

УДК 004.514

СИСТЕМА РАСПОЗНАВАНИЯ ЖЕСТОВОГО ЯЗЫКА НА ОСНОВЕ АНАЛИЗА ВИДЕОПОТОКА В РЕАЛЬНОМ ВРЕМЕНИ

А. С. ГОЛЯС

Научный руководитель Д. В. РОГОЛЕВ, канд. физ.-мат. наук
Белорусско-Российский университет
Могилев, Беларусь

Эффективная коммуникация играет важную роль в современном мире, однако для людей с ограниченными возможностями слухового восприятия она остаётся серьёзной проблемой. По данным всемирной организации здравоохранения, более 466 млн человек имеют нарушения слуха и их основным средством общения является жестовый язык.

Представленное приложение Gestotalk, разработанное на Python, призвано преодолеть эти барьеры, используя технологии компьютерного зрения и машинного обучения для автоматического распознавания жестов. Это решение способствует созданию инклюзивной среды, упрощая взаимодействие между людьми с разными возможностями слуха.

Gestotalk использует библиотеку компьютерного зрения OpenCV для захвата видеопотока с веб-камеры в реальном времени. Для детекции рук и их ключевых точек применяется MediaPipe Hand Landmarker – инструмент на основе машинного обучения, разработанный Google. Он определяет положение рук и для каждой обнаруженной руки возвращает координаты трехмерной ключевой точки (landmarks). Эти точки представляют собой суставы и кончики пальцев, а также запястье, формируя детальный «скелет» кисти.

Приложение Gestotalk имеет собственную библиотеку жестов в формате JSON, где каждая буква русского дактильного алфавита (кроме ё, й, щ) представлена с помощью 21 ручного ориентира, представляющих собой трёхмерные координаты. Они нормализуются относительно запястья. Такой подход обеспечивает устойчивость к изменению положения руки в кадре.

Распознавание жестов происходит в несколько этапов. Сначала видеопоток обрабатывается с помощью OpenCV, затем MediaPipe Hand Landmarker выделяет ключевые точки руки. Полученные координаты сравниваются с эталонными значениями из библиотеки с использованием метрики евклидова расстояния. Наилучшее совпадение, не превышающее заданный порог, считается распознанным жестом. На рис. 1 продемонстрирована блок-схема работы приложения Gestotalk.

Приложение поддерживает многопоточную обработку, позволяя одновременно анализировать жесты обеих рук. При обнаружении жеста «ввод» левой руки распознанный в тот же момент жест правой руки записывается в текстовое поле. При необходимости текст может быть озвучен, записан и сохранён в форматах .txt и MP3. Распознанные жесты последовательно формируют текст, который отображается в графическом пользовательском интерфейсе (GUI), созданном с помощью PyQt5.

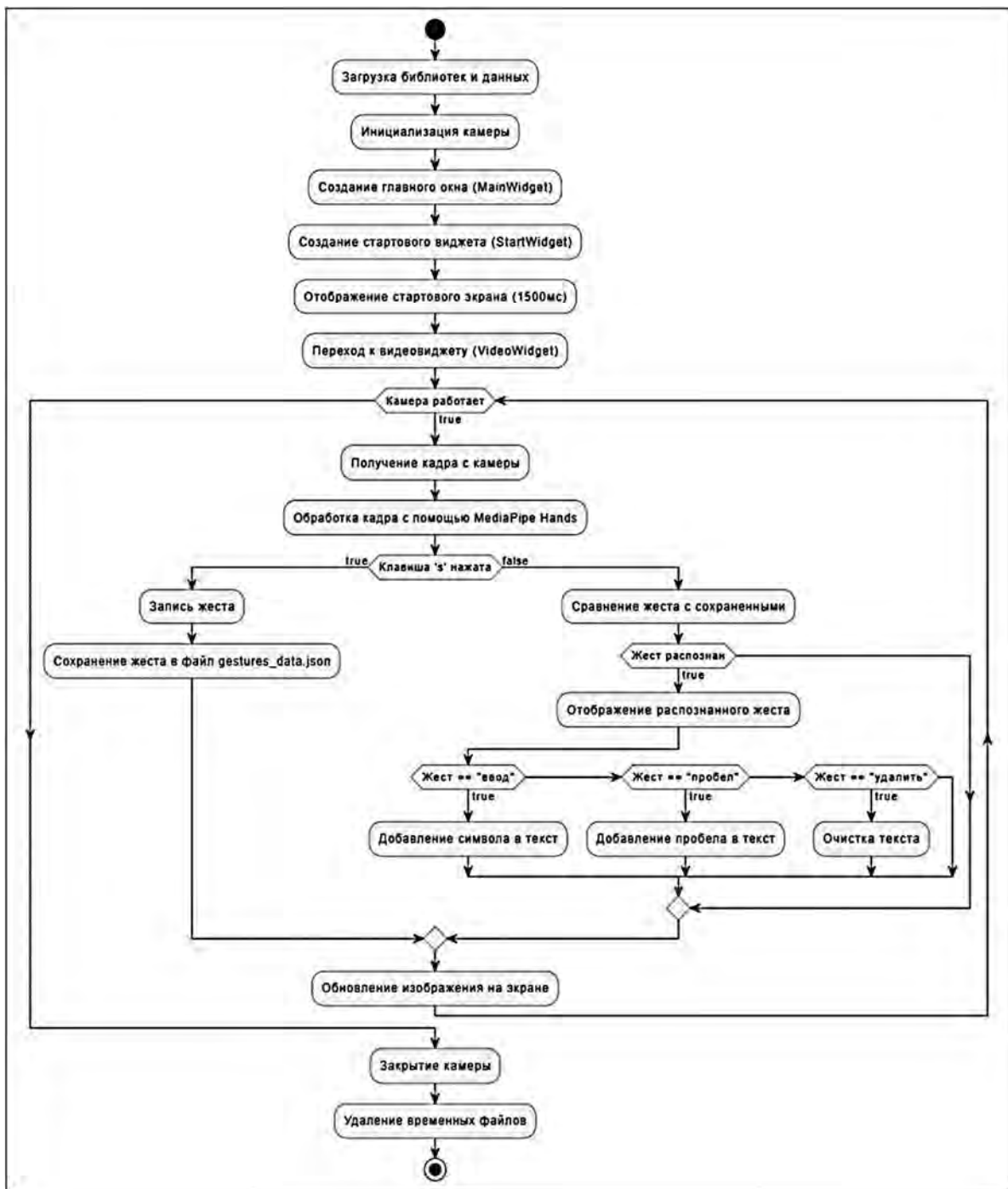


Рис. 1. Блок-схема работы приложения Gestotalk

Gestotalk представляет собой комплексное решение для распознавания жестового языка, сочетающее современные технологии компьютерного зрения и машинного обучения. Перспективы развития системы включают интеграцию временных моделей для распознавания динамических жестов, улучшение устойчивости к изменению освещенности и фона, разработку мобильной версии с оптимизированными нейросетевыми моделями, а также добавление поддержки дополнительных жестовых алфавитов. Приложение способствует улучшению коммуникации для людей с нарушениями слуха, делая взаимодействие более доступным и удобным.