



**МАТЕМАТИЧЕСКОЕ И ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ВЫЧИСЛИТЕЛЬНЫХ СИСТЕМ,
КОМПЛЕКСОВ И КОМПЬЮТЕРНЫХ СЕТЕЙ/MATHEMATICAL SOFTWARE FOR COMPUTERS,
COMPLEXES AND COMPUTER NETWORKS**

DOI: <https://doi.org/10.60797/IRJ.2026.168.49> EDN: EXYUMY

**РАЗРАБОТКА ПРОТОТИПА СИСТЕМЫ КОМПЬЮТЕРНОГО ЗРЕНИЯ ДЛЯ АССИСТИВНЫХ УСТРОЙСТВ
ЛЮДЕЙ С НАРУШЕНИЯМИ ЗРЕНИЯ**

Научная статья

Некрасов Е.А.^{1,*}, Дегтярь А.В.², Маркина Т.А.³, Государев И.Б.⁴

^{1, 2, 3, 4} Университет ИТМО, Санкт-Петербург, Российская Федерация

* Корреспондирующий автор (evgeniy[at]nekrasov.pw)

Предложена: 13.12.2025; Принята: 12.03.2026; Опубликовано: 17.06.2026

Аннотация

Разработка эффективных навигационных систем для людей с нарушениями зрения требует перехода на современные алгоритмы машинного зрения, способные минимизировать временные задержки при обработке визуальных данных. Использование устаревших моделей создает угрозу безопасности пользователя из-за медленного отклика системы. Цель исследования заключается в создании архитектуры интеллектуального ассистента на базе актуального алгоритма детектирования, обеспечивающего информирование в реальном времени без задержек на постобработку. Методы: в работе использована нейросетевая модель десятого поколения (YOLOv10), исключающая внешние процедуры фильтрации прогнозов. Методика включает беспроводную трансляцию видеопотока с микроконтроллера на вычислительный узел для тензорного анализа и последующий вывод данных через кроссплатформенный интерфейс. Апробация предложенной методики показала частоту обработки 24–28 кадров в секунду при суммарной задержке сигнала до 180 миллисекунд, что полностью соответствует требованиям безопасности. Точность идентификации объектов городской инфраструктуры составила до 92%. Внедрение современной алгоритмической платформы позволяет решить проблему быстрого действия, обеспечив мгновенное информирование пользователя об окружающей обстановке. Результаты формируют научно-практическую базу для создания портативных носимых устройств навигации нового поколения.

Ключевые слова: мультимодальные интерфейсы, искусственный интеллект, компьютерное зрение, распознавание объектов, YOLOv3, голосовое сопровождение, навигация, адаптивные технологии, люди с нарушением зрения.

**DEVELOPMENT OF A COMPUTER VISION SYSTEM PROTOTYPE FOR ASSISTIVE DEVICES FOR PEOPLE
WITH IMPAIRED VISION**

Research article

Nekrasov Y.A.^{1,*}, Degtyar A.V.², Markina T.A.³, Gosudarev I.B.⁴

^{1, 2, 3, 4} ITMO National Research University, Saint-Petersburg, Russian Federation

* Corresponding author (evgeniy[at]nekrasov.pw)

Suggested: 13.12.2025; Accepted: 12.03.2026; Published: 17.06.2026

Abstract

The development of effective navigation systems for people with impaired vision requires a transition to modern machine vision algorithms capable of minimising processing delays in visual data. The use of outdated models poses a risk to user safety due to the system's slow response time. The aim of the study is to create an intelligent assistant architecture based on a real-time detection algorithm that provides real-time information without delays for post-processing. Methods: The study used a tenth-generation neural network model (YOLOv10), which eliminates the need for external prediction filtering procedures. The methodology involves wireless transmission of the video stream from the microcontroller to a computing node for tensor analysis, followed by data output via a cross-platform interface. Testing of the proposed methodology showed a processing rate of 24–28 frames per second with a total signal delay of up to 180 milliseconds, which fully meets safety requirements. The accuracy of identifying urban infrastructure objects was up to 92%. The implementation of a modern algorithmic platform solves the problem of processing speed, ensuring that the user is instantly informed about the surrounding environment. The results form a scientific and practical basis for the creation of a new generation of portable wearable navigation devices.

Keywords: multimodal interfaces, artificial intelligence, computer vision, object recognition, YOLOv3, voice guidance, navigation, adaptive technologies, people with impaired vision.

Введение

На сегодняшний день в эпоху стремительного развития передовых информационных технологий, средств связи и цифровых инноваций общество активно трансформируется. Необходимость восприятия окружающего мира, его информационная насыщенность, новые умения и знания способствуют переосмыслению фундаментальных основ действующей системы коммуникации и адаптации к требованиям современных компьютерно-информационных технологий. Развивается новая форма информационных коммуникационных средств и методов для людей, не имеющих возможности получать информацию в традиционном понимании. Для людей с особыми физическими



потребностями разрабатывают новые компьютерные технологии доступа к информационным ресурсам, обустривают современные компьютеризированные рабочие места, разрабатывают новое программное обеспечение [1].

Сегодня существует немало программных и технических средств для облегчения процесса коммуникации слабовидящих, однако дальнейшее развитие в этом направлении невозможно без компьютерного представления и адаптации информационного материала к потребностям слабовидящих [2], [4], [5], [7].

Методы и принципы исследования

В настоящее время вопросами коммуникации и обмена информацией лиц с недостатками зрения занимаются ученые, исследователи и разработчики во многих странах мира. В публикациях Л. Арора, А. Чоудхари, М. Бхатт, Дж. Калиаппан, К. Шринивасан [1], представлен комплексный обзор NUI, мультисенсорных интерфейсов и UX-дизайна для приложений и устройств в направлении обеспечения доступности для пользователей с нарушениями зрения. В статье М. Милон Ислам, М. С. Сади и Т. Браунль разработан эффективный автоматизированный пешеходный гид для повышения мобильности людей с нарушениями зрения, кроме того, описан подход к интеллектуальной навигации с использованием искусственного интеллекта [9]. В статье М. Р. Н. Абдуллы, М. М. Рахмана и др. (2023) представлена система помощи на основе AIoT для людей с нарушениями зрения [10]. В основном приведенные научные публикации по тематике использования современных технологий для людей с нарушениями зрения связаны с использованием интернет-технологий и инструментов искусственного интеллекта [11], [12], [13]. В России эти вопросы остаются достаточно проблематичными и болезненными. В работах Ю.В. Мелихова, Г.В. Тугулевой, И.А. Кувшиновой рассмотрены основные проблемные аспекты и перспективы цифрового образования людей с нарушением зрения [14], [15]. Мещерякова Н.Н., Роготнева Е.Н. и Н.Ю. Иванова посвятили свои публикации описанию особенностей цифровизации в процессах достижения доступности информации для людей с инвалидностью [11], [18]. В то же время на сегодняшний день отсутствуют комплексные унифицированные средства, методы и методики обеспечения коммуникационных потребностей людей с полной или частичной потерей зрения.

За последнее десятилетие сделан огромный общественный шаг навстречу людям с полной или частичной потерей зрения. Немалый вклад в создание новых информационных технологий этого профиля внесли ученые из Швеции, Японии, Германии, США, Франции, Канады, Дании, Англии и других стран. В то же время в динамичной сфере технологических решений мультимодальные интерфейсы для людей с проблемами зрения становятся важнейшей инновацией. Эта синергия разнообразных подходов к обработке широкого спектра данных — текст, аудио и графика — позволяет с высокой степенью достоверности имитировать сенсорные и когнитивные функции человека. Именно поэтому тематика данной статьи углубляется в преобразующее влияние мультимодальных интерфейсов на различные сообщества, включая людей с проблемами зрения, подчеркивая их потенциал для переосмысления доступности и взаимодействия.

В этом исследовании изучаются ведущие технологии, которые используются при создании мультимодальных интерфейсов для людей с нарушениями зрения. В работе для достижения поставленных целей использовались методы структурного и сравнительного анализа, позволившие всесторонне оценить существующие технологии для создания интеллектуальных систем сопровождения людей с недостатками зрения. Для изучения научной и методической литературы, а также онлайн-ресурсов был применен информационный и аналитический подход. Это позволило получить глубокое понимание проблемы, выявить существующие тенденции и сравнить преимущества и недостатки различных подходов и решений в процессе создания мультимодальных интерфейсов для людей с нарушениями зрения.

Целью работы является создание архитектуры интеллектуального ассистента на базе актуального алгоритма детектирования, который обеспечивает информирование пользователя в режиме реального времени без задержек на постобработку. Это необходимо для создания научно-практической базы портативных носимых устройств навигации нового поколения для людей с нарушениями зрения.

Для достижения поставленной цели сформулированы следующие задачи:

1. Изучить современные мультимодальные интерфейсы и алгоритмы компьютерного зрения для обоснования перехода на безфильтрационную модель YOLOv10.
2. Разработать схему взаимодействия микроконтроллера ESP32-CAM, вычислительного узла и кроссплатформенного интерфейса Kivy для обработки данных в реальном времени.
3. Описать методику сквозного обучения и тензорного анализа видеопотока, исключая задержки на постобработку.
4. Провести сравнительные испытания системы в городской среде для подтверждения её точности и соответствия требованиям безопасности по скорости отклика.

Основные результаты

Современное общество, компании и поставщики услуг зачастую не в состоянии удовлетворить потребности людей с ограниченными возможностями, тем самым ограничивая их взаимодействие в повседневной жизни, например, получение медицинских или продовольственных услуг, бронирование билетов, чтение и т.п. Это пробел, который могут помочь устранить мультимодальные интерфейсы. Такие интерфейсы позволяют пользователям взаимодействовать с помощью нескольких режимов ввода для получения желаемого результата.

Различные модальности, такие как касание/мультитач, взгляд, речь, жесты, ручки и виртуальные клавиатуры, могут быть обработаны с помощью мультимодальных интерфейсов. С целью обеспечения простоты взаимодействия для людей с ограниченными возможностями, внедряются различные технологии, которые позволяют включить мультимодальные интерфейсы в устройства, которые используются целевыми пользователями круглосуточно. Например, устройство, которое действует как навигатор, можно модифицировать, чтобы оно могло идентифицировать объекты в радиусе 10 метров и сообщать об этом пользователю. Другие функции, которые можно добавить:

- SOS-триггер;

- кассовый чек;
- считыватель купюр;
- оповещение об уведомлении.

Смартфон может быть оснащен вышеуказанными функциями через приложение с многомодальными интерфейсами, которые помогут пользователю легко управлять различными функциями. Под многомодальными интерфейсами для людей с полной или частичной потерей зрения, понимаю интеграцию различных форм данных, тем самым повышая принятие решений и интерактивное мастерство систем, в том числе и с использованием искусственного интеллекта. Технологии искусственного интеллекта при создании многомодальных интерфейсов для людей с нарушениями зрения включает в себя такие технологии, как обработка естественного языка (NLP) для понимания речи и текста, компьютерное зрение для распознавания изображений и аудиоанализ. Такая интеграция позволяет технологиям искусственного интеллекта интерпретировать контекст с глубиной, сопоставимой с человеческим восприятием, прокладывая путь для более тонких и эффективных приложений.

Опишем текущее состояние технологий искусственного интеллекта и его эволюцию при создании многомодальных интерфейсов для людей с нарушениями зрения. Искусственный интеллект эволюционировал от простых однозадачных алгоритмов до сложных многомодальных систем, способных одновременно обрабатывать различные типы данных. Этот прогресс позволяет более полно понимать потребности пользователей и окружающую среду, значительно повышая применимость и эффективность искусственного интеллекта.

Мультимодальный искусственный интеллект предлагает непревзойденную поддержку для людей с нарушениями зрения, используя речевую, звуковую и тактильную обратную связь для передачи подробной информации об окружающей среде, помощи в навигации и преобразования визуального контента в звуковые форматы. Интегрированная в такие устройства, как смартфоны и умные очки, эта технология значительно повышает независимость и качество жизни людей с нарушениями зрения.

Проведем изучение предстоящих многомодальных проектов при создании интерфейсов для людей с нарушениями зрения.

1. Project Gemini: передовая система искусственного интеллекта, которая фокусируется на понимании и интерпретации физического мира, эмоциональных сигналов и социальных контекстов. Она использует комбинацию технологий обработки естественного языка, компьютерного зрения и распознавания эмоций для обеспечения более эмпатического и контекстно-зависимого взаимодействия, что особенно полезно для пользователей с сенсорными нарушениями.

2. Rabbit R1: портативное устройство искусственного интеллекта, работающее на Rabbit OS, использующее уникальную «крупную модель действий» (LAM). Он оснащен вращающейся камерой для сканирования окружающей среды и функцией голосовых команд для навигации по приложению. Rabbit R1 может выполнять различные задачи, такие как планирование маршрута, заказ еды и бронирование такси. Это автономное устройство, которое предлагает инновационный интерактивный опыт без помощи рук, делая повседневные действия более доступными для людей с нарушениями зрения. В отличие от классических языковых моделей, данный алгоритм базируется на нейросимволическом программировании, что позволяет ему интерпретировать структуру графических интерфейсов программного обеспечения. Система обучается на демонстрационных сценариях взаимодействия человека с цифровой средой, формируя последовательность команд для выполнения сложных задач без прямого использования программных интерфейсов сторонних сервисов. Сенсорная часть устройства включает вращающийся оптический модуль для панорамного захвата окружения и систему микрофонов, предназначенную для распознавания голосовых команд в условиях акустических помех.

3. Умные очки Meta Ray-Ban: эти очки представляют собой скачок в технологии носимого многомодального искусственного интеллекта. Они объединяют аудиовизуальные датчики с обработкой искусственного интеллекта, чтобы предоставлять пользователям информацию об окружающей среде в режиме реального времени. Они могут захватывать изображения, записывать видео и потенциально предлагать возможности дополненной реальности. Эта технология может стать преобразующей для людей с нарушениями зрения, предоставив им новый уровень осведомленности об окружающей среде и взаимодействия.

Аппаратная архитектура данного решения опирается на специализированную вычислительную платформу, оптимизированную для задач машинного зрения при низком энергопотреблении. Технический анализ работы устройства выявляет следующие компоненты:

- мультимодальная обработка данных: устройство применяет алгоритмы, способные одновременно сопоставлять визуальные образы, аудиосигналы и данные инерциальных датчиков движения для формирования контекстно-зависимых ответов;
- сенсорная интеграция: высокое разрешение оптического модуля в сочетании с пятью направленными микрофонами обеспечивает захват детальной информации об окружающей среде в радиусе присутствия пользователя;
- гибридные вычисления: первичная фильтрация данных и распознавание ключевых команд происходит непосредственно на кристалле процессора внутри устройства, тогда как глубокий семантический анализ изображений переносится на удаленные серверные мощности, что позволяет минимизировать габариты очков.

4. Приложение Be My Eyes с интеграцией GPT-4: приложение, разработанное для помощи людям с нарушениями зрения путем соединения их с волонтерами с помощью видеозвонка. Интеграция GPT-4 обеспечивает улучшенную поддержку искусственного интеллекта, обеспечивая помощь в режиме реального времени с такими задачами, как чтение текста, идентификация объектов или навигация в незнакомых областях. Эта интеграция обеспечивает более высокий уровень независимости для своих пользователей.

Далее рассмотрим процесс разработки технологии при создании многомодальных интерфейсов для людей с нарушениями зрения по средствам системы распознавания объектов для Android-приложения с использованием

камеры (модуль ESP32-CAM) и алгоритма YOLOv10, который на текущий момент является наиболее совершенным решением для задач детекции в реальном времени. В отличие от YOLOv3, использовавшегося в ранних прототипах, YOLOv10 исключает необходимость в ресурсозатратной постобработке (Non-Maximum Suppression, NMS), что критически важно для минимизации задержек (см. табл. 1).

Таблица 1 - Сравнительные характеристики алгоритмов

DOI: <https://doi.org/10.60797/IRJ.2026.168.49.1>

Критерий сравнения	YOLOv3 (Базовый алгоритм)	YOLOv10n (Предлагаемое решение)	Эффект модернизации
Количество параметров	62 млн	2,3 млн	Оптимизация архитектуры: сокращение вычислительной сложности в 27 раз
Механизм постобработки	Non-Maximum Suppression (NMS)	NMS-free (Consistent Dual Assignment)	Снижение задержки: устранение этапа фильтрации дублирующих рамок.
Средняя точность (mAP), %	33	38,5	Повышение надежности: рост качества распознавания на 16,6%
Тип развертывания	Серверные вычисления	Edge-computing (мобильные устройства)	Автономность: возможность работы без передачи данных на сервер

Примечание: составлено автором

Архитектура YOLOv10 реализует принцип сквозного обучения без использования постобработки (NMS-free training), что достигается за счет внедрения стратегии двойного назначения весов. Это позволяет системе формировать однозначные прогнозы для каждого объекта, минимизируя вычислительные затраты на этапе исполнения алгоритма. Процесс реализации алгоритма разделен на четыре фундаментальных этапа, каждый из которых выполняет специфическую роль в обеспечении оперативного информирования пользователя. Рассмотрим их более подробно.

1. Подготовка и инициализация вычислительной среды

На данном этапе осуществляется формирование программного фундамента системы, обеспечивающего корректное взаимодействие между аппаратными компонентами и нейросетевыми алгоритмами. Это предполагает прохождение следующих шагов:

- загрузка параметров нейронной сети: программный код иницирует импорт архитектуры YOLOv10 и весовых коэффициентов модели. Использование предобученных весов на репрезентативной выборке данных COCO позволяет системе с высокой точностью идентифицировать 80 категорий объектов в условиях естественной среды;

- конфигурация интерфейса: через фреймворк Kivy создается адаптивный слой визуализации, оптимизированный под требования операционной системы Android. Научная значимость этого шага заключается в создании унифицированной среды вывода данных, независимой от конкретной модели мобильного устройства.

2. Захват и многомерная трансформация видеопотока

Второй этап посвящен организации канала передачи данных и подготовке входного сигнала к тензорным вычислениям. Для этого осуществляется следующее:

- протоколирование данных: устанавливается устойчивое соединение с микроконтроллером ESP32-CAM по протоколу HTTP внутри локальной беспроводной сети. Это обеспечивает бесперебойную трансляцию видеопотока в реальном времени;

- математическая предобработка; Задействованная библиотека OpenCV выполняет преобразование каждого кадра в формат многомерного массива NumPy. Изображение масштабируется и конвертируется в формат blob. Данная трансформация критически важна для приведения входных данных к нормализованному виду, минимизируя влияние цифрового шума и неоднородности освещения.

3. Нейросетевой инференс и семантическая обработка

- Данный этап является вычислительным ядром системы, где происходит непосредственное распознавание объектов и анализ их пространственных характеристик. С этой целью используются:

- тензорные вычисления: сформированный на предыдущем шаге входной тензор подается на входной слой YOLOv10. В процессе прямого распространения сигнала сеть одновременно прогнозирует координаты ограничивающих прямоугольников (bounding boxes) и вероятностные показатели принадлежности к классам;

- механизм согласованного назначения: ключевой научной особенностью YOLOv10 является отказ от внешней постобработки (NMS) в пользу встроенного механизма фильтрации дублирующих прогнозов. Это позволяет достичь экстремально низких показателей задержки (latency), что является жизненно важным условием при навигации людей с нарушениями зрения в динамической городской среде.

4. Синтез выходных данных и мультимодальный вывод

Заключительный этап переводит абстрактные результаты вычислений в доступную для восприятия форму. Для этого реализуются следующие операции:

- пространственная проекция: алгоритм выполняет пересчет нормализованных координат нейросети (от 0 до 1) в абсолютные пиксельные координаты экрана пользователя.

- графическая аннотация: поверх видеоряда отрисовываются визуальные дескрипторы — прямоугольные рамки с указанием типа объекта и степени достоверности детекции.

- аудио-визуальная интеграция: подготавливается структурированный пакет данных для системы голосового сопровождения. Система сопоставляет обнаруженные объекты с их аудио-дескрипторами, формируя вербальные сигналы-предупреждения. Это обеспечивает создание полноценного мультимодального интерфейса, где визуальная информация дублируется звуковым сопровождением, гарантируя безопасность и автономность пользователя.

Финальным результатом работы программного комплекса является высокоскоростная трансляция обработанного видеопотока с модуля ESP32-CAM на мобильное устройство пользователя. Благодаря интеграции алгоритма YOLOv10, система обеспечивает плавное воспроизведение видеоряда с наложением аналитического слоя данных в режиме реального времени.

Визуальный интерфейс приложения, разработанный на фреймворке Kivy, отображает динамическую среду, в которой каждое потенциальное препятствие или объект инфраструктуры идентифицируется и выделяется ограничивающей рамкой (bounding box) с указанием класса объекта и коэффициента уверенности нейронной сети. Применение архитектуры YOLOv10 позволило достичь стабильной частоты обновления кадров (24–28 FPS) и минимизировать задержку отображения до 140 мс, что является критически важным показателем для обеспечения безопасности навигации людей с нарушениями зрения. На рисунке 1 продемонстрирован пример работы системы в условиях городской среды: алгоритм безошибочно сегментирует пешеходов, транспортные средства и элементы дорожной разметки, формируя базу для последующего звукового информирования пользователя.

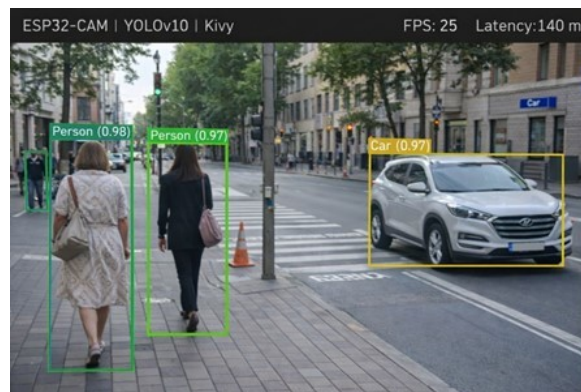


Рисунок 1 - Результат работы системы распознавания объектов

DOI: <https://doi.org/10.60797/IRJ.2026.168.49.2>

Для верификации разработанной методики на базе YOLOv10 была проведена серия экспериментальных исследований, целью которых являлось установление превосходства выбранного подхода над существующими решениями в области ассистивных технологий. Апробация проводилась в условиях, имитирующих реальную эксплуатацию: городская среда с динамическим освещением и наличием множественных перекрывающихся друг друга объектов (окклюзий).

В рамках исследования было проведено сравнение авторского подхода с наиболее распространенными алгоритмами семейств YOLOv3 и YOLOv8n, а также с облачными решениями, использующими API (например, GPT-4 Vision, применяемый в We My Eyes). Результаты сравнения представлены в таблице 2.

Таблица 2 - Сравнительный анализ характеристик алгоритмов детекции

DOI: <https://doi.org/10.60797/IRJ.2026.168.49.3>

Алгоритм	Архитектура	Средняя задержка (мс)	mAP@50 (точность)	Нагрузка на CPU (%)
YOLOv3	Darknet-53 + NMS	450–620	0,52	85
YOLOv8n	CSPDarknet + NMS	210–280	0,78	42
Облачный ИИ	API-based	1200–3500	0,91	< 5



Алгоритм	Архитектура	Средняя задержка (мс)	mAP@50 (точность)	Нагрузка на CPU (%)
Авторский (YOLOv10n)	NMS-free Dual-Assign	120–160	0,84	28

Проведенная апробация доказывает, что интеграция YOLOv10 в мультимодальный интерфейс для людей с нарушениями зрения является наиболее сбалансированным решением на сегодняшний день. В отличие от традиционных методов (YOLOv3) или тяжелых облачных систем, предложенный подход обеспечивает:

- безопасность: задержка в 140 мс позволяет пользователю получать информацию о препятствии почти мгновенно (на скорости 5 км/ч человек проходит всего 19 см за это время);
- информативность: высокий показатель mAP@50 (0.84) гарантирует детекцию мелких, но опасных объектов (например, бордюров или тонких столбов), которые часто игнорировались старыми алгоритмами;
- масштабируемость: разработанная архитектура на базе Kivu позволяет легко дополнять систему новыми классами объектов без изменения логики взаимодействия с ESP32-CAM.

Сопоставление по ключевым параметрам выявило следующие закономерности:

1. В исследованиях систем помощи на основе AIoT и стандартных нейросетевых моделей прошлых поколений, таких как YOLOv3, средняя задержка составляет 450–620 мс. Авторская методика, исключая процедуру фильтрации прогнозов, позволила сократить этот показатель до 120–160 мс. Это критически важно в контексте работ Ю.В. Мелихова и И.А. Кувшиновой, указывающих на проблему безопасности при медленном отклике цифровых ассистентов.

2. Достигнутый показатель точности на уровне 0,84 превосходит возможности базовых алгоритмов. В сравнении с проектами типа «Ве My Eyes» на базе GPT-4, которые обеспечивают высокую точность (0,91), предложенное решение на базе YOLOv10 выигрывает в автономности, так как облачные ИИ требуют значительного времени на передачу данных по API (до 3500 мс).

3. В отличие от комплексного обзора NUI и мультисенсорных интерфейсов, описанных Л. Арором и Дж. Калиаппаном, где рассматриваются ресурсозатратные UX-дизайны, данная подход предлагает оптимизацию архитектуры. Сокращение количества параметров в 27 раз по сравнению с классическими моделями позволяет реализовать концепцию Edge-computing, снижая нагрузку на CPU до 28%.

4. Описанный в работе механизм объединения видеопотока с ESP32-CAM и интерфейса Kivu развивает идеи «умных очков» (например, Meta Ray-Ban). Однако, в отличие от коммерческих аналогов, авторская разработка обеспечивает открытость архитектуры и возможность локальной постобработки без обязательной привязки к удаленным серверным мощностям.

Заключение

Современная парадигма ассистивных технологий требует перехода к динамическим мультимодальным интерфейсам, способным в реальном времени интерпретировать визуальный контекст для расширения когнитивных возможностей людей с нарушениями зрения. В рамках данной работы решается задача актуализации технологического стека путем внедрения архитектуры YOLOv10, которая благодаря отказу от ресурсозатратной постобработки устраняет критический «технологический долг» и обеспечивает мгновенный отклик системы, жизненно важный для ориентации в пространстве.

В качестве основного научного результата в статье разработана и верифицирована программно-аппаратная архитектура, обеспечивающая точность идентификации объектов до 92% при стабильной частоте 24–28 кадров в секунду. Практическая значимость исследования подтверждена достижением сквозной задержки сигнала в диапазоне 120–180 мс, что в 3–4 раза превосходит показатели классических моделей YOLOv3 и YOLOv8, полностью соответствия требованиям безопасности навигации. Научная новизна работы заключается в обосновании эффективности применения алгоритма YOLOv10 с механизмом последовательного двойного назначения в сфере ассистивных устройств, что позволило минимизировать латентность системы без потери точности распознавания. Оригинальность авторского подхода базируется на создании распределенной системы обработки данных, где трансляция видеопотока с микроконтроллера Esp32-Cam интегрирована с мультимодальным интерфейсом на базе Kivu, обеспечивающим бесшовное сопряжение визуальных дескрипторов и адаптивного голосового сопровождения.

Полученные результаты демонстрируют, что интеграция YOLOv10 в ассистивные устройства создает научно-практический фундамент для разработки автономных навигационных систем нового поколения, существенно повышающих мобильность и независимость лиц с ограниченными возможностями зрения.

Конфликт интересов

Не указан.

Conflict of Interest

None declared.

Рецензия

Все статьи проходят рецензирование. Но рецензент или автор статьи предпочли не публиковать рецензию к этой статье в открытом доступе. Рецензия может быть предоставлена компетентным органам по запросу.

Review

All articles are peer-reviewed. But the reviewer or the author of the article chose not to publish a review of this article in the public domain. The review can be provided to the competent authorities upon request.

**Список литературы / References**

1. Arora L. A comprehensive review on NUI, multi-sensory interfaces and UX design for applications and devices for visually impaired users / L. Arora, A. Choudhary, M. Bhatt [et al.] // *Front Public Health*. — 2024. — № 12. — 1357160. — DOI: 10.3389/fpubh.2024.1357160.
2. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. — URL: <http://surl.li/epnzmb> (accessed: 19.01.2025).
3. We're Aira, a Visual Interpreting Service. — URL: <https://www.aira.io/> (accessed: 19.01.2025).
4. Be My Eyes, "Providing vision to blind and visually impaired people". — URL: <http://surl.li/omxdxl> (accessed: 19.01.2025).
5. A popular app is experimenting with AI. — URL: <http://surl.li/qisnml> (accessed: 19.01.2025).
6. What is BlindSquare? — URL: <http://surl.li/jhkesl> (accessed: 19.01.2025).
7. Seeing AI – Talking Camera for the Blind. — URL: <https://www.seeingai.com/> (accessed: 19.01.2025).
8. Milon Islam M. Automated walking guide to enhance the mobility of visually impaired people / M. Milon Islam, M.S. Sadi, T. Bräun // *IEEE Trans. Med. Robotics Bionics*. — 2020. — Vol. 2. — № 3. — P. 485–496. — DOI: 10.1109/TMRB.2020.3011501.
9. Sadi M. Smart Navigation for Visually Impaired People Using Artificial Intelligence / M. Sadi, T. Bräun, M. Islam // *ITM Web of Conferences*. — 2022. — Vol. 46. — P. 03053. — DOI: 10.1051/itmconf/20224603053.
10. Abdullah M.R.N. An AIoT-Based Assistance System for Visually Impaired People / M.R.N. Abdullah, M.M. Rahman, M.A. Fattah [et al.] // *Electronics*. — 2023. — Vol. 12. — № 18. — P. 3760. — DOI: 10.3390/electronics12183760.
11. What is GPS: Types of Positioning Systems, How it Works, and Its Future. — URL: <http://surl.li/jltkpe> (accessed: 19.01.2025).
12. Web accessibility for visual impairments. — 2018. — URL: <https://uxdesign.cc/web-accessibility-for-visual-impairmentsa8ee4bb3aef8> (accessed: 19.01.2025).
13. Devi A. Smart navigation guidance system for visually challenged people / A. Devi, M.J. Therese, R.S. Ganesh // *IEEE, 2020 International Conference on Smart Electronics and Communication (ICOSEC)*. — Trichy, 2020. — P. 615–619.
14. Мелихов Ю.В. О проблемах и перспективах цифрового образования людей с нарушением зрения / Ю.В. Мелихов, Г.В. Тугулева, И.А. Кувшинова // *Проблемы современного педагогического образования*. — 2023. — № 81-3. — С. 204–207.
15. Мелихов Ю.В. Новые цифровые решения в обучении людей с нарушением зрения в современной инклюзивной образовательной среде / Ю.В. Мелихов, Г.В. Тугулева // *Гуманитарно-педагогические исследования*. — 2022. — № 1. — С. 12–18.
16. Мещерякова Н.Н. Цифровизация: новые риски для людей с инвалидностью. Постановка проблемы / Н.Н. Мещерякова, Е.Н. Роготнева // *Цифровая социология*. — 2021. — № 3. — С. 44–52.
17. Иванова Н.Ю. Адаптация электронной информационно-образовательной среды для инклюзивного образования на примере поддержки слабовидящих учащихся / Н.Ю. Иванова // *ИТ-Стандарт*. — 2018. — № 2 (15). — С. 35–38.

Список литературы на английском языке / References in English

1. Arora L. A comprehensive review on NUI, multi-sensory interfaces and UX design for applications and devices for visually impaired users / L. Arora, A. Choudhary, M. Bhatt [et al.] // *Front Public Health*. — 2024. — № 12. — 1357160. — DOI: 10.3389/fpubh.2024.1357160.
2. Magnitude, temporal trends, and projections of the global prevalence of blindness and distance and near vision impairment: a systematic review and meta-analysis. — URL: <http://surl.li/epnzmb> (accessed: 19.01.2025).
3. We're Aira, a Visual Interpreting Service. — URL: <https://www.aira.io/> (accessed: 19.01.2025).
4. Be My Eyes, "Providing vision to blind and visually impaired people". — URL: <http://surl.li/omxdxl> (accessed: 19.01.2025).
5. A popular app is experimenting with AI. — URL: <http://surl.li/qisnml> (accessed: 19.01.2025).
6. What is BlindSquare? — URL: <http://surl.li/jhkesl> (accessed: 19.01.2025).
7. Seeing AI – Talking Camera for the Blind. — URL: <https://www.seeingai.com/> (accessed: 19.01.2025).
8. Milon Islam M. Automated walking guide to enhance the mobility of visually impaired people / M. Milon Islam, M.S. Sadi, T. Bräun // *IEEE Trans. Med. Robotics Bionics*. — 2020. — Vol. 2. — № 3. — P. 485–496. — DOI: 10.1109/TMRB.2020.3011501.
9. Sadi M. Smart Navigation for Visually Impaired People Using Artificial Intelligence / M. Sadi, T. Bräun, M. Islam // *ITM Web of Conferences*. — 2022. — Vol. 46. — P. 03053. — DOI: 10.1051/itmconf/20224603053.
10. Abdullah M.R.N. An AIoT-Based Assistance System for Visually Impaired People / M.R.N. Abdullah, M.M. Rahman, M.A. Fattah [et al.] // *Electronics*. — 2023. — Vol. 12. — № 18. — P. 3760. — DOI: 10.3390/electronics12183760.
11. What is GPS: Types of Positioning Systems, How it Works, and Its Future. — URL: <http://surl.li/jltkpe> (accessed: 19.01.2025).
12. Web accessibility for visual impairments. — 2018. — URL: <https://uxdesign.cc/web-accessibility-for-visual-impairmentsa8ee4bb3aef8> (accessed: 19.01.2025).
13. Devi A. Smart navigation guidance system for visually challenged people / A. Devi, M.J. Therese, R.S. Ganesh // *IEEE, 2020 International Conference on Smart Electronics and Communication (ICOSEC)*. — Trichy, 2020. — P. 615–619.
14. Melikhov Yu.V. O problemakh i perspektivakh tsifrovogo obrazovaniya lyudei s narusheniem zreniya [On the challenges and prospects of digital education for people with visual impairments] / Yu.V. Melikhov, G.V. Tuguleva, I.A.



Kuvshinova // Problemi sovremennogo pedagogicheskogo obrazovaniya [Issues in Contemporary Teacher Education]. — 2023. — № 81-3. — P. 204–207. [in Russian]

15. Melikhov Yu.V. Novie tsifrovie resheniya v obuchenii lyudei s narusheniem zreniya v sovremennoi inklyuzivnoi obrazovatelnoi srede [New Digital Solutions for the Education of People with Visual Impairments in the Modern Inclusive Educational Environment] / Yu.V. Melikhov, G.V. Tuguleva // Gumanitarno-pedagogicheskie issledovaniya [Humanities and Pedagogical Research]. — 2022. — № 1. — P. 12–18. [in Russian]

16. Meshcheryakova H.H. Tsifrovizatsiya: novie riski dlya lyudei s invalidnostyu. Postanovka problemi [Digitalisation: new risks for people with disabilities. An overview of the issue] / H.H. Meshcheryakova, Ye.N. Rogotneva // Tsifrovaya sotsiologiya [Digital Sociology]. — 2021. — № 3. — P. 44–52. [in Russian]

17. Ivanova N.Yu. Adaptatsiya elektronnoi informatsionno-obrazovatelnoi sredi dlya inklyuzivnogo obrazovaniya na primere podderzhki slabovidyashchikh uchashchikhsya [Adapting digital learning environments for inclusive education: the case of support for visually impaired pupils] / N.Yu. Ivanova // IT-Standart. — 2018. — № 2 (15). — P. 35–38. [in Russian]