

УДК 004.4+004.048

ПРОГРАММНОЕ СРЕДСТВО ПЕРЕВОДА ЯЗЫКА ЖЕСТОВ В РЕЖИМЕ РЕАЛЬНОГО ВРЕМЕНИ

Бобров Б.С., студент, Потоцкий Д.С., магистрант

*Белорусский государственный университет информатики и радиоэлектроники,
Институт информационных технологий,
г. Минск, Республика Беларусь*

Потоцкий Д.С. – ассистент каф. ИСиТ

Аннотация. Данная работа является результатом исследования проблем людей с ограничениями в общении и разработки программного средства для решения исследованных проблем. Подробно описаны подходы и технологии, используемые для построения программного средства для работы в режиме реального времени с использованием сверточных сетей.

Ключевые слова. искусственный интеллект, сверточные сети, жесты, компьютерное зрение, OpenCV, Tensorflow, mediapipe.

Введение. Сфера поддержки людей с ограничениями в области речи и слуха переживает значительные трансформации в свете стремительного технологического развития. Современные технологии, такие как системы распознавания и синтеза речи, а также специализированные мобильные приложения, играют неотъемлемую роль в улучшении коммуникации и обмене информацией для данной группы людей. Эти инновационные решения в значительной степени способствуют преодолению трудностей и барьеров, с которыми сталкиваются индивиды с ограничениями в области речи и слуха.

Эти инновации не только улучшают сам процесс коммуникации, но и играют важную роль в содействии социальной интеграции. Технологический прогресс предоставляет новые перспективы для разработки инструментов, способствующих более активному участию в обществе и улучшению качества жизни данной аудитории. В результате, технологии, ориентированные на поддержку людей с ограничениями в области речи и слуха, становятся не только средством преодоления барьеров, но и ключевыми компонентами в стремлении создать более инклюзивное и поддерживающее общество для всех.

Современные средства коммуникации, включая виртуальную и дополненную реальность, а также системы видеозвонков, играют важную роль в создании инклюзивных сред для обучения и общения. Эти инновационные технологии открывают новые горизонты, обеспечивая возможность взаимодействия в виртуальных пространствах и обеспечивая доступ к образованию и общению для всех. Однако, несмотря на свой потенциал, эти средства не всегда гарантируют более естественные и доступные способы общения для людей с ограничениями.

Таким образом, несмотря на значительные достижения в области коммуникационных технологий, существует потребность в дальнейших исследованиях и разработках, направленных на создание более натуральных и универсальных средств общения, способных удовлетворять потребности людей с различными ограничениями.

Для повышения инклюзивности и обеспечения более эффективного общения для людей с ограничениями в области коммуникации, эффективными инструментами могут стать программные средства, приложения, модули и плагины, специально разработанные для улучшения взаимодействия между людьми, владеющими языком жестов, и теми, кто не владеет.

Эти технологии могут предоставлять различные функции, такие как распознавание и интерпретация языка жестов, автоматический перевод на различные языки, визуальные подсказки и поддержку альтернативных методов общения. Программные средства и приложения могут быть спроектированы так, чтобы создавать барьеры между разными формами коммуникации, обеспечивая более натуральное и доступное взаимодействие для пользователей с разными коммуникативными потребностями.

В настоящее время не существует готового и полностью поддерживаемого программного средства, предназначенного для эффективного решения данной проблемы. Однако, с другой стороны, наблюдается стремительный рост в области технологий компьютерного зрения, нейронных сетей и других передовых технологий. Этот динамичный прогресс создает уникальные возможности для разработки нового программного средства, способного эффективно решать данную проблему.

Технологии компьютерного зрения, вместе с развитием нейронных сетей, открывают новые перспективы в области распознавания и интерпретации визуальной информации. Это позволяет создавать более сложные и интеллектуальные системы, способные эффективно взаимодействовать с окружающей средой и выполнять сложные задачи, такие как распознавание жестов, языка жестов и других визуальных коммуникационных элементов.

В свете этих технологических достижений сегодня представляет собой оптимальный момент для разработки программного средства, направленного на решение указанной проблемы. Наличие

передовых технологических средств и растущий интерес к инновационным решениям создают благоприятные условия для успешной разработки инновационного и эффективного решения, способного значительно улучшить сферу коммуникации для тех, кто сталкивается с ограничениями.

Основная часть. Для разработки программного средства был выбран язык программирования Python из-за его простоты разработки, обширного набора исследований, образовательных материалов и широкой аудитории разработчиков. Python предоставляет удобное окружение для создания и тестирования программ, а также обеспечивает быстрое прототипирование благодаря своему интуитивному синтаксису.

Для реализации компьютерного зрения были использованы следующие технологии:

1. OpenCV – библиотека алгоритмов компьютерного зрения, обработки изображений и численных алгоритмов общего назначения с открытым исходным кодом. OpenCV реализована на языках программирования C/C++, но также имеет доступные обвязки для Python и других языков. Эта библиотека предоставляет обширный инструментарий для анализа изображений, распознавания объектов и других задач компьютерного зрения [1].

2. MediaPipe – пакет предоставляет набор инструментов для реализации компьютерного зрения. Он работает с двумерными массивами для анализа, поиска и разметки объектов на изображении. MediaPipe также упрощает создание мультимодальных конвейеров машинного обучения, предоставляя инструменты для обработки данных и взаимодействия с различными модулями [2].

3. TensorFlow – библиотека для машинного обучения, предоставляющая различные технологии для обучения искусственного интеллекта в решении разнообразных задач. Изначально разработанная для Python, TensorFlow широко используется в области машинного обучения и глубокого обучения. Он предоставляет удобные средства для создания, обучения и развертывания моделей машинного обучения [3].

Выбор этих технологий обусловлен их функциональностью, эффективностью и широкой поддержкой в сообществе разработчиков, что обеспечивает оптимальные условия для разработки программного средства, направленного на решение конкретной проблемы в области компьютерного зрения.

Сверточная нейронная сеть в данной системе обучается на основе предварительно записанных, загруженных и размеченных кадров, охватывающих различные условия. Прежде чем подвергнуться обучению, изображения программно корректируются, что позволяет оптимизировать и стандартизировать входные данные. Затем эти скорректированные изображения передаются в библиотеку MediaPipe, совместно с OpenCV, где формируется двумерный массив точек, представляющих положение рук или руки в пространстве.

В процессе обучения модель TensorFlow использует эти размеченные данные для выявления паттернов и особенностей, связанных с различными жестами и движениями рук в различных условиях. Таким образом, сеть учится автоматически извлекать признаки из изображений, необходимые для точного распознавания положения рук.

После завершения обучения модели часть данных, которые не использовались в процессе обучения, отделяется и используется для тестирования обученной модели. Этот этап тестирования позволяет оценить эффективность модели на новых данных и дает представление о ее обобщающей способности, то есть способности обрабатывать неизвестные данные, которые не были представлены в ходе обучения.

Работа программного средства следует определенному алгоритму, обеспечивающему эффективный процесс обработки и перевода жестовых данных:

1. Захват изображения с веб-камеры: Программное средство начинает свою работу с захвата изображения с веб-камеры устройства, на котором оно запущено. Этот этап предоставляет программе поток видео, который будет анализироваться для выявления жестов и движений рук пользователя.

2. Анализ каждые 30 кадров: В ходе работы программы происходит анализ видеопотока с периодичностью каждые 30 кадров. Этот временной интервал, хотя может различаться в зависимости от характеристик устройства, составляет от 0.5 до 1.5 секунд. В течение этого интервала программа обрабатывает видеоданные, выделяя жесты и движения рук.

3. Перевод координат рук в слова или буквы: На основе анализа изображения и выделенных координат рук, программа приступает к переводу этой информации в текстовые символы, слова или фразы. Этот этап представляет собой важный шаг в процессе коммуникации, где жесты пользователя преобразуются в понятный для программы формат.

4. Выбор наилучшего перевода: После получения текстовой информации из координат рук, программа выбирает наилучший перевод. Это может включать в себя использование моделей машинного обучения или других алгоритмов, направленных на выбор наиболее вероятного и соответствующего значения для переданных координат.

5. Отображение результатов пользователю: Выбранный наилучший перевод отображается пользователю на экране. Это обеспечивает мгновенный обратный отклик и позволяет пользователю видеть результаты коммуникации в режиме реального времени.

6. Повторение процесса до закрытия программы: Программа продолжает повторять вышеописанный процесс анализа, перевода и отображения результатов до момента закрытия программного средства пользователем.

Схема работы алгоритма перевода языка жестов представлена на рисунке 1.

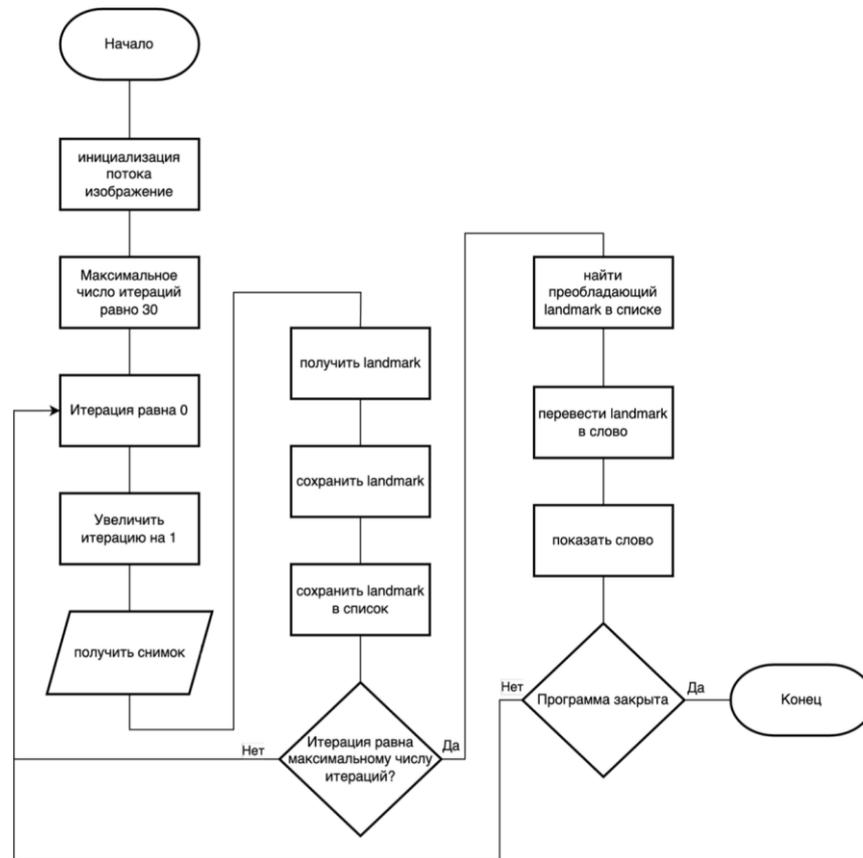


Рисунок 1 - схема работы алгоритма перевода языка жестов

Заключение. Разработанная система на базе Python, OpenCV, MediaPipe и TensorFlow представляет собой эффективный инструмент, способный переводить жесты рук в текстовые символы в реальном времени. Обучение нейронной сети на основе предварительно размеченных данных позволяет системе точно распознавать положение рук в различных условиях.

Список использованных источников:

1. OpenCV начальная информация [Электронный ресурс] – Режим доступа: <https://github.com/opencv/opencv/wiki>. – Дата доступа: 04.01.2024.
2. MediaPipe определение рук [Электронный ресурс] – Режим доступа: <https://developers.google.com/mediapipe>. – Дата доступа: 04.01.2024.
3. Tensor Flow обзор [Электронный ресурс] – Режим доступа: <https://www.tensorflow.org/overview>. – Дата доступа: 04.01.2024.

UDC 004.4+004.048

REAL-TIME GESTURE LANGUAGE TRANSLATION SOFTWARE

Bobrov B.S., Potosky D.S.

*Institute of Information Technologies of the Belarusian State University of Informatics and Radioelectronics,
Minsk, Republic of Belarus*

Potosky D.S. – Assistant

Annotation. This work is the result of researching the issues faced by individuals with communication restrictions and developing a software tool to address the identified problems. Approaches and technologies employed in constructing the real-time software tool using convolutional neural networks are detailed.

Keywords. Artificial intelligence, convolutional networks, gesture translation, computer vision, OpenCV, TensorFlow, MediaPipe.