

**Двадцать третья международная конференция
«СОВРЕМЕННЫЕ ПРОБЛЕМЫ ДИСТАНЦИОННОГО ЗОНДИРОВАНИЯ ЗЕМЛИ ИЗ КОСМОСА»**

Военно-космическая академия имени А.Ф. Можайского

**Виртуальный ассистент для геоинформационных сервисов на
базе аудиовизуальных технологий искусственного интеллекта**



**доктор технических наук Григорьев Андрей Николаевич
кандидат технических наук Коршунов Денис Сергеевич
Дементьев Денис Сергеевич**

Современные геоинформационные системы занимают ключевое место в управлении пространственными данными, их анализе и визуализации, обеспечивая основу для принятия решений. Однако традиционные способы взаимодействия с ГИС через клавиатуру, мышь и стандартные графические интерфейсы зачастую оказываются недостаточно удобными и оперативными. Возрастающий объём пространственных данных и усложнение аналитических задач формируют потребность в создании новых, более интуитивных и естественных способов взаимодействия с системами, способных существенно повысить эффективность работы специалистов.

Одним из наиболее перспективных направлений решения этой проблемы является разработка мультимодальных ассистентов, объединяющих голосовое и жестовое управление. Интеграция подобных технологий в ГИС позволяет не только повысить скорость и точность выполнения операций, но и обеспечить автономность работы за счёт офлайн-распознавания речи и анализа жестов в реальном времени. Внедрение такого подхода в платформу QGIS делает её более адаптивной к современным требованиям, расширяет возможности применения в профессиональной деятельности и образовательных проектах, а также открывает новые перспективы для автоматизации процессов пространственного анализа и принятия решений в условиях ограниченных ресурсов.



Идея исследования состоит в создании мультимодального ассистента для геоинформационных систем, который объединяет возможности офлайн-распознавания речи и анализа жестов для управления функционалом QGIS. Такое решение направлено на формирование нового уровня взаимодействия с пространственными данными, где голосовые команды и жесты служат естественными инструментами работы с картографическим интерфейсом. В основе концепции лежит интеграция нейросетевых технологий Vosk и MediaPipe в модульную архитектуру, обеспечивающую масштабируемость и возможность адаптации под различные прикладные задачи — от анализа геоданных и полевых исследований до образовательных и демонстрационных целей.

Целью исследования является разработка и внедрение мультимодального ассистента для геоинформационных систем на базе QGIS, обеспечивающего интуитивное и эффективное взаимодействие с пространственными данными посредством голосового и жестового управления. Реализация данной цели направлена на повышение удобства, скорости и автономности работы пользователей с геоинформационными системами, а также на расширение функциональных возможностей QGIS за счёт интеграции современных технологий офлайн-распознавания речи и компьютерного зрения.



Нейросетевые архитектуры детекции ключевых точек на ладони и обработки последовательности жестов

3

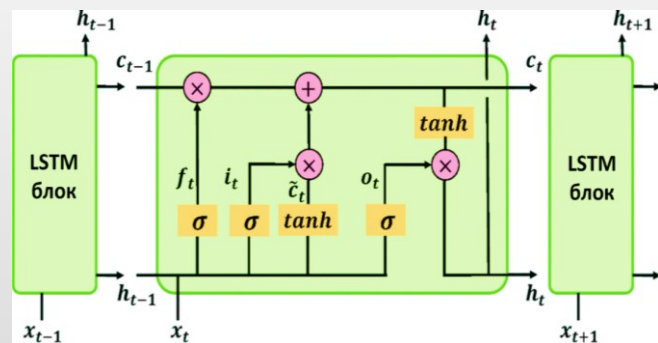


Рис. 1 «Архитектура LSTM»

Архитектура LSTM применяется для анализа последовательностей движений ключевых точек кисти. Она учитывает временные зависимости в жестах, что делает возможным их корректную интерпретацию и различение даже при сложных комбинациях движений. В связке с SSD сеть LSTM обеспечивает надёжное распознавание жестов и преобразование их в команды пользователя для управления ГИС.

Архитектура SSD используется в работе для быстрого и точного обнаружения ладоней на изображении. Она позволяет в реальном времени выделять области интереса, даже в условиях сложного фона или переменного освещения. Благодаря этому система получает координаты ладоней, которые далее передаются для уточнения положения ключевых точек кисти.

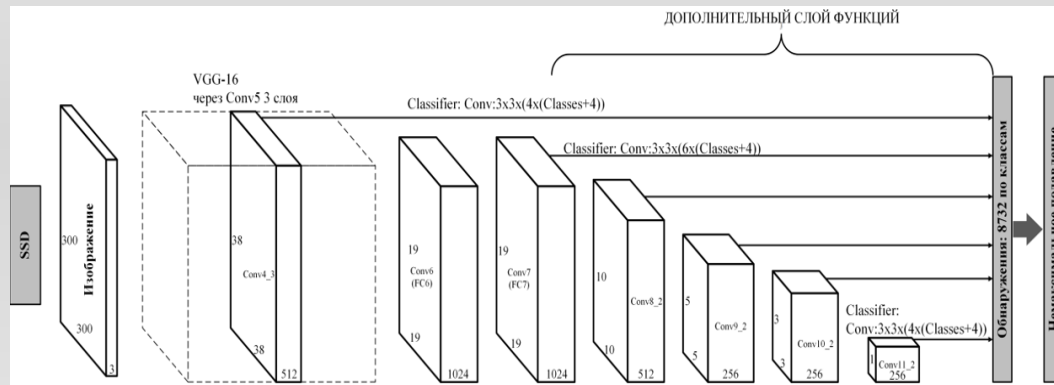


Рис. 2 «Архитектура SSD»



Нейросетевые архитектуры детекции ключевых точек на ладони и обработки последовательности жестов

4

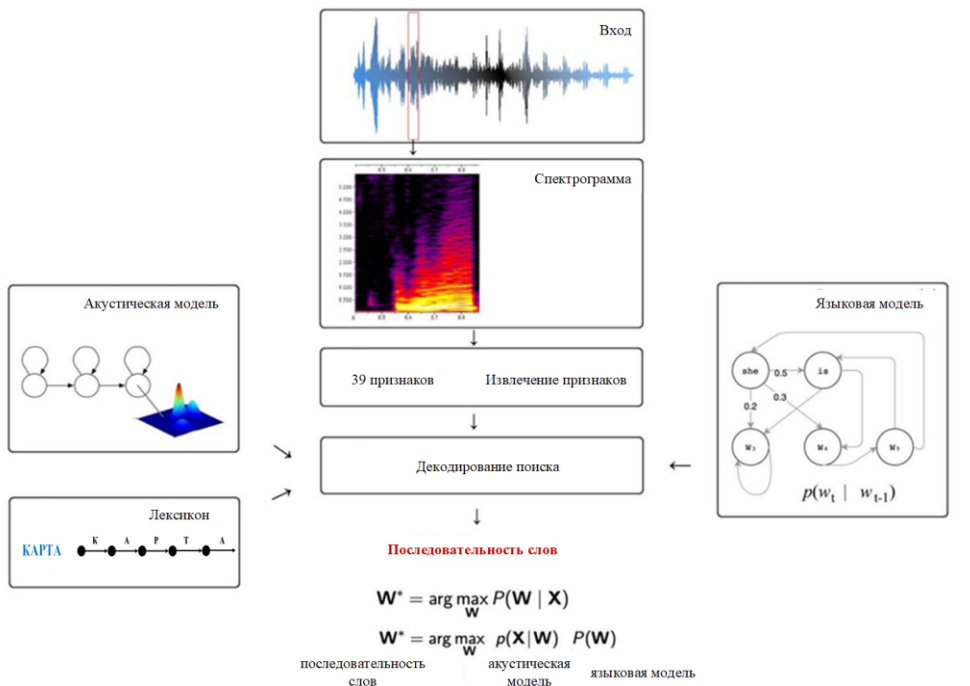


Рис. 3 «Архитектура модели распознавания речи»

В работе используется офлайн-модель распознавания речи Vosk, основанная на акустических и нейросетевых архитектурах. Она преобразует голосовые команды пользователя в текст и интерпретирует их как действия в ГИС. Благодаря использованию алгоритма Витерби и методов сглаживания вероятностных распределений достигается высокая точность распознавания даже в условиях шума. В связке с жестовым управлением голосовая модель обеспечивает полноценное мультимодальное взаимодействие с системой.

Результаты сравнения точных и эвристических методов комбинаторики в зависимости от количества съемочных участков

5

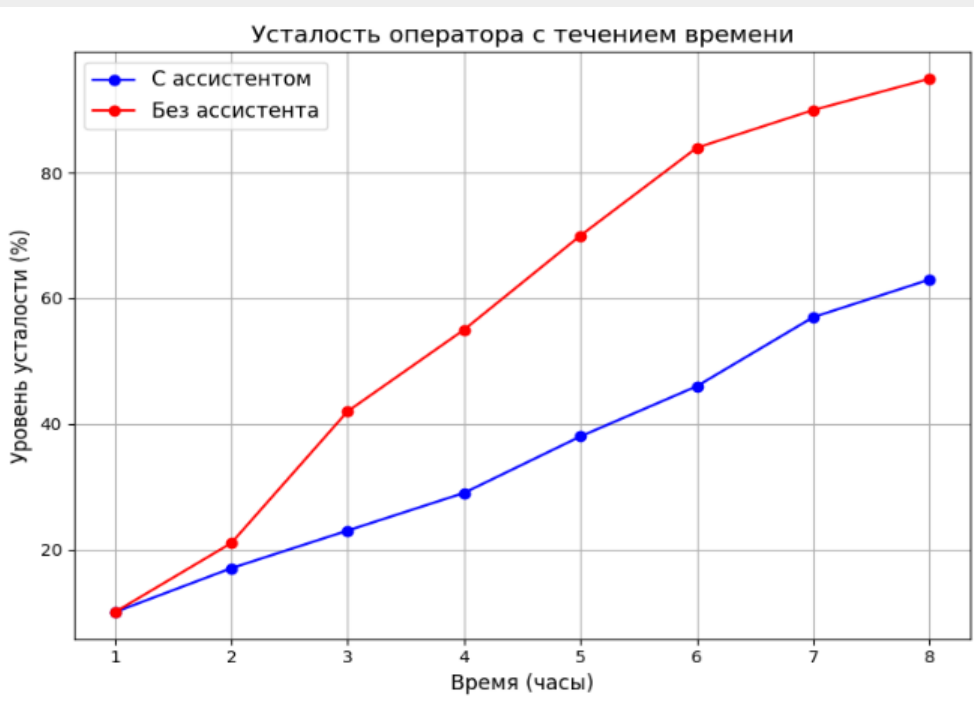


Рис. 4 «Зависимость времени работы оператора от усталости»

Главным показателем эффективности разработанного виртуального ассистента выступает уровень усталости оператора в процессе работы. Представленные на графике данные показывают, что при использовании ассистента динамика накопления усталости существенно ниже по сравнению с традиционным взаимодействием с системой. Это объясняется снижением когнитивной нагрузки и упрощением выполнения операций за счёт голосового и жестового управления.

Таким образом, в результате проведённого исследования была разработана и реализована концепция мультимодального ассистента для геоинформационных систем, основанная на сочетании голосового и жестового управления. Использование библиотеки Vosk обеспечило возможность офлайн-распознавания речи с высокой точностью и устойчивостью к шумам, а интеграция технологий MediaPipe позволила реализовать надёжное и удобное управление картой с помощью жестов. Созданный модуль успешно интегрирован в среду QGIS, что подтверждает практическую применимость предложенного подхода для работы с пространственными данными.

Полученные результаты демонстрируют, что мультимодальное управление существенно повышает удобство и эффективность взаимодействия с ГИС, снижает когнитивную нагрузку на пользователя и позволяет использовать систему в условиях ограниченного интерфейса и отсутствия интернет-соединения. Разработанный ассистент может быть использован в полевых исследованиях, при решении градостроительных и инфраструктурных задач, в образовательной и демонстрационной деятельности. В дальнейшем развитие данного направления предполагает расширение набора команд и жестов, внедрение адаптивных алгоритмов распознавания и интеграцию с интеллектуальными средствами анализа пространственных данных, что позволит ещё более повысить уровень автоматизации и эффективности применения ГИС.



1. Бондаренко, А. В. Геоинформационные системы: теория и практика / А. В. Бондаренко. – Москва: КДУ, 2020. – 352 с.
2. Ушаков, Д. В., Панова, А. С. Человеко-машинные интерфейсы: современные концепции и технологии. – Москва: Юрайт, 2021. – 280 с.
3. Котов, В. Е. Рекуррентные нейронные сети и их применение в задачах анализа последовательностей. // Труды МФТИ. – 2019. – Т.11. №1. – С. 55–67.