя кистями

Результат: Сервис способен распознать только две кисти, одновременно находящиеся в кадре. Количество распознаваемых кистей в кадре можно изменять в коде программы.

**ЗАКЛЮЧЕНИЕ**

Цель курсовой работы – моделирование кисти и её анатомических областей методом визуальных признаков – достигнута.

В теоретической части курсовой работы описаны основные принципы технологии распознавания и идентификации объектов, в том числе в видеопотоке.

Выполнен реферативный обзор существующих решений распознавания жестов, выделены их преимущества и недостатки. Рассмотрены особенности идентификации жестов в видеопотоке.

В работе предложен проект и реализация сервиса моделирования кисти в видеопотоке на основе определения визуальных признаков анатомических областей. Основной функцией сервиса является идентификация и подсчет количества пальцев руки как основной анатомической области руки.

При программной реализации функционала использовались библиотеки языка Python OpenCV и Mediapipe.

Разработанный сервис может являться составным элементом крупных проектов по работе с людьми с ограниченными способностями.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Аричи, Т. Надежное распознавание жестов с использованием предварительной обработки признаков и взвешенной динамической деформации времени, Мультимедийные инструменты и приложения // Т. Аричи, С. Челеби, А. С. Айдын, Т. Т. Темиз. – 2014. – 72 c.
2. Бабенко, Б. Визуальное отслеживание с многократным онлайн-обучением, в: Компьютерное зрение и распознавание образов // Б. Бабенко, М.-Х. Ян, С. Белонги. – 2009. – 983–990 с.
3. Балакришан, А. В. Теория фильтрации Калмана // А. В. Балакришан. – 2021. – 5–13 с.
4. Белонги, С. Сопоставление форм и распознавание объектов с использованием контекстов форм, Анализ шаблонов и машинный интеллект // С. Белонги, Д. Малик, Д. Пузича. – 2002. – 509–522 c.
5. Летесье, Ж. Визуальное отслеживание пальцев для интерактивных поверхностей, в: Материалы 17-го ежегодного симпозиума ACM по программному обеспечению и технологиям пользовательского интерфейса // Ж. Летесье, Ф. Берар. – 2004. – 119–122 с.
6. Микилотта, А. С. Обнаружение и отслеживание людей с помощью вероятностной сборки частей тела // А.С. Микилотта, Э.-Дж. Онг, Р. Боуден. – 2005. – 9 c.
7. Митра, C. Распознавание жестов: обзор, Системы, человек и кибернетика, Часть C: Приложения и обзоры // С. Митра, Т. Ачарья. – 2007. – 311–324 c.
8. Петерсон, Л. Э. K-ближайший сосед // Л. Э. Петерсон. – 2009. – 18–20 с.
9. Уилсон, А. Д. Параметрические скрытые марковские модели для распознавания жестов, Анализ шаблонов и машинный интеллект // А. Д. Уилсон, А. Ф. Бобик. – 1999. – 884–900 с.
10. Урмсон, К. П. Распознавание образов вождения и контроль безопасности // К. П. Урмсон, Д. А Долгов, Ф. Немек. – 2011. – 1 с.
11. Шёлкопф, Б. Машины опорных векторов, Энциклопедия биостатистики // Б. Шёлкопф, А. Смола. – 1998. – 3–5 с.
12. Хасан, Х. Статическое распознавание жестов рук с использованием нейронных сетей // Х. Хасан, С. Абдул-Карим. – 2014. – 147–181 с.
13. NND, Высокоточная система управления жестами. – 2012 – . URL: <https://nnd.name/2012/10/vysokotochnaya-sistema-zhestovogo-upravleniya/> (дата обращения 14.10.2022).
14. 3dnews, CES 2016: новая медиасистема Volkswagen с жестовым управлением. – 2016 – . URL: <https://3dnews.ru/926244> (дата обращения 16.10.2022).