

Всероссийский конкурс исследовательских и проектных работ
школьников «Высший пилотаж»

**ТЕХНОЛОГИЯ И СИСТЕМА ДЛЯ ОБУЧЕНИЯ ЯЗЫКУ ЖЕСТОВ НА ОСНОВЕ
КОМПЬЮТЕРНОГО ЗРЕНИЯ**

Проект

Направление «*Бизнес-информатика*»

Выполнил:

Растопшин Андрей

учащийся 10 класса

ГАУО АО ДО «Региональный школьный технопарк»

Руководитель:

Тараскин Дмитрий Владиславович

преподаватель

ГАУО АО ДО «Региональный школьный технопарк»

2023 г.

СОДЕРЖАНИЕ

<i>АННОТАЦИЯ</i>	3
ОБОСНОВАНИЕ АКТУАЛЬНОСТИ	3
<i>Рисунок 1 – Процентное соотношение глухих в мире (2022 г).</i>	3
<i>Рисунок 2 – Влияние языка жестов на слуховое восприятие.</i>	4
<i>Рисунок 3 – Распределение сурдопереводчиков по странам.</i>	5
ЦЕЛЬ И ЗАДАЧИ.....	6
АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ	6
<i>Таблица 1 – сравнение аналогов.</i>	7
<i>Рисунок 4 – метод работы обучения на сайте аналога.</i>	7
ОПИСАНИЕ РАЗРАБОТАННОГО РЕШЕНИЯ	7
<i>Рисунок 6 – примеры функции выделения.</i>	8
МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОЦЕНКИ КОРРЕКТНОСТИ ЖЕСТОВ.....	9
<i>Рисунок 7 – система координат ключевых точек в кадре.</i>	9
<i>Рисунок 8 – прямоугольный треугольник со сторонами a, b, c</i>	9
<i>Рисунок 9 – схематичная кисть и её ключевые точки.</i>	10
<i>Рисунок 10 – проблема первой версии алгоритма.</i>	11
<i>Рисунок 11 – новая проблема при первом методе решить проблему алгоритма.</i>	12
<i>Рисунок 12 – угол α на координатной оси.</i>	13
<i>Рисунок 13 – работа алгоритма программы проекта (белые линии показывают корректную часть жеста, красные линии, наоборот).</i>	13
ЗАКЛЮЧЕНИЕ	14
СПИСОК ИСТОЧНИКОВ.....	14

Ключевые слова: язык жестов, компьютерное зрение, обучение, кисть, жест, цвет, ключевые точки, python, камера, программа, глухие, дактиль.

АННОТАЦИЯ

Проект решает задачу по организации процесса самостоятельного изучения языка жестов в условиях отсутствия контакта с педагогом. Разработанная программа реализует интерактивный метод обучения языку жестов в режиме реального времени, координируя движение кисти человека, и, таким образом, ускоряет этот процесс и повышает его эффективность в сравнении с традиционными способами (печатное пособие, видео-уроки) и известными программными решениями, не имеющими возможностей оценки и направленной коррекции жеста. В нашей программе пользователь видит себя на экране устройства и непосредственно на этом изображении проецируются линии, показывающие какие фрагменты жеста совпадают с эталонными, а какие являются ошибочными. Приложение позволяет оценить степень ошибки – цвет линий прямо пропорционально зависит от процента неточности. Данные о положении кисти постоянно обновляются и по изменению цветовой палитры проецируемых линий пользователь может самостоятельно выявить и исправить свои недочеты. Таким образом, реализована полноценная модель «учитель-ученик». Такой подход кратно повышает эффективность и скорость изучения языка жестов, что ускоряет процесс преодоления социализации людей, страдающих глухотой.

ОБОСНОВАНИЕ АКТУАЛЬНОСТИ

По данным всемирной организации здравоохранения, более 5% населения мира — 364 миллионов взрослых и 36 миллионов детей — страдают от потери слуха (диаграмма распределения взрослых и детей в этой группе представлена на рисунке 1).



Рисунок 1 – Процентное соотношение глухих в мире (2022 г).

В результате объемного сравнительного исследования с участием 327 глухих учеников из школ, работавших в Германии на основе словесной речи, и школ с тотальной коммуникацией были получены показатели, в каком формате эффективнее воспринимается информация глухими людьми (таблица отношения формы лексического материала к процентным показателям представлена на рисунке 2).

Форма лексического материала	Показатели
Чтение с губ + восприятие на слух	30,2%
Чтение с губ + жесты + восприятие на слух	26,2%
Чтение с губ + жесты	26,2%
Только чтение с губ	20,8%
Только жесты	17,0%
Только восприятие на слух	12,6%
Жесты + восприятие на слух	10,2%

Рисунок 2 – Влияние языка жестов на слуховое восприятие.

Дети с нарушениями слуха попадают в специальную школу интернат, а взрослым приходится оплачивать за обучение языка жестов (в среднем 2000 рублей в час) и за обучение у сурдопедагога (в среднем 2500 рублей в день). В случае, когда от патологии страдает ребенок, расходы на лечение и содержание берут на себя родители, а если человек потеряет слух во взрослом возрасте, ему придется оплачивать обучение самостоятельно. Развивающие центры есть далеко не в каждом городе, а специалистов, которые в них работают, тоже относительно немного. Получается, что человек потенциально может столкнуться с двумя проблемами:

1. Отсутствие специализированного развивающего центра в его регионе или в его окрестностях.

2. Отсутствие материальных средств для оплаты работы специалистов, так как с приобретением подобной патологии, люди как правило лишаются основного источника дохода.

В нашей стране по статистике на 800 глухих людей приходится 1 сурдопереводчик, а значит примерно столько же и сурдопедагогов. Очевидно, что спрос сильно превышает предложение, а жесты оказывают существенное влияние на восприятие информации. Следовательно, многим людям приходится изучать язык жестов самостоятельно (количество и положение сурдопереводчиков в странах СНГ продемонстрирована на рисунке 3).



Рисунок 3 – Распределение сурдопереводчиков по странам.

Глухие люди во время обучения должны тренировать мимику, выучить жестовый язык, в котором каждый жест – это целое слово или понятие, и выучить дактиль, в котором каждый жест обозначает одну букву из алфавита. Кроме того, такая же задача по изучению стоит и перед их окружением для обеспечения коммуникации. Когда человек изучает язык жестов без педагога, у него нет обратной связи, насколько правильно или неправильно он показал жест, где именно он допустил ошибки. Это негативно сказывается на скорости и эффективности обучения. Также, большинство людей с нарушением слуха стараются пользоваться записками или телефонами, чтобы передать другим информацию, но это может не сработать в экстренных ситуациях (например, если у человека резко ухудшилось самочувствие) [6].

Решением данной проблемы является доступное бесплатное программное обеспечение, позволяющее быстро и точно обучать языку жестов. Приложение, на момент разработки, умеет распознавать любую букву из дактиля, получая изображение кисти человека через камеру. Чтобы исключить неопределенность, термин «язык жестов» будет означать исключительно дактиль.

Алгоритм работы приложения можно описать следующим образом:

1. Пользователю демонстрируется изображение камеры устройства и эталонного примера в виде жеста.
2. Изображение камеры устройства, при обнаружении кисти человека на нем, дополняется линиями, соединяющие ключевые точки.
3. Цвет каждой линии зависит от корректности положения соседних ключевых точек относительно эталонного примера. Красный цвет – данная часть руки неправильно расположена. Белый цвет – данная часть руки расположена корректно. Цвет линии меняет оттенок равномерно в зависимости от степени правильности текущей части жеста.
4. Эталонный пример, при корректном положении каждой ключевой точки меняет цвет линий на зеленый и запускает таймер на 2 секунды.
5. По окончании таймера, эталонный пример меняется на следующую букву по списку.

ЦЕЛЬ И ЗАДАЧИ

Цель проекта - разработка приложения для самостоятельного изучения языка жестов путем отслеживания положения кисти человека через веб-камеру устройства с реализацией обратной связи и направленной коррекции жестов в условиях отсутствия контакта со специализированным педагогом.

Задачи проекта:

1. Произвести сравнительный анализ существующих методов решения рассматриваемой проблемы.
2. В сотрудничестве с педагогом по обучению языку жестов составить техническое задание к функционалу проекта.
3. Описать ключевые параметры статического жеста и реализовать алгоритм их нахождения на изображении.
4. Разработать метод определения оценки корректности составленного жеста в результате сравнения с эталонным примером.
5. Реализовать обратную связь с пользователем по критерию оценки корректности жеста для достижения правильного результата.

АНАЛИЗ СУЩЕСТВУЮЩИХ РЕШЕНИЙ

В интернете было найдено 3 аналога данного проекта. «ASL Coach» и «Marlee Signs» - бесплатные приложения, предназначенные для изучения языка жестов на базовом уровне. И «fingerspelling.xyz» - бесплатный ресурс на сайте разработчиков «ASL». Однако есть два больших недостатка: обучение проходит только на английском языке жестов, что исключает их из списка возможных методов обучения людей, разговаривающих на русском языке, и пользователь не имеет обратной связи, насколько корректно были выстроены отдельные части жеста. Стоит обратить особое внимание, что отечественных аналогов приложения не существует несмотря на высокий спрос и очевидный недостаток педагогов в этой отрасли (характеристика аналогов представлена на таблице 1 и демонстрация работы последнего аналога на рисунке и 4).

Название	Способ обучения	Преимущества	Недостатки
ASL Coach	Демонстрация эталонного жеста в виде изображений	-Удобный интерфейс -Концентрация только на жест	-Приложение только на телефон -Отсутствие проверки корректности жеста пользователя -Приложение только для обучения английского языка жестов

Marlee Signs	Демонстрация эталонного жеста на видео	-Демонстрация жеста на реальной руке человека в формате видео-образца	-Приложение только на телефон -Отсутствие проверки корректности жеста пользователя -Приложение только для обучения английского языка жестов
Fingerspelling	Сравнение жеста эталонного примера и жеста с веб-камеры устройства с использованием нейросети	-Веб-ресурс для всех устройств -Проверка жеста пользователя через веб-камеру пользователя	-Отсутствие функционала коррекции пользователя при построении жеста -Для обучения доступен только английский язык жестов

Таблица 1 – сравнение аналогов.

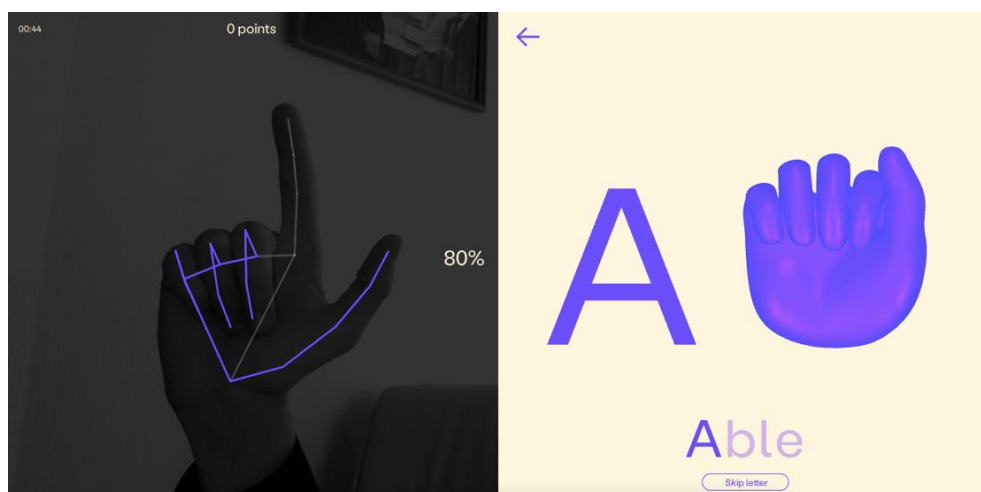


Рисунок 4 – метод работы обучения на сайте аналога.

ОПИСАНИЕ РАЗРАБОТАННОГО РЕШЕНИЯ

В качестве языка программирования был выбран Python. Основные преимущества этого языка заключаются в его легкой синтаксисе и в большом количестве библиотек [1]. Работа программы разбивается на 2 этапа: инициализация и непосредственно работа алгоритма.

На первом этапе проводится подключение всех библиотек:

1. OpenCV-python – библиотека компьютерного зрения, которая предназначена для анализа, классификации и обработки изображений. Широко используется в таких языках как C, C++, Python и Java [3]. Используется в программе для захвата изображения с камеры и его вывода на экран с измененными параметрами.

2. Mediapipe – фреймворк для создания кроссплатформенных мультимодальных конвейеров машинного обучения, который состоит из алгоритмов машинного обучения, классического компьютерного зрения и обработки медиа информации [4]. В Python это обширная библиотека со многими возможностями, такие как распознавание лиц, тела, положения зрачков, рук, объектов и др. В данном проекте используется только технология Hand Tracking (Отслеживание рук). Она позволяет определять заданное количество кистей на изображении, выделять их и возвращать координаты каждой ключевой точки кисти относительно разрешения входного изображения (схема расположения ключевых точек на кисти руки на рисунке 6 и пример выделения кистей руки на рисунке 7).

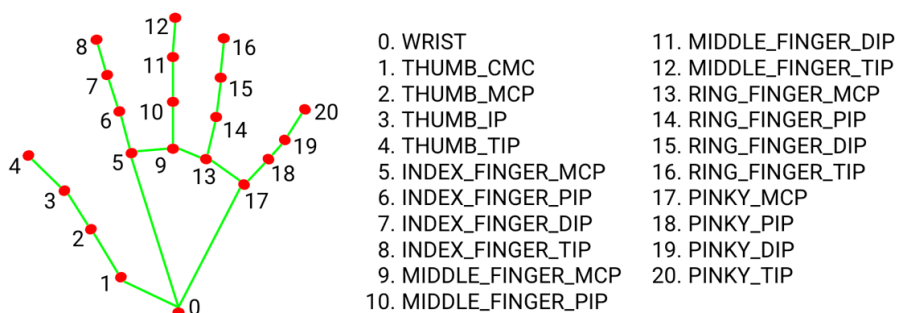


Рисунок 5 – расположение ключевых точек.



Рисунок 6 – примеры функции выделения.

Также помимо библиотек на первом этапе загружается список координат ключевых точек кисти каждого жеста, полученный из эталонных примеров на сайте всероссийского общества глухих.

Все эти действия происходят один раз непосредственно при каждом запуске программы и занимают порядка 3-6 секунд.

На втором этапе циклически работает основная часть программы, пока не выполнит поставленную задачу – пройтись по каждой букве из языка жестов. По окончании цикла, программа завершается.

МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОЦЕНКИ КОРРЕКТНОСТИ ЖЕСТОВ

Чтобы определить степень правильности жеста нужно ввести некоторые параметры, по значениям которых возможно сделать тот или иной вывод. Входными данными являются координаты ключевых точки кисти (21 пар координат по оси x и по оси y). Началом отсчёта является точка в левом верхнем углу экрана $(0;0)$, а единичным отрезком является пиксель изображения камеры (схема представлена на рисунке 8).

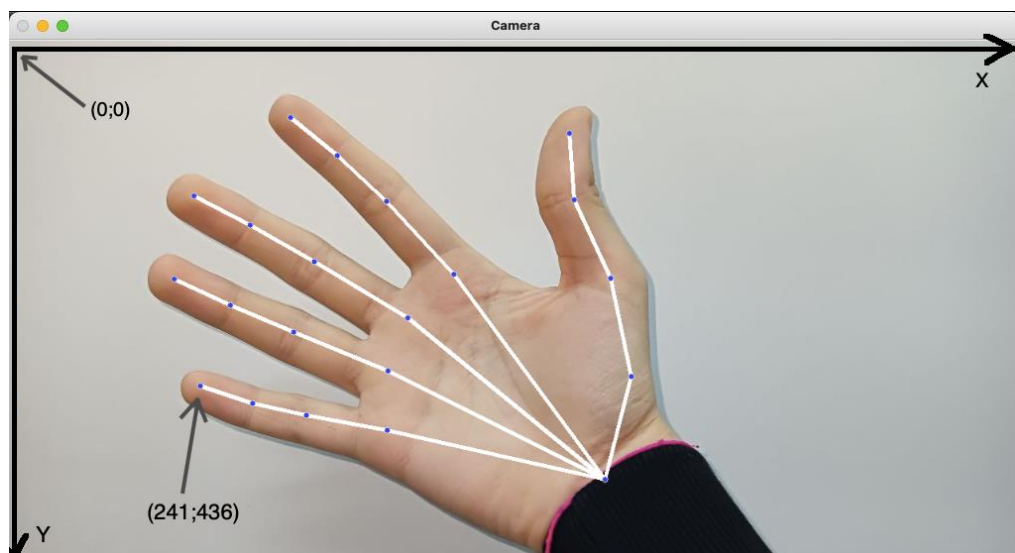


Рисунок 7 – система координат ключевых точек в кадре.

Представим, что отрезок между двумя соседними соединенными ключевыми точками является гипотенузой. Проекции гипотенузы по оси x и оси y являются соответственно её катетами. Мы получаем прямоугольный треугольник. Длины катетов находятся как модуль разницы координат двух соседних соединенных ключевых точек.

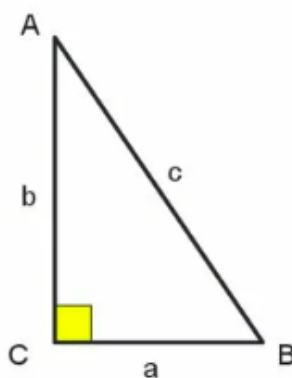


Рисунок 8 – прямоугольный треугольник со сторонами a , b , c

Формула нахождения длины катетов опираясь на рисунок 9:

$$a = |x_1 - x_2|$$

$$b = |y_1 - y_2|$$

Где x_1 – координата первой ключевой точки по оси x ;

x_2 – координата второй ключевой точки по оси x ;

y_1 – координата первой ключевой точки по оси y ;

y_2 – координата второй ключевой точки по оси y .

Теперь, зная длину каждого катета, можно найти длину самой гипотенузы через теорему Пифагора.

Формула теоремы Пифагора опираясь на рисунок 5:

$$c^2 = a^2 + b^2$$

Где a , b , c – стороны прямоугольного треугольника.

Проделав такие расчёты между всеми соседними соединёнными ключевыми точками, мы получаем длины каждого отрезка между ними (схема расчёта длин отрезков вручную представлена на рисунке 10).

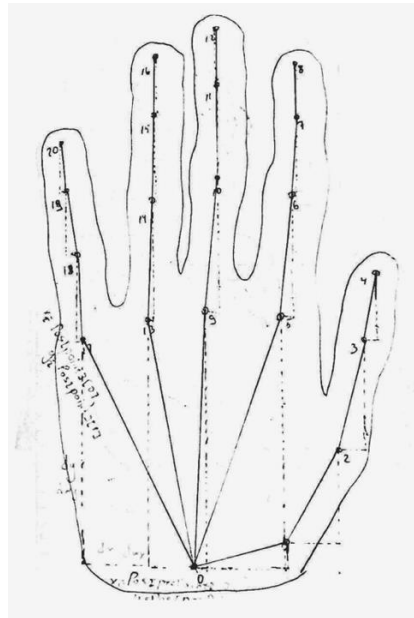


Рисунок 9 – схематичная кисть и её ключевые точки.

Проделав аналогичную работу на эталонном примере, мы можем сравнить длины гипотенуз на кадре с веб-камеры устройства с длинами гипотенуз на этом же примере.

Формула процентного соотношения длин гипотенуз:

$$\min(a_i, b_i) = \begin{cases} a_i, & a_i < b_i \\ b_i, & a_i \geq b_i \end{cases}$$

$$\max(a_i, b_i) = \begin{cases} a_i, & a_i > b_i \\ b_i, & a_i \leq b_i \end{cases}$$

$$p_i = \frac{\min(a_i)}{\max(b_i)} \times 100$$

Где p – список последовательных элементов ($p_1, p_2, p_3, \dots, p_{21}$);

min – функция по поиску минимального числа среди двух значений;

max – функция по поиску максимального числа среди двух значений;

a – список длин гипотенуз на кадре с веб-камеры устройства ($a_1, a_2, a_3, \dots, a_{21}$);

b – список длин гипотенуз на эталонном примере ($b_1, b_2, b_3, \dots, b_{21}$).

На первый взгляд мы получаем отличный результат. Если каждое p_i значение выше порогового числа, то текущий жест корректно построен и можно переходить к следующему. Однако появляется проблема – если кисть относительно веб-камеры повернуть на определенный угол, то жест по-прежнему останется корректно построенным, что на самом деле не так. (пример текущей проблемы на рисунке 11, гипотенуза 1 и гипотенуза 4 равны, но углы a и c не равны).

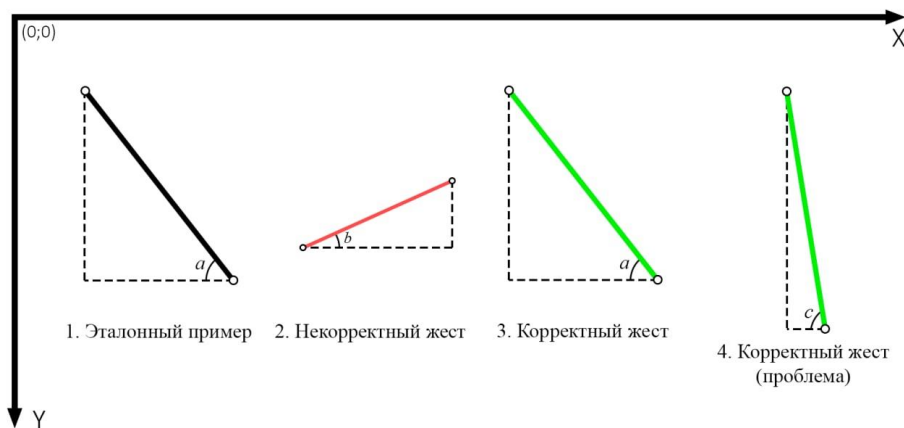


Рисунок 10 – проблема первой версии алгоритма.

Для утверждения, что выявленные прямоугольные треугольники являются подобными, требуется обратить внимание на угол.

Можно доказать путём аналогичного сравнения по формуле процентного соотношения длин гипотенуз, но уже углов между гипотенузами и катетами. Однако тогда появится ещё одна проблема – при повороте кисти, вектор катета (-ов) треугольника может поменять направление, а его длина не изменится (за счёт модуля в формуле), тем самым угол также останется неизменённым [2] (пример новой проблемы на рисунке 12, при текущих формулах в итоге получится что угол $a = b$, что неверно).

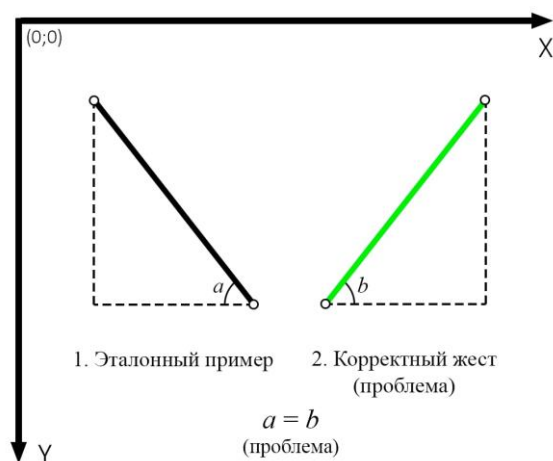


Рисунок 11 – новая проблема при первом методе решить проблему алгоритма.

Второй метод представляет более правдоподобный успешный исход решения проблемы. Представим, что наши гипотенузы являются векторами, у которых координата x и y равны длине их катетов на оси x и y соответственно. Тогда формулы нахождения угла между двумя векторами, у которых номера соседних соединённых ключевых точек равны, эталонного примера и кадра веб-камеры (демонстрация угла на рисунке 13):

$$\bar{a}_i \times \bar{b}_i = a_{ix} \times b_{ix} + a_{iy} \times b_{iy}$$

$$\cos \alpha = \frac{\bar{a}_i \times \bar{b}_i}{|\bar{a}_i| \times |\bar{b}_i|}$$

Где α – угол между векторами;

\bar{a} – список векторов на кадре с веб-камеры устройства ($a_1, a_2, a_3, \dots, a_{21}$);

\bar{b} – список векторов на эталонном примере ($b_1, b_2, b_3, \dots, b_{21}$).

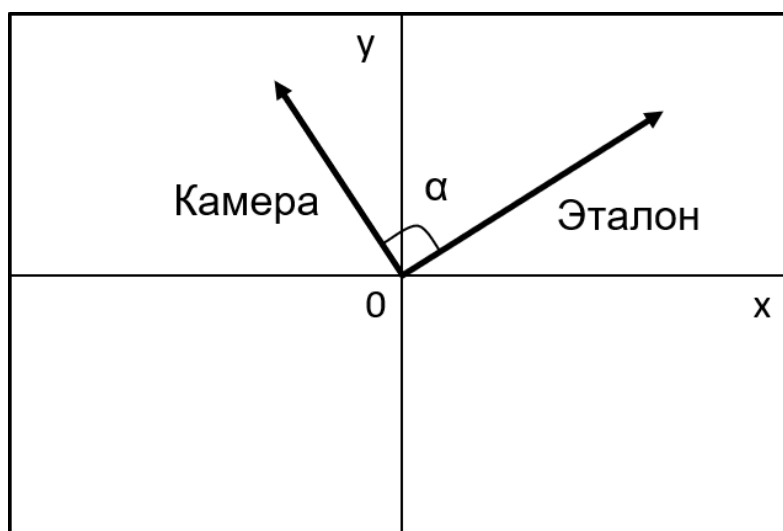


Рисунок 12 – угол α на координатной оси.

Для получения аналогичного списка p , только теперь для углов, произведём формулу:

$$m_i = \left(1 - \frac{c_i}{180}\right) \times 100$$

Где m – список последовательных элементов ($m_1, m_2, m_3, \dots, m_{21}$);

c – список полученных углов (α) между векторами ($c_1, c_2, c_3, \dots, c_{21}$).

Теперь, когда мы имеем два списка – p (проценты корректности длин гипотенуз) и m (проценты корректности углов) можем получить общий процент через формулу среднего геометрического двух чисел [5]:

$$r_i = \sqrt{p_i \times m_i}$$

Где r – список последовательных элементов ($r_1, r_2, r_3, \dots, r_{21}$).

В итоге мы получаем приемлемый результат. Теперь, если каждое r_i значение выше порогового числа (на данный момент значения 60), то текущий жест корректно построен и можно переходить к следующему. На данный момент алгоритм работает именно по такому принципу, но в процессе тестирования постоянно дорабатывается, т.к. периодически возникают разные нестандартные ситуации, где модель может давать погрешность (демонстрация работы алгоритма на рисунке 14).

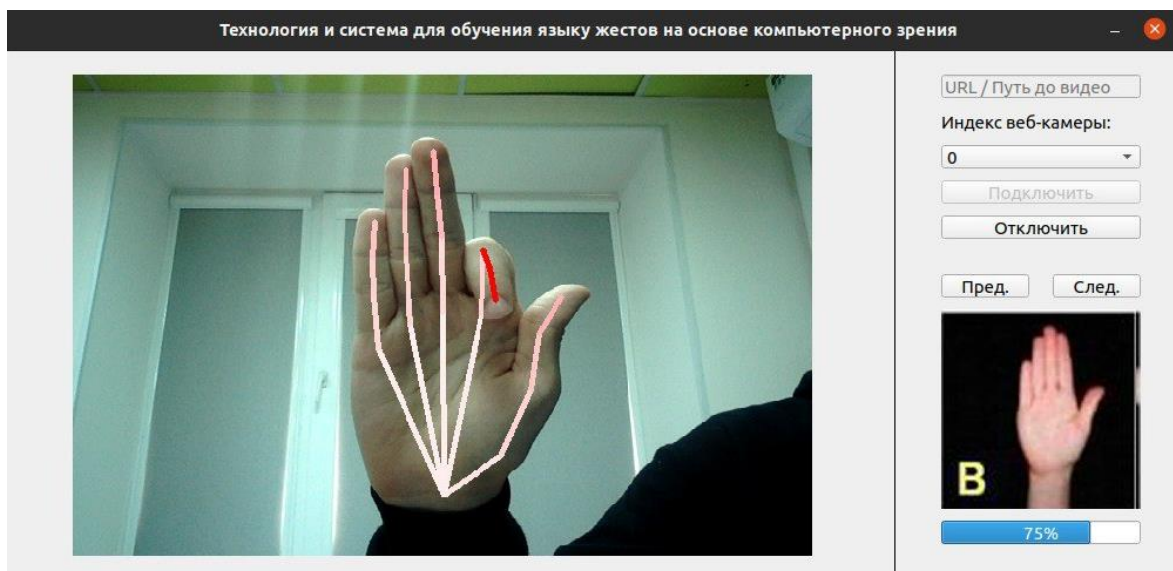


Рисунок 13 – работа алгоритма программы проекта (белые линии показывают корректную часть жеста, красные линии, наоборот).

ЗАКЛЮЧЕНИЕ

На сегодняшний день реализовано определение статичных жестов, обозначающих отдельные буквы. Распознавание происходит в режиме реального времени, изображение сравнивается с эталонным, показываются сходства и различия. Уникальность нашей работы состоит в том, что мы реализовали полноценную модель «учитель-ученик», которая не представлена ни в одном из аналогов. Как правило, обучение происходит с педагогом, но как уже было сказано, далеко не во всех регионах можно найти такого узкого специалиста и стоимость подобных услуг довольно высока. В случае, если педагога найти не получилось людям приходится пользоваться видео-уроками или обучающими пособиями, но очевидно, что такой подход не подразумевает наличие обратной связи, а значит снижается скорость и эффективность обучения. Наш проект имеет социальную направленность, это решение должно помочь социализироваться людям с нарушением слуха независимо от региона и финансового обеспечения.

В перспективе предполагается:

1. Реализовать базу данных со всеми словами из языка жестов и дополнить алгоритм на возможность считывать мимику и движение рук. Это необходимо для анализа более сложных и продвинутых жестов.
2. От эксперта поступила рекомендация добавить ряд упражнений на координацию рук, пальцев, и на развитие мимических мышц лица.
3. Добавить режим «соревнований», где нужно как можно быстрее указать жест, требующий на экране устройства, чем твой оппонент.
4. Разработать переводчик жестов в режиме реального времени.
5. Провести реальные испытания с привлечением целевой группы слабослышащих людей и их педагогов, чтобы оценить, насколько удобен такой формат освоения материала.
6. Реализовать отображение жеста не только в двухмерном формате, но и в объёме.
7. Протестировать возможность корректировать пользователя не только через экран, но и с использованием физических объектов или устройств (например, перчатка с датчиками, надетая на руку перед тренировкой).

СПИСОК ИСТОЧНИКОВ

1. Д. Ю. Федоров, Программирование на языке высокого уровня Python: учебное пособие для прикладного бакалавриата / Д. Ю. Федоров. – 2Й-е изд., перераб. и доп. – Москва: Издательство: Юрайт, 2019. – 161 с.

2. Г. В. Калашников, С. В. Макеев, Основы тригонометрии: учебное пособие. Издательство: Воронежский государственный университет инженерных технологий (Воронеж), 2020. – 74 с.
3. Joseph Howse, OpenCV Computer Vision with Python / #1326 in Computer Vision & Pattern Recognition. Издательство: Packt Publishing, 2013. – 31 с.
4. Arpita Halder, Akshit Tayade, Real-time Vernacular Sign Language Recognition using MediaPipe and Machine Learning. Издательство: International Journal of Research Publication and Reviews, 2011. – 12 с.
5. Т. М. Татарникова, Анализ данных: учебное пособие по теоретическим и алгоритмическим основам анализа данных. Издательство: Санкт-Петербургский государственный экономический университет (Санкт-Петербург), 2018. – 58 с.
6. Ф. А. Игебаева, Язык жестов в деловом общении: научная статья о жестикуляции. Издательство: журнал novainfo.ru, 2016. – 274 с.