

## УПРАВЛЕНИЕ ИНТЕРАКТИВНЫМ ДИСПЛЕЕМ ЖЕСТАМИ ПРИ ПОМОЩИ АЛГОРИТМОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ

Белорусский государственный университет, Минск, Беларусь

В рамках данной работы разработана программа для взаимодействия человека с компьютером посредством жестов с помощью алгоритмов компьютерного зрения. При проектировании приложения выбран язык программирования Python, а также фреймворки компьютерного зрения OpenCV и машинного обучения MediaPipe.

Люди в значительной степени зависят от пяти органов чувств, чтобы интерпретировать происходящие события в окружающем нас мире. Широко известно, что с помощью зрения человек получает наибольший объем информации об окружающих предметах, причем без непосредственного взаимодействия с ними. Глаза, через которые мы видим и воспринимаем многое, помогают нам видеть путь, по которому мы идем, дорогу, по которой мы едем, и следить за любым возможным столкновением.

Однако первые же эксперименты выявили огромные трудности, связанные с наделением машины способностью видеть. Оказалось, что изображения одного и того же объекта, полученные в естественных условиях, настолько не похожи друг на друга, что их очень сложно распознать. Не так просто оказалось осуществлять и навигацию в пространстве по изображениям, поскольку в явном виде плоские изображения не содержат информацию о трехмерных характеристиках окружающих объектов и расстояниях до них. Попытки решить данные проблемы привели к возникновению новой области знаний – компьютерного зрения [1].

Компьютерное зрение – это молодая, перспективная и развивающаяся область робототехники. Хотя многие проекты в этой области находятся в стадии разработки, способность создавать машины, обрабатывающие визуальные данные, уже давно продемонстрирована в современном обществе. Эта способность использовалась для анализа штрих-кодов, были разработаны системы безопасности, сканирующие глаза человека и отпечатки пальцев для идентификации личности. Также проводились эксперименты по управлению автомобилями с помощью компьютера, обрабатывающего данные с видеокамер.

Исходя из перечисленного выше, была поставлена задача: разработать программу, благодаря которой человек мог бы взаимодействовать с компьютером посредством жестов рук. Разработка системы управления компьютером велась на языке программирования Python с использованием библиотеки OpenCV и ML фреймворка от компании Google – MediaPipe. Кратко расскажем, что и для чего применяется из использованных инструментов.

### **MediaPipe**

MediaPipe – это кроссплатформенная платформа машинного обучения с открытым исходным кодом, используемая для построения сложных и мультимодальных прикладных конвейеров машинного обучения. Его можно использовать для создания передовых моделей машинного обучения, таких как распознавание лиц, отслеживание несколькими руками, обнаружение объектов и отслеживание, и многие другие. MediaPipe в основном выступает в качестве посредника для обработки реализации моделей для систем, работающих на любой платформе, что помогает разработчику больше сосредоточиться на экспериментировании с моделями, чем на системе.

Одной из ключевых особенностей MediaPipe является его поддержка передовых моделей машинного обучения. Модели машинного обучения – это алгоритмы, которые могут извлекать уроки из данных и делать прогнозы на основе этих данных. MediaPipe включает в себя ряд предварительно обученных моделей машинного обучения, которые можно использовать для таких задач, как обнаружение объектов, распознавание лиц и оценка позы. Эти

модели отличаются высокой точностью и могут выполнять эти задачи в режиме реального времени, что делает их идеальными для использования в мультимедийных приложениях.

### OpenCV

OpenCV (Open Source Computer Vision Library) – это открытая библиотека для работы с алгоритмами компьютерного зрения, машинным обучением и обработкой изображений. Написана на C++, но существует также для Python, JavaScript, Ruby и других языков программирования. Работает на Windows, Linux и MacOS, iOS и Android.

Также библиотека работает с машинным обучением — отраслью, которая обучает алгоритмы действовать тем или иным образом.

OpenCV применяется:

- в робототехнике – для ориентирования робота в пространстве, распознавания объектов и взаимодействия с ними;
- медицинских технологиях – для создания точных методов диагностики, например, 3D-визуализации органа при МРТ;

Применимо к поставленной задаче использовался фреймворк, который дал возможность отслеживания рук человека. Разработчики реализовали технологию точного отслеживания рук и пальцев с помощью машинного обучения (ML). Программа определяет 21 ключевую точку руки в 3D пространстве (высоту, длину и глубину) и на основании этих данных классифицирует жесты, которые показывает рука. Все это на основании всего одного фрейма видео, работает в реальном времени и масштабируется на несколько рук.

Решение состоит из 3 основных моделей, работающих вместе:

#### 1) Детектор ладони (BlazePalm):

- a) Принимает полное изображение из видео
- b) Возвращает ориентированный bounding box (ограничивающая рамка)

#### 2) Модель для определения ключевых точек на руке:

- a) Принимает обрезанную картинку руки;
- b) Возвращает 21 ключевую точку руки в 3D пространстве, как показано на рисунке 1

+ показатель уверенности

#### 3) Алгоритм распознавания жестов

- a) Принимает ключевые точки руки;
- b) Возвращает название жеста, который показывает рука [2].

На рисунке 1 представлены ключевые точки руки.



Рисунок 1 – Ключевые точки руки

Комплект моделей ручных ориентиров содержит модель обнаружения ладоней и модель обнаружения ручных ориентиров. Модель обнаружения ладоней находит руки на входном изображении, а модель обнаружения ориентиров рук идентифицирует конкретные ориентиры рук на обрезанном изображении руки, определяемом моделью обнаружения ладоней.

## Секция 2. Прикладные проблемы информатики

Поскольку запуск модели обнаружения ладоней занимает много времени в режиме воспроизведения видео или прямой трансляции, используется ограничивающая рамка, определенная моделью ориентиров рук в одном кадре, чтобы локализовать область рук для последующих кадров. Модель обнаружения ладоней запускается повторно только в том случае, если модель ручных ориентиров больше не определяет присутствие рук или не может отслеживать руки в кадре.

С помощью описанных выше библиотек и фреймворков была разработана программа со следующим функционалом и интерфейсом, представленном на рисунке 2.

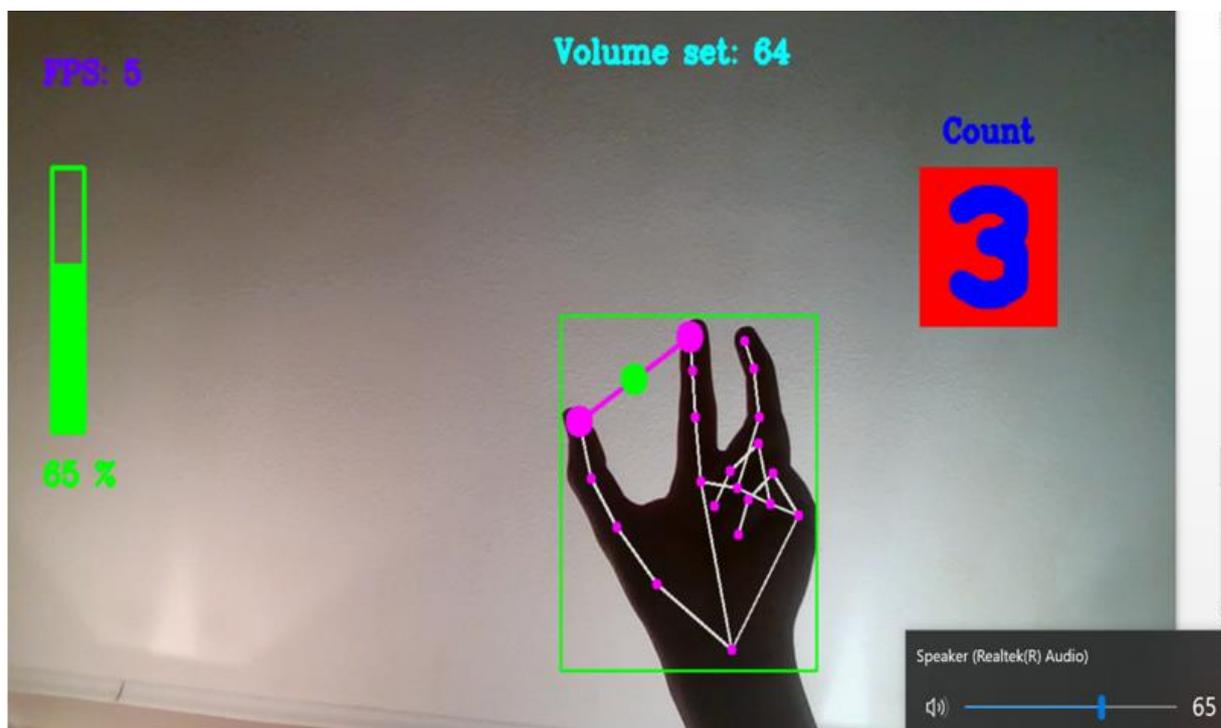


Рисунок 2 – Интерфейс разработанного приложения

1. Нахождение и отображение 21 ключевой точки руки в пространстве в режиме реального времени.
2. Подсчёт количества пальцев, которые присутствуют на изображении руки как показано на рисунке.
3. Изменение громкости компьютера посредством считывания расстояния между большим и указательным пальцами руки.
4. Управление курсором мыши компьютером при помощи следования за указательным пальцем руки.
5. Имитация клика ЛКМ при совмещении большого и указательного пальцев руки.

Распознавание жестов – одна из важнейших задач достижения взаимопонимания между человеком и компьютером. Руки и тело человека являются естественными манипуляторами и обладают большим числом степеней свободы, поэтому успешное решение задачи распознавания выполняемых ими жестов раскрывает перспективы для решения широкого круга прикладных задач.

### Список литературы

1. Machine Learning Mastery [Электронный ресурс]. – Режим доступа: <https://machinelearningmastery.com/what-is-computer-vision/>. – Дата доступа: 10.04.2023.
2. MediaPipe [Электронный ресурс]. – Режим доступа: Hand landmarks detection guide | MediaPipe | Google Developers. – Дата доступа: 10.04.2023.