

Федеральное государственное автономное образовательное учреждение
высшего образования «Национальный исследовательский технологический
университет «МИСИС»

На правах рукописи

Гужва Никита Сергеевич

**Идентификация светофоров на основе 3D измерений нейронных сетей в
мультикамерных системах помощи водителя трамвая**

Специальность 2.3.1 — «Системный анализ, управление и обработка
информации, статистика»

Диссертация на соискание учёной степени кандидата технических наук

Научный руководитель
д.т.н., доцент Садеков Р. Н.

Москва — 2025

Оглавление

Введение	5
Глава 1. Анализ существующих решений и постановка задачи	12
1.1 Анализ существующих подходов идентификации светофоров . . .	12
1.1.1 Автоматическое распознавание светофоров в задачах беспилотного транспорта	12
1.1.2 Уровни автономности транспортных средств в контексте подсистемы идентификации светофоров	16
1.1.3 Подходы к решению задачи	21
1.2 Постановка задачи и этапы ее решения	35
Глава 2. Методика сопоставления детекций в двухкамерной разнофокусной системе	41
2.1 Обзор методов трекинга и ассоциации детекций	41
2.1.1 Межкадровое сопоставление детекций	41
2.1.2 Объединение результатов детекций между камерами	44
2.1.3 Метрики схожести для сопоставления треклетов	47
2.1.4 Обработка асинхронных измерений	48
2.2 Трекинг	50
2.2.1 Основные этапы трекинга	50
2.2.2 Представление состояния и модель движения	51
2.2.3 Способ ассоциации данных	53
2.3 Объединение результатов детекций между камерами	56
2.3.1 Анализ существующих подходов	56
2.3.2 Предлагаемый подход	59

Глава 3. Методика фильтрации пространственных координат

светофоров	72
3.1 Обзор методов фильтрации и оценки координат	72
3.1.1 Методы оптимальной фильтрации	72
3.1.2 Тесносвязанные схемы комплексирования	78
3.2 Оптимальная фильтрация координат местоположения светофоров	80
3.2.1 Описание алгоритмов фильтрации	80
3.2.2 Оценка точности фильтрации	84
3.3 Сопоставление результатов с картой	89
3.3.1 Семантическая цифровая карта	89
3.3.2 Алгоритм сопоставления с картой	91
3.4 Классификация сигналов светофоров	93
3.4.1 Классификационная нейронная сеть	93
3.4.2 Обеспечение временной стабильности сигналов	95

Глава 4. Методика оценки точности алгоритма идентификации

светофоров	100
4.1 Анализ существующих подходов оценки точности работы подсистемы распознавания светофоров	100
4.1.1 Наборы данных для оценки точности	100
4.1.2 Метрики оценки точности	102
4.2 Действия и сценарии работы системы	105
4.3 Входная информация и критерии формирования событий	108
4.3.1 Входная информация и эталонное поведение	109
4.3.2 Критерии формирования событий	109
4.4 Результаты оценки точности системы	111

Заключение	118
-------------------	------------

Список сокращений и условных обозначений	121
---	------------

Словарь терминов	124
Список рисунков	150
Список таблиц	152
Приложение А. Акт о внедрении	153

Введение

Актуальность работы. Для повышения безопасности движения и удобства управления транспортным средством используются системы активной помощи водителю. В последнее время внедрение таких систем производится и на общественном транспорте, к которым относятся трамваи. Система активной помощи трамвая должна осуществлять полную остановку (у стоп-линии) в случае, если трамвай едет через перекресток на запрещающий сигнал светофора. Принятие решения как при подъезде трамвая к перекрестку, так и трогании с него требует разработки высокоточных и надежных алгоритмов идентификации светофоров. Под идентификацией светофоров понимается процесс, включающий непосредственно детектирование светофоров на изображениях с камер; связывание результатов детектирования с разных камер для получения “единых” детекций сцены; определение сигнала светофора; локализацию детекций светофоров в трехмерном пространстве (уточнение положения) и “привязку” к соответствующим светофорам на карте. Все эти шаги являются необходимыми для последующего принятия решения о действиях транспортного средства на перекрестке.

В условиях города реализация данной задачи не является тривиальной в силу многочисленных проблем: большое разнообразие сигналов и типов светофоров, “кучное” размещение светофоров на перекрестках, наличие многочисленных “ложных” источников света, расположение светофоров в разных частях дорожной сцены и ошибки навигационных систем. Данные обстоятельства требуют разработки сложных алгоритмов идентификации светофоров.

Существует инженерный подход к решению задачи идентификации сигналов светофора - технология V2X (Vehicle-to-Everything). Данная технология представляет собой систему коммуникации транспортных средств с окружающей инфраструктурой, позволяющую напрямую получать информацию о состоянии сигналов светофора через беспроводные каналы связи. Однако решение

данной задачи на основе технологии V2X не является массовым. На практике города не оборудованы соответствующими системами, поэтому текущие решения используют исключительно бортовые сенсоры. Системы помощи водителя, как правило, оснащены одной или двумя камерами. Использование высокоточных лидарных датчиков не представляется возможным прежде всего по экономическим причинам. В городской среде наиболее рациональной является схема компоновки с использованием камер ближнего и дальнего диапазона, что обеспечивает принципиальную возможность распознавания основных и дублирующих светофоров. Данное обстоятельство особенно полезно при остановке на перекрестке, когда основной управляющий светофор не виден и принятие решения должно производиться по сигналам дублирующего светофора.

Для нахождения светофоров на изображении применяются нейронные сети. При этом традиционно для целей распознавания сигналов используются 2D-детекции. Современным трендом развития нейронных сетей является возможность 3D-детектирования положения объектов сцены. Это особенно полезно в задачах анализа дорожной сцены, так как не требует использования различных инженерных алгоритмов восстановления глубины и позволяет “видеть” объект в условиях его частичного заслонения.

Таким образом, алгоритмы идентификации светофоров с использованием нескольких разнофокусных камер на основе 3D-детекций, полученных нейронной сетью, являются актуальной темой исследования.

Объектом исследования являются мультикамерные системы помощи водителю трамвая, процессы идентификации светофоров в условиях: асинхронной работы камер, имеющих разные углы обзора; динамических помех от городской инфраструктуры и ошибок навигационных систем.

Предметом исследования являются алгоритмы комплексирования асинхронных измерений, методы оптимальной фильтрации и идентификации объектов дорожной инфраструктуры.

Целью данного исследования является решение задачи идентификации светофоров бортовой системой помощи водителя трамвая на основе 3D-детекций,

полученных нейронной сетью, с учётом особенностей их объединения с разнофокусных камер и процедур оценивания координат их местоположения, влияющих на точность решения задачи.

Для достижения поставленной цели необходимо решить следующие **задачи**:

1. Провести анализ существующих моделей и алгоритмов, направленных на решение задачи идентификации светофоров в различных условиях обстановки и разном наборе сенсоров. Обосновать наиболее рациональную архитектуру системы, исходные данные для формирования алгоритмов и преимущества использования нейронных сетей с выходными данными в виде 3D-детекций;
2. Сформировать общую постановку задачи и провести её декомпозицию на этапы выполнения. По каждому предложенному этапу: межкадровое сопоставление, межкамерное сопоставление, фильтрация измерений, сопоставление с картой, предложить возможные алгоритмы реализации;
3. Разработать алгоритм объединения результатов детекции светофоров на изображениях каждой камеры в отдельности (межкадровый трекинг) и вместе (межкамерный трекинг) на основе “взвешенных” комбинированных метрик схожести;
4. Разработать алгоритм оценивания (фильтрации) положения светофоров в пространстве на основе модифицированного фильтра Калмана (Unscented Kalman Filter) с учётом детекций на изображении;
5. Разработать алгоритм сопоставления полученных координат детекций с координатами светофоров из карты;
6. Предложить методику оценки точности алгоритма идентификации, учитывающую действия водителя и не требующую проведения процедур разметки данных.

Идея работы заключается в одновременном учёте процедуры межкамерного сопоставления измерений на основе нескольких метрик их соответствия и “тесносвязанных” уравнений измерений, учитывающих положение светофоров

на изображении, что в совокупности обеспечивает минимизацию ошибок первого и второго рода работы алгоритма идентификации светофоров и проезда перекрёстков.

Основные научные положения, выносимые на защиту:

1. Методика сопоставления детекций в двухкамерной разнофокусной системе, базирующаяся на новой двухфакторной метрике сопоставления измерений каждой камеры на основе вектора признаков (эмбеддингов) и пространственного положения детекций, позволяющая связывать измерения с разных камер, что обеспечивает большее количество уравнений измерений и устойчивость работы в условиях асинхронных измерений и бликов сигналов. Соответствует П4 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта»;
2. Методика фильтрации и идентификации светофоров на карте, организованная на основе «тесносвязанной» схемы комплексирования, учитывающей прямые измерения на изображении, что позволяет повысить точность определения координат местоположения светофоров и вероятность нахождения соответствий между оценками координат светофоров и их местоположением на карте. Соответствует П5 «Разработка специального математического и алгоритмического обеспечения систем анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта»;
3. Методика оценки точности работы алгоритма идентификации светофоров, учитывающая новый перечень режимов, состояний и внешних событий системы, обеспечивающая оценку её точности на основе действий водителя и не требующая создания особых датасетов. Соответствует П3 «Разработка критериев и моделей описания и оценки эффективности решения задач системного анализа,

оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта».

Новизна научных результатов исследований заключается в следующем:

1. Методика сопоставления детекций в двухкамерной разнофокусной системе, отличающаяся от известных двухфакторной процедурой сопоставления, учитывающей похожесть вектора признаков детекций и “похожесть” их пространственного положения, в условиях асинхронных измерений, позволяет сформировать избыточные измерения для оценки местоположения светофоров на сцене;
2. Методика фильтрации и идентификации светофоров, отличающаяся от известных использованием инструментов нелинейной фильтрации на основе UKF-фильтра, использующего расширенный вектор состояния и учитывающего прямые измерения с изображений («тесно связанные» уравнения), разную точность детекций с каждой камеры, а также семантические данные цифровых карт для реализации процедур идентификации светофоров;
3. Методика оценки точности работы алгоритма идентификации светофоров отличается от известных отсутствием процедур сравнения результатов с эталонными данными в виде разметки на изображениях, а базируется на новой идее сопоставления результатов работы системы с эталонными действиями водителя трамвая в заданных сценариях, что позволяет более точно и просто оценить точность разработанных алгоритмов.

Теоретическая значимость диссертации заключается в разработке оригинальных методик и алгоритмов идентификации светофоров, учитывающих факторы асинхронности измерений, разнофокусность датчиков, ошибки пространственного положения и вектора признаков измерений, что позволяет повысить качество работы подсистемы проезда перекрестков и, как следствие, безопасность движения общественного транспорта.

Практическая значимость результатов диссертации заключается в том, что:

1. Разработан программный модуль, реализующий методику сопоставления детекций светофоров с разных камер, обеспечивающий повышение точности сопоставления по сравнению с традиционным методом, использующим только сравнение векторов признаков детекций;
2. Разработан программный модуль, реализующий методику фильтрации пространственных координат светофоров и сопоставление светофоров с картой, обеспечивающий повышение точности на 11% и полноты на 6% при использовании одной камеры и точности на 11% и полноты на 12% при использовании двух камер по сравнению с традиционным подходом;
3. Разработана программа, реализующая оценку точности системы, что на практике позволяет оперативно оценить вероятности ошибок первого и второго рода системы, не прибегая к трудоёмкому процессу разметки данных.

Описанные результаты используются в программном обеспечении современных систем помощи водителю трамвая, установленных в количестве нескольких сотен единиц и эксплуатирующихся СПб ГУП “Горэлектротранс” в г. Санкт-Петербург.

Обоснованность и достоверность результатов исследований определяется корректным использованием теории оптимальной фильтрации, интеллектуальных методов анализа данных и методов сопоставления, значительным объёмом вычислительных экспериментов и данных, собранных на протяжении нескольких лет эксплуатации систем помощи водителя трамвая в масштабах мегаполиса (г. Санкт-Петербург). О надёжности результатов свидетельствует их повторяемость в процессе эксплуатации в различных условиях обстановки.

Методы исследования включают методы компьютерного зрения и глубокого обучения, использование сверточных нейросетевых моделей (CNN) для 3D-детекции светофоров, генерации векторов признаков для семантического сопоставления детекций между камерами и трекинга, алгоритмы оптимальной

фильтрации с применением Unscented Kalman Filter (UKF) для оценки 3D-положений светофоров в условиях асинхронных измерений с разнофокусных камер, методы семантической корреляции сцены с цифровыми картами, экспериментальные и оценочные методы с использованием статистического анализа точности детекций (метрики Precision, Recall, F1-score) и устойчивости системы.

Апробация работы. Основные положения диссертации докладывались на международных конференциях “Санкт-Петербургской международной конференции по интегрированным навигационным системам” (с 2022 по 2024 год, ежегодно), на ежегодных страт-сессиях компании ООО “Когнитив Роботикс”, мероприятиях СПб ГУП “Горэлектротранс”, ежегодных семинарах и аттестациях аспирантов кафедры “Инженерной кибернетики”.

Реализация и внедрение работы. В компании ООО “Когнитив Роботикс” при производстве систем помощи водителю трамвая. В СПб ГУП “Горэлектротранс” для оценки качества работы эксплуатирующихся систем помощи водителю трамвая в части подсистемы распознавания светофоров. На ежегодном конкурсе водительского мастерства в 2024 году в г. Санкт-Петербург (см. приложение **A**).

Объем и структура работы. Диссертация состоит из введения, 4 глав, заключения и 1 приложения. Полный объем диссертации составляет 154 страницы, включая 31 рисунок и 9 таблиц. Список литературы содержит 111 наименований.

Глава 1. Анализ существующих решений и постановка задачи

1.1 Анализ существующих подходов идентификации светофоров

1.1.1 Автоматическое распознавание светофоров в задачах беспилотного транспорта

Светофоры являются неотъемлемой частью транспортной инфраструктуры, предназначенной для регулирования дорожного движения и обеспечения безопасности участников. Согласно статистическим данным Всемирной организации здравоохранения, ежегодно в дорожно-транспортных происшествиях погибает около 1,35 миллиона человек по всему миру, при этом значительная доля аварий происходит на перекрестках, оборудованных светофорами [1]. Эффективное функционирование светофорной сети напрямую влияет на снижение аварийности и повышение пропускной способности.

Роль светофоров особенно важна в условиях интенсивного транспортного потока, характерного для современных мегаполисов, где синхронизация движения транспортных средств и пешеходов достигается именно благодаря сигналам светофоров. В крупных городах, таких как Москва, Санкт-Петербург, Лондон или Токио, количество светофорных объектов может исчисляться десятками тысяч, что создает сложную многоуровневую систему управления дорожным движением. В системах автономного вождения корректное распознавание светофоров — одна из ключевых задач компьютерного зрения, напрямую влияющая на безопасность навигации и принятие решений [2; 3].

Исторически развитие светофорных систем прошло путь от простейших механических устройств начала XX века до современных интеллектуальных систем управления движением. Первый электрический светофор был установлен в 1912 году в Солт-Лейк-Сити, а к настоящему времени светофорные системы

эволюционировали в сложные адаптивные комплексы, способные изменять алгоритмы работы в зависимости от интенсивности движения и времени суток [4]. Современные светофоры оснащаются светодиодными источниками света, системами дистанционного мониторинга и могут интегрироваться в цифровые городские системы управления транспортом.

В контексте автономных транспортных средств задача детектирования светофоров осложняется рядом технических и методологических факторов. Техническое разнообразие светофорного оборудования представляет собой серьезный вызов для систем автоматического распознавания. В различных странах и регионах используются значительно отличающиеся типы и конструкции светофоров с существенными различиями в геометрии и компоновке сигнальных элементов, включая как вертикальное, так и горизонтальное расположение. Отдельную сложность создают нестандартные светофоры, предназначенные для трамвайного движения и пешеходных переходов. Кроме того, в светофорах применяются различные технологии освещения — от традиционных галогенных ламп до современных светодиодных систем и матричных дисплеев.

Международные стандарты светофорного оборудования существенно различаются между регионами. Венская конвенция о дорожных знаках и сигналах 1968 года устанавливает общие принципы, однако на практике каждая страна имеет свои особенности реализации. Например, в Великобритании и некоторых странах Содружества используется четырехфазная система сигналов (красный, красный+желтый, зеленый, желтый), в то время как в большинстве европейских стран применяется трехфазная система. В Японии светофоры часто имеют горизонтальное расположение и синий оттенок зеленого сигнала, что создает дополнительные вызовы для систем компьютерного зрения [5; 6].

Светофоры можно классифицировать на несколько основных типов в зависимости от их назначения и конструктивных особенностей. Стандартные автомобильные светофоры представляют собой трехсекционные устройства с красным, желтым и зеленым сигналами, регулирующие основные потоки

транспорта. Стрелочные светофоры для поворотов включают дополнительные секции с контурными стрелками, управляющие поворотными движениями. Трамвайные светофоры являются специализированными устройствами с уникальной символикой (буквы Т, белые полосы), предназначенными для управления движением рельсового транспорта.

Светофоры для пешеходов и велосипедистов оснащаются пиктограммами человека или велосипеда и часто дополняются звуковыми сигналами. Реверсивные светофоры предназначены для управления движением по полосам с изменяемым направлением. Железнодорожные светофоры представляют собой специализированные сигнальные устройства для регулирования движения поездов на переездах. Каждый тип имеет отличительные характеристики — форму корпуса, цветовую гамму, расположение сигнальных элементов, тип используемых пиктограмм и специфические геометрические особенности.

Современные светофоры также могут оснащаться цифровыми табло обратного отсчета, дополнительными информационными панелями и системами адаптивного регулирования яркости в зависимости от условий освещенности. Автоматическая система должна быть способна не только к бинарной классификации («горит/не горит»), но и к полноценной интерпретации сигнала в зависимости от типа светофора и дорожной ситуации [7; 8].

Условия окружающей среды также существенно влияют на качество распознавания. Особые трудности возникают при работе в условиях недостаточной освещенности, характерной для ночного времени, сумерек или при движении в тоннелях [9—11]. Значительное влияние оказывают неблагоприятные погодные условия, такие как туман, дождь, снег или град. Серьезной проблемой является засветка от прямых солнечных лучей и бликов на мокром асфальте. Сезонные изменения, включая появление листвы на деревьях или накопление снежных отложений на светофорах, также могут существенно затруднять процесс распознавания.

Динамические помехи городской среды создают дополнительный уровень сложности для систем компьютерного зрения. Светофоры часто оказываются

частично или полностью перекрыты объектами городской инфраструктуры, такими как деревья, здания или другие транспортные средства. Существенные помехи создают иные источники света, включая рекламные конструкции, фары автомобилей и освещенные витрины магазинов. При движении транспортного средства постоянно меняется угол обзора светофоров, а вибрации и тряска негативно влияют на качество получаемого изображения [12; 13].

Сложность задачи усиливается тем, что на одном кадре может присутствовать несколько светофоров, относящихся к разным полосам движения. Это требует от алгоритма не только способности распознать объект и определить активный сигнал, но и понимания контекста, то есть выделения именно тех сигналов, которые релевантны текущей полосе движения транспортного средства. Такое когнитивное восприятие сцены традиционно считается задачей более высокого уровня и требует интеграции визуальных данных с картографической и навигационной информацией [14; 15].

Особую сложность представляют перекрестки с множественными светофорными группами, где может одновременно функционировать до 10-15 различных светофоров, каждый из которых управляет определенными потоками движения. В таких условиях система должна не только детектировать все видимые светофоры, но и корректно определить их принадлежность к конкретным направлениям движения и полосам. Ошибочная интерпретация сигнала может привести к критическим ситуациям, включая нарушение правил дорожного движения или создание аварийных ситуаций.

Дополнительную сложность представляет необходимость работы в реальном времени с ограниченными вычислительными ресурсами бортовых систем. В отличие от лабораторных условий, где возможна обработка данных мощными серверными системами, бортовые компьютеры автономных транспортных средств имеют ограничения по энергопотреблению, тепловыделению и вычислительной мощности. Это требует разработки эффективных алгоритмов, способных обеспечивать высокую точность распознавания при минимальных вычислительных затратах [16].

Ко всем перечисленным сложностям добавляется необходимость обработки данных с нескольких камер, поскольку одна камера не всегда может обеспечить полный обзор перекрестка и всех светофоров. Это требует не только синхронизации и объединения данных с различных ракурсов, но и сопоставления полученной визуальной информации с данными от навигационных систем для точного геопозиционирования светофоров. Сами же навигационные системы, в основном полагающиеся на GPS, подвержены значительным ошибкам в условиях плотной городской застройки, что создает дополнительные трудности для надежной привязки светофоров к цифровой карте.

Решение такой комплексной задачи требует применения современных методов компьютерного зрения, включающих глубокие нейросетевые архитектуры, алгоритмы трекинга объектов, методы фильтрации и предсказания, а также интеграцию с геоинформационными системами. Многомодальный подход, объединяющий данные различных сенсоров (камеры, лидары, радары) и источников информации (цифровые карты, V2X коммуникации), позволяет повысить надежность и точность системы идентификации светофоров. Это делает задачу идентификации светофоров одной из наиболее технологически насыщенных и исследовательски значимых в области автономного транспорта [3; 17].

1.1.2 Уровни автономности транспортных средств в контексте подсистемы идентификации светофоров

Развитие автономных транспортных средств требует соблюдения строгих стандартов функциональной безопасности. основополагающим международным стандартом является ISO 26262 “Автотранспортные средства. Функциональная безопасность”, определяющий требования к разработке и производству автомобильных электрических и электронных систем (Е/Е) [18]. Хотя данный

стандарт изначально разрабатывался для автомобилей, его принципы и требования в равной степени применимы и к другим типам автономных транспортных средств, включая автобусы, грузовые автомобили и трамваи. Стандарт признаёт сложность современных транспортных технологий и обеспечивает согласование практик обеспечения безопасности с техническими достижениями в области автономного вождения. В Российской Федерации действует ГОСТ Р ИСО 26262 серии стандартов, являющийся адаптацией международного стандарта ISO 26262 для российских условий [19—21].

Стандарт определяет четыре уровня целостности автомобильной безопасности (ASIL – Automotive Safety Integrity Level): от ASIL A (наименее критичный) до ASIL D (наиболее критичный). Подсистемы идентификации светофоров для автономных транспортных средств обычно классифицируются как ASIL B или ASIL C в зависимости от их роли в обеспечении безопасности [22].

В то время как классификация ASIL относится к оценке рисков и надежности отдельных компонентов, для описания общего уровня автономности транспортного средства используется иная система. Наиболее распространенной и общепринятой в мировой практике является классификация, предложенная сообществом автомобильных инженеров (SAE International). В отличие от стандартов функциональной безопасности, сфокусированных на предотвращении сбоев, стандарт SAE J3016 [23] определяет уровни автоматизации транспортных средств (от 0 до 5), исходя из степени вовлеченности водителя в процесс управления. Несмотря на наличие государственных стандартов, именно классификация SAE стала де-факто международным стандартом для определения возможностей любых типов автономных транспортных средств (см. таблицу 1).

Таблица 1 — Уровни автоматизации транспортных средств по классификации SAE J3016:2021

Уровень	Название	Описание
L0	Отсутствие автоматизации	Полное ручное управление. Водитель выполняет все задачи по управлению транспортным средством.
L1	Помощь водителю	Система помогает водителю с рулевым управлением ИЛИ ускорением/торможением. Примеры: круиз-контроль, система помощи удержания в полосе.
L2	Частичная автоматизация	Система может контролировать и рулевое управление, И ускорение/торможение. Водитель должен постоянно контролировать систему и быть готовым к перехвату управления.
L3	Условная автоматизация	Автомобиль способен самостоятельно управлять движением в определённых условиях, но водитель должен быть готов взять управление по запросу системы.
L4	Высокая автоматизация	Полная автоматизация в определённых условиях. Участие водителя не требуется в штатных ситуациях.
L5	Полная автоматизация	Полная автоматизация в любых условиях. Человек выступает только в роли пассажира.

В зависимости от уровня автоматизации системы ADAS роль подсистемы идентификации светофоров существенно различается. На начальных уровнях, L0-L2, данная подсистема реализует преимущественно пассивные функции информационной поддержки, включающие обнаружение светофоров в поле зрения, определение их текущих сигналов и своевременное информирование водителя через интерфейс оповещения [7; 8].

Начиная с уровня L3, подсистема идентификации светофоров становится активным компонентом управления транспортным средством [24; 25]. Она не

только детектирует и классифицирует сигналы светофоров, но и непосредственно влияет на поведение автомобиля (см. рисунок 1.1):

- автоматическое торможение на красном сигнале светофора;
- контроль скорости при приближении к перекрёстку;
- принятие решений о проезде перекрёстка на основе анализа сигналов;
- координация с другими подсистемами для безопасного маневрирования.

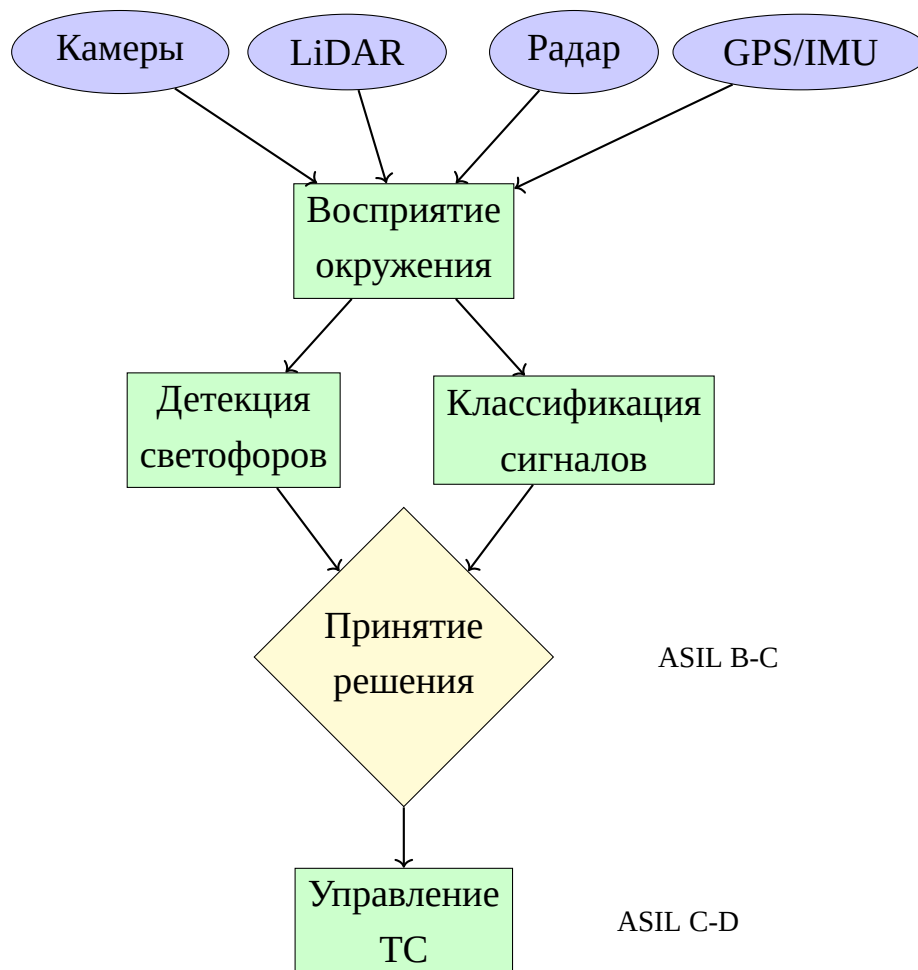


Рисунок 1.1 — Архитектура подсистемы идентификации светофоров в автономных транспортных средствах

Особая значимость подсистемы идентификации светофоров обусловлена комплексом взаимосвязанных факторов. Прежде всего, безопасность движения напрямую зависит от корректности распознавания сигналов — любые ошибки в работе системы могут привести к критическим ситуациям и авариям [3;

14]. Существенную роль играет и сложность самой задачи, требующей учёта множества внешних факторов, таких как условия освещения, погодные явления, возможные перекрытия объектами, а также разнообразие типов светофоров. Не менее важным является требование надёжности — система должна обеспечивать высокую точность работы в режиме реального времени при любых условиях эксплуатации. Кроме того, эффективность подсистемы во многом определяется качеством её интеграции с другими системами автомобиля и элементами дорожной инфраструктуры [26].

Именно из-за этих факторов подсистема идентификации светофоров является одной из наиболее исследовательски значимых в области ADAS. Её полноценная интеграция в качестве активного компонента управления происходит только на высоких уровнях автоматизации L3-L5, когда системы достигают достаточной надёжности и способности к комплексному анализу дорожной ситуации [3].

На уровнях L4 и L5 подсистема идентификации светофоров должна обеспечивать безошибочное распознавание в любых условиях, что требует применения передовых технологий компьютерного зрения, искусственного интеллекта и мультисенсорной интеграции [27—29].

Применение подсистем идентификации светофоров в трамваях имеет ряд специфических особенностей и сложностей. Это накладывает дополнительные требования на точность и надёжность распознавания сигналов, особенно учитывая более длинный тормозной путь трамвая по сравнению с автомобилями. Примером успешной реализации таких систем является проект автономного трамвая от Siemens, где распознавание сигналов является одной из ключевых технологий [30].

1.1.3 Подходы к решению задачи

В данном разделе рассматриваются различные источники данных и подходы, используемые для решения задачи автоматического распознавания светофоров. Подходы можно условно разделить на две категории: основанные на взаимодействии с дорожной инфраструктурой, и использующие бортовые сенсорные системы транспортного средства.

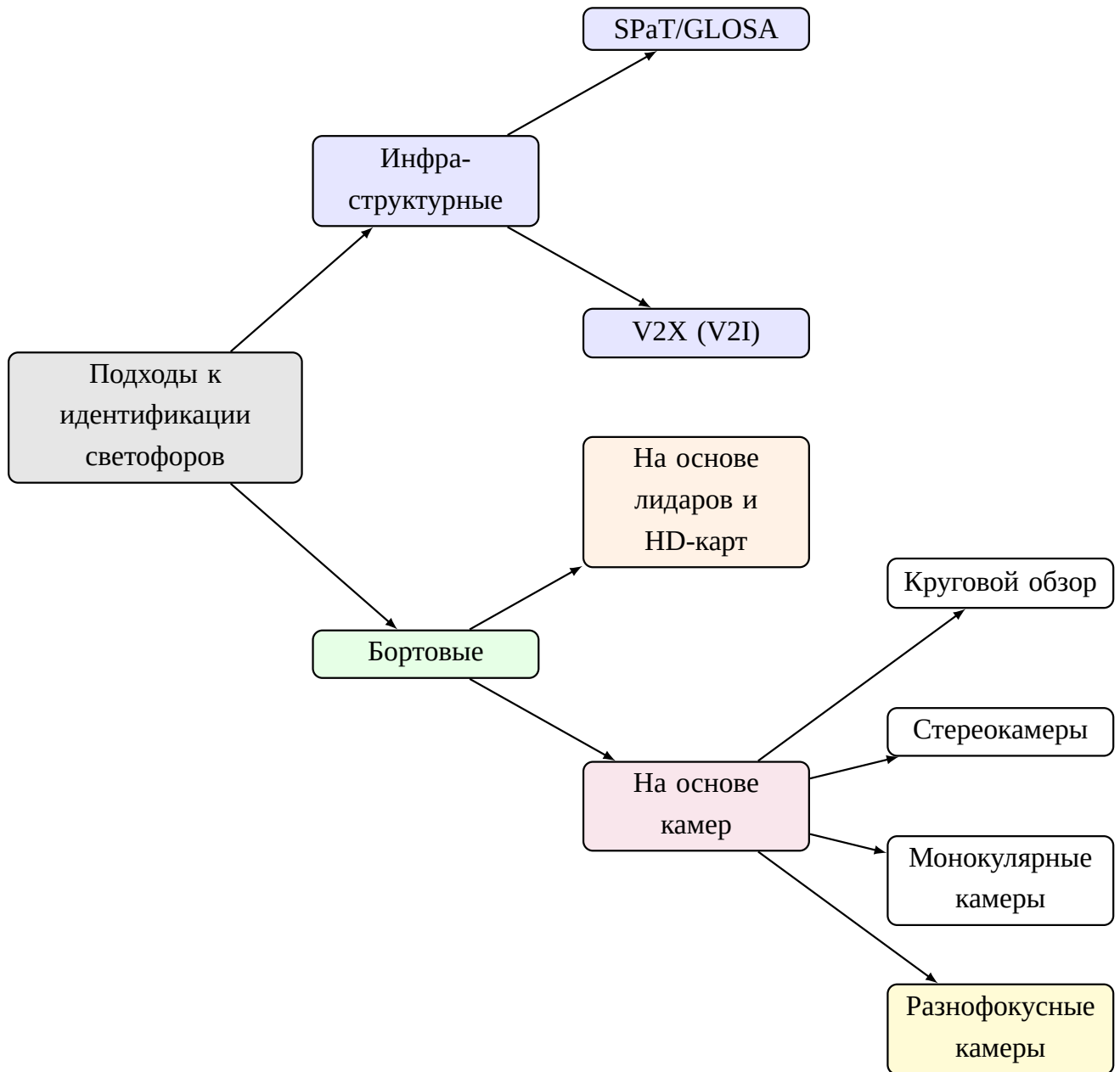


Рисунок 1.2 — Классификация подходов к идентификации светофоров

Взаимодействие с инфраструктурой. Технологии Vehicle-to-Everything (V2X) представляют собой системы коммуникации, позволяющие любым транспортным средствам обмениваться информацией с другими участниками дорожного движения и элементами дорожной инфраструктуры (см. рисунок 1.3). Это взаимодействие включает Vehicle-to-Infrastructure (V2I), Vehicle-to-Vehicle (V2V), Vehicle-to-Pedestrian (V2P) и Vehicle-to-Network (V2N) коммуникации [31]. Основная цель V2X – повышение безопасности дорожного движения,

оптимизация транспортных потоков и предоставление водителям и системам автономного вождения актуальной информации об окружающей обстановке.

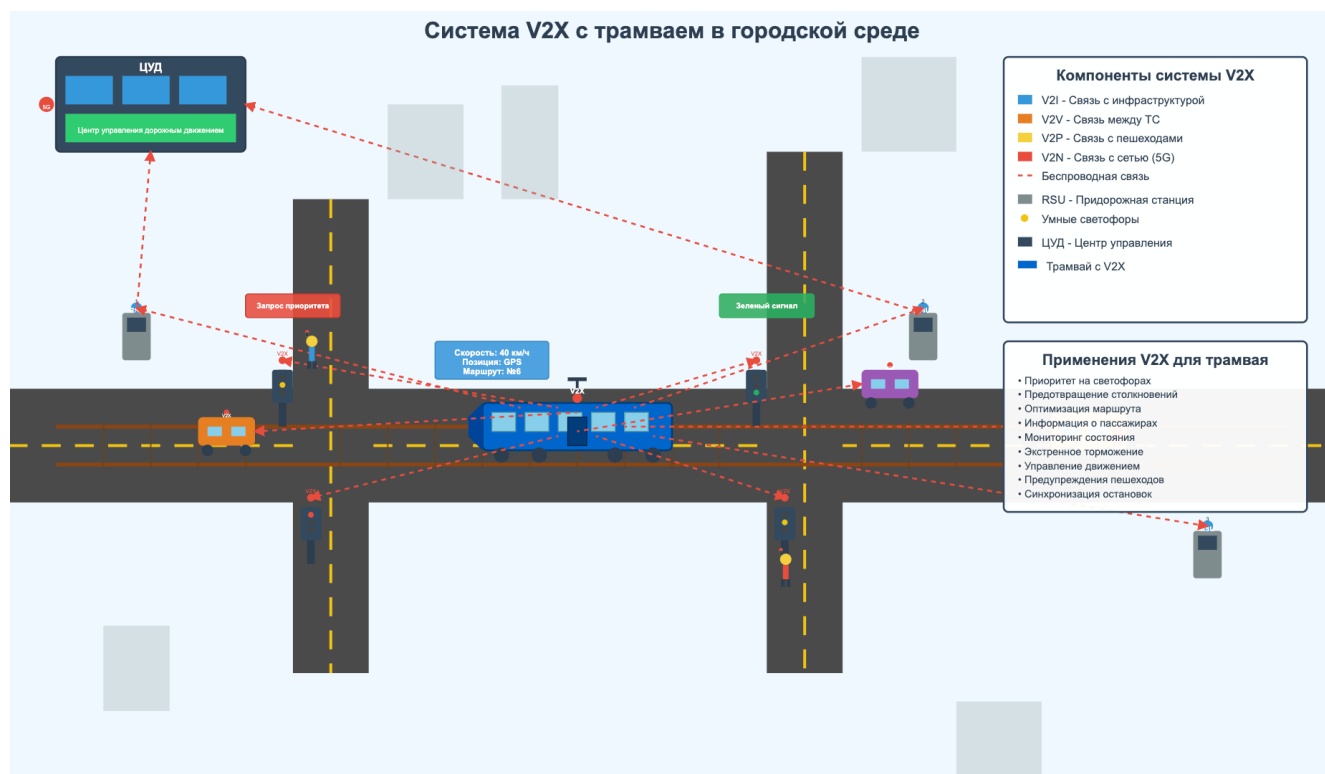


Рисунок 1.3 — Взаимодействие транспортного средства с дорожной инфраструктурой

В контексте распознавания светофоров ключевую роль играет V2I взаимодействие. Современные интеллектуальные транспортные системы (ИТС) активно развивают стандарты и протоколы для передачи информации от светофорных объектов непосредственно на борт транспортного средства. Одним из таких механизмов является передача сообщений SPaT (Signal Phase and Timing), которые содержат данные о текущей фазе сигнала светофора и прогнозируемом времени до ее окончания для каждой полосы движения на перекрестке [32]. Эта информация позволяет системам помощи водителю (СПВ) и автономным транспортным средствам (АТС) заранее адаптировать скорость и манеру движения, что способствует экономии топлива и повышению плавности проезда перекрестков. Системы типа GLOSA (Green Light Optimized

Speed Advisory) используют SPaT-сообщения для предоставления водителю рекомендаций по оптимальной скорости движения, чтобы прибыть к перекрестку на зеленую фазу сигнала [33].

Исследования показывают, что использование V2X технологий, и в частности SPaT и GLOSA, может значительно улучшить координацию транспортных потоков на перекрестках и повысить общую эффективность дорожного движения [34]. Получение точной информации о фазах светофоров напрямую от инфраструктуры позволяет снизить зависимость от систем компьютерного зрения, особенно в сложных погодных условиях или при ограниченной видимости, а также уменьшить задержки на перекрестках.

Несмотря на очевидные преимущества, массовое внедрение V2X-подходов для идентификации светофоров сталкивается с рядом существенных проблем. Ключевым ограничением является необходимость создания и поддержания сложной и дорогостоящей дорожной инфраструктуры. Развертывание такой инфраструктуры — это масштабная задача, требующая значительных капиталовложений и координации на государственном уровне. По некоторым оценкам, стоимость прокладки 1 км волоконно-оптических линий связи, необходимых для V2X, может достигать 1-2 миллионов рублей, а общие инвестиции в создание покрытия на основных магистралях могут исчисляться сотнями миллиардов [35].

Помимо прямых финансовых затрат, существуют и другие серьезные технологические и организационные вызовы [36—38]. Одной из ключевых проблем является отсутствие единых стандартов для оборудования и протоколов связи, что создает барьеры для совместимости решений разных производителей и замедляет повсеместное внедрение технологии. Не менее важным аспектом является обеспечение кибербезопасности, поскольку централизованная система связи становится потенциальной мишенью для кибератак, и требуется высочайший уровень защиты для исключения злонамеренного вмешательства в управление трафиком. Наконец, для приложений, критически важных для безопасности, необходимо гарантированное покрытие сети и минимальная

задержка сигнала, чего сложно достичь на всей протяженности дорожной сети, особенно в удаленных или труднодоступных районах.

Таким образом, хотя технологии V2X и открывают большие перспективы для интеллектуального управления светофорным регулированием, их широкое применение в ближайшем будущем ограничено инфраструктурными и экономическими факторами. Вследствие этого, современные СПВ и АТС в большинстве своем вынуждены полагаться на существующую дорожную инфраструктуру и использовать бортовые сенсорные системы (камеры, лидары, радары) для автономного детектирования и распознавания состояния светофоров. Дальнейшие разделы будут посвящены именно таким бортовым решениям.

Бортовые решения на основе лидаров и HD-карт. Следующий класс подходов к восприятию и анализу окружающей обстановки автономными транспортными средствами, особенно на высоких уровнях автоматизации L4/L5, связан с использованием бортовых сенсорных систем в сочетании с высокоточными цифровыми картами (HD-картами, см. рисунок 1.4). Такие компании, как Waymo, активно применяют комбинацию из множества лидаров, радаров и камер, данные которых комплексировются и сопоставляются с HD-картой для точной локализации и восприятия окружающей обстановки [39].

HD-карты представляют собой детализированные многослойные цифровые представления дорожной среды, выполненные с сантиметровой точностью. В отличие от традиционных навигационных карт, они содержат не только дорожный граф, но и семантическую информацию о всех элементах дорожной инфраструктуры. Структура таких карт обычно включает несколько ключевых слоев:

- Геометрический слой описывает трехмерную геометрию дорожного полотна, включая бордюры, разделительные полосы и рельеф местности.
- Слой дорожной разметки содержит информацию о полосах движения, их границах (сплошные, прерывистые линии), направлениях движения, а также о стоп-линиях и пешеходных переходах.

- Слой объектов включает точное расположение и тип статических объектов, таких как дорожные знаки, столбы, мачты светофоров и здания.
- Семантический слой определяет взаимосвязи между объектами, например, к какой полосе относится тот или иной светофор, какие маневры разрешены на перекрестке и скоростные ограничения.

Для представления этой сложной информации были разработаны различные форматы, среди которых выделяется Lanelet2 [40]. Этот формат использует атомарные элементы (точки и полилинии) для построения более сложных структур — «лейнлетов», представляющих собой участки полос движения с постоянными свойствами. Лейнлеты объединяются в дорожный граф, а регуляторные элементы, такие как светофоры или знаки приоритета, привязываются к ним, определяя правила дорожного движения. Альтернативные форматы, такие как OpenDRIVE, также широко используются в индустрии, однако они в большей степени ориентированы на моделирование геометрии дорог для симуляций, в то время как Lanelet2 предлагает более гибкую семантическую аннотацию, необходимую для принятия решений в реальном времени. Использование таких карт позволяет системе управления предвидеть изменение дорожной ситуации, планировать траекторию движения и повышать общую надежность системы. Лидары, в свою очередь, обеспечивают точное трехмерное сканирование окружающего пространства при различных погодных и световых условиях, позволяя детектировать объекты и измерять расстояние до них с высокой точностью [41; 42].

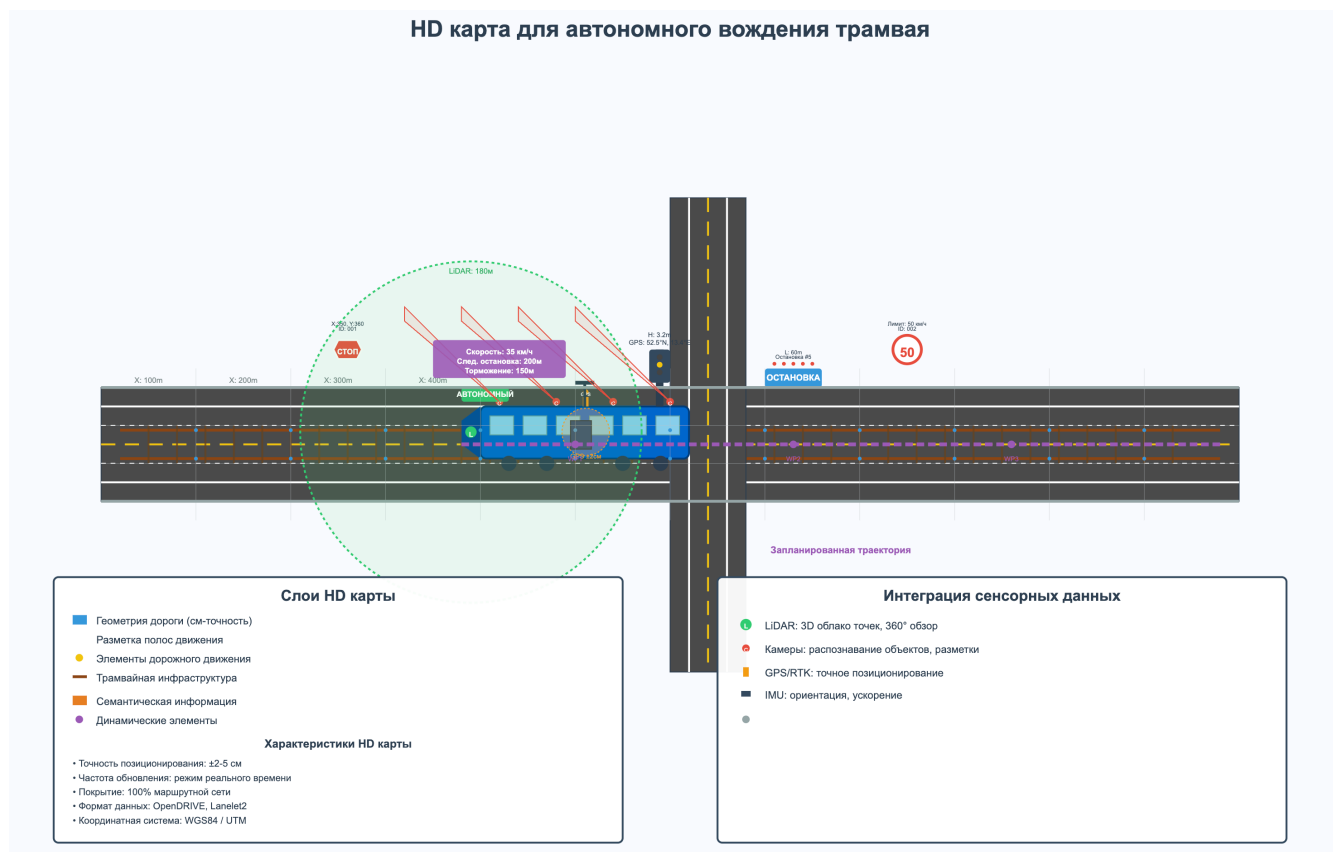


Рисунок 1.4 — Высокоточная цифровая карта дорожной сцены

Системы, основанные на HD-картах и лидарах, при всех своих достоинствах, сталкиваются с серьезными практическими ограничениями при массовом внедрении. Ключевой проблемой выступает необходимость поддержания HD-карт в актуальном состоянии. Начальное создание высокоточных карт для обширных территорий является ресурсоемким процессом, требующим привлечения дорогостоящих мобильных картографических систем и значительных временных затрат [39]. Ситуация осложняется динамичным характером городской среды - регулярно прокладываются новые маршруты, меняются схемы организации движения, происходит реконструкция светофоров и установка новых дорожных знаков. Своевременное отражение этих изменений в HD-картах требует создания сложной системы постоянного мониторинга и оперативного обновления картографических данных, что существенно повышает стоимость эксплуатации таких решений [39].

Что касается лидаров, то, несмотря на их способность точно определять геометрию объектов и расстояние до них, они имеют ограничения в контексте задачи распознавания состояния светофоров. Лидары не предоставляют информацию о цвете объектов, что делает невозможным прямое определение активного сигнала светофора (красный, желтый, зеленый) только на основе их данных [43]. Распознавание цвета сигнала является прерогативой систем технического зрения, использующих видеокамеры. Лидары могут помочь в обнаружении самого объекта светофора и определении его положения, но для полной идентификации состояния требуется комплексирование с другими сенсорами, в первую очередь с камерами [43].

Кроме того, стоимость высокопроизводительных лидаров, обеспечивающих необходимое разрешение и дальность для уровней L4/L5, остается высокой [42; 44]. Хотя цены на лидары постепенно снижаются, их интеграция в системы помощи водителю более низких уровней (L2/L3), к которым относится СПВ для трамваев, часто оказывается экономически нецелесообразной. Такие системы должны быть доступны для массового внедрения и иметь разумную стоимость. Интересно отметить, что некоторые компании, например Tesla, изначально сделавшие ставку на системы L4/L5, в настоящее время активно развивают СПВ, полагаясь преимущественно на техническое зрение, избегая использования дорогостоящих лидаров в своих серийных автомобилях [44], что подчеркивает стремление к поиску более масштабируемых и экономически эффективных решений.

Таким образом, несмотря на высокую точность и потенциал систем, основанных на HD-картах и лидарах, проблемы их создания, дорогостоящего поддержания и ограниченной применимости лидаров для непосредственного распознавания сигналов светофоров делают их использование затруднительным для СПВ уровня L2/L3, ориентированных на массовое применение. В связи с этим, для решения задачи идентификации светофоров в таких системах основной упор делается на бортовые сенсоры, способные напрямую воспринимать и интерпретировать состояние светофоров, т.е. камеры.

Бортовые решения на основе камер и проблема релевантности светофоров. На городских перекрестках зачастую присутствует множество светофоров, предназначенных для различных полос движения, направлений поворота, а также для разных категорий участников движения (например, автомобили, пешеходы, велосипедисты, трамваи), как представлено на рисунке 1.5. Для безопасного и корректного принятия решений система управления транспортным средством должна однозначно идентифицировать, какой именно из видимых светофоров регулирует движение по ее полосе и намеченному маршруту. Эта задача, известная как «проблема сопоставления светофора с полосой движения» или определение релевантности светофора и является нетривиальной [45]. Ошибочное определение релевантного светофора может привести к проезду на запрещающий сигнал или, наоборот, к необоснованной остановке, что напрямую влияет на безопасность и эффективность движения.

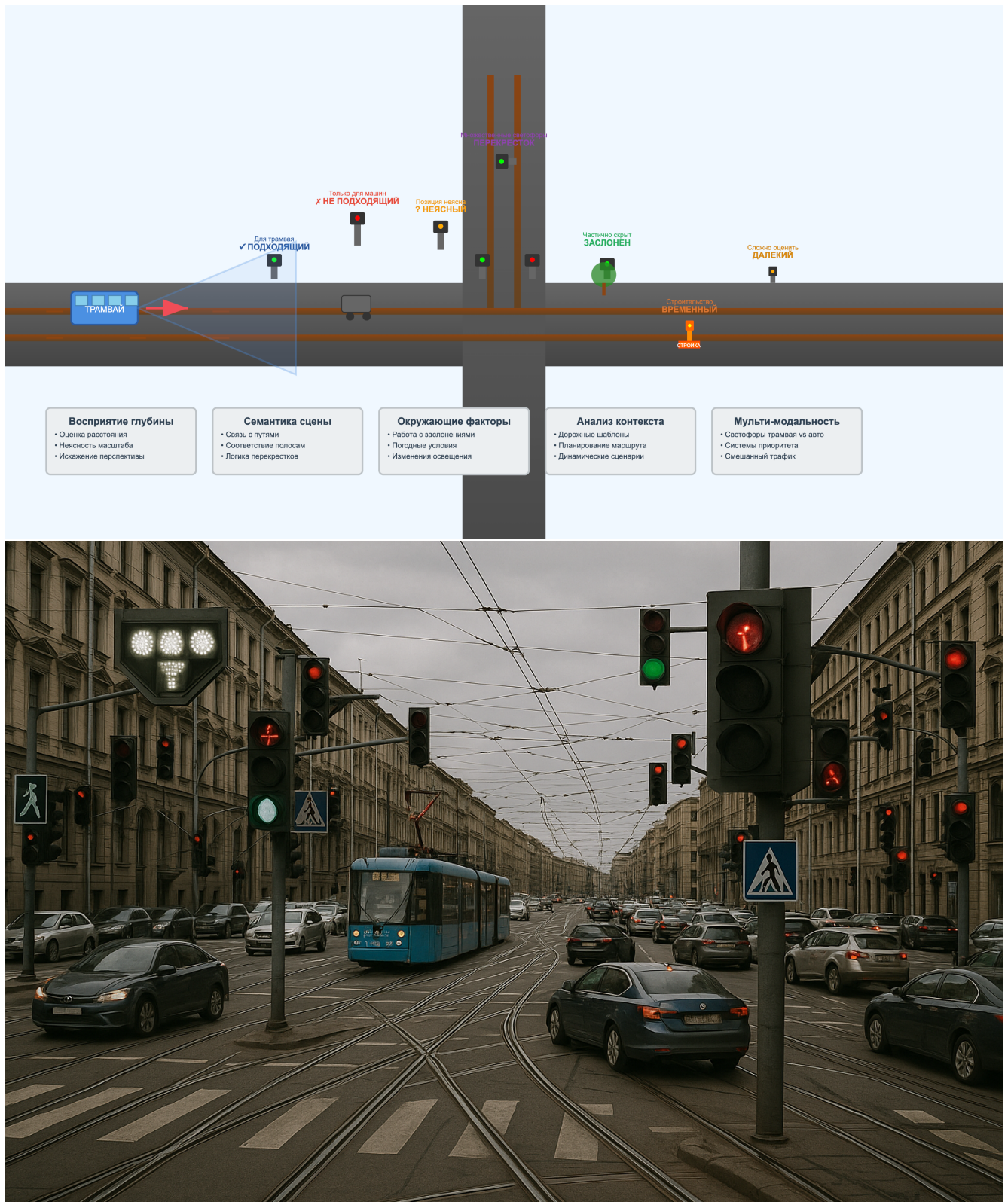


Рисунок 1.5 — Проблема сопоставления светофора с полосой движения

Исследователи предлагают различные подходы для решения проблемы релевантности. Некоторые работы фокусируются на анализе визуальных признаков, таких как «стрелки направления» на сигнальных секциях светофора или анализе дорожной разметки, для установления связи между светофором и полосой движения [15]. Однако такие методы могут быть ненадежны в условиях плохой видимости, отсутствия или нечеткой разметки, а также на сложных перекрестках с нестандартной конфигурацией светофоров. Попытки полностью автоматического формирования семантической информации о соответствии светофоров полосам движения только на основе сенсорных данных без какой-либо априорной информации сталкиваются со значительными трудностями в обеспечении требуемой для систем L2/L3 надежности [45]. Как отмечается в [46], создание надежной системы восприятия светофоров включает в себя не только их детекцию, но и сложный процесс ассоциации и принятия решений.

Для повышения надежности определения релевантных светофоров, особенно в системах ADAS L2/L3, где использование дорогостоящих HD-карт и лидаров ограничено, перспективным направлением является применение «упрощенных» цифровых карт дорог (ЦКД) с векторным представлением светофоров. В отличие от полномасштабных HD-карт, требующих непрерывного обновления и высокой точности позиционирования, такие карты могут содержать базовую геометрию перекрестка, приблизительное расположение светофорных объектов и информацию о том, какие полосы или направления они регулируют. При этом карты могут быть созданы и обновляться с меньшими затратами, например, на основе анализа спутниковых снимков или данных, собранных коммунальными службами [14].

Интеграция данных с бортовых камер с информацией из таких «упрощенных» карт позволяет существенно снизить неоднозначность при определении релевантного светофора. Например, зная примерное положение автомобиля (полученное от GPS/ГЛОНАСС приемника) и геометрию перекрестка из карты, система может ограничить поиск релевантного светофора определенной зоной и сопоставить детекции с камер с картографическими данными [14]. Такой

подход позволяет идентифицировать светофоры, отфильтровать относящиеся к другим полосам или направлениям движения, и сосредоточиться на тех, которые действительно важны для транспортного средства. Методы, объединяющие нейронные сети для детекции светофоров и карты, показывают хорошие результаты в решении задач идентификации [14]. Даже при создании HD-карт используются GPS и стереокамеры для точного определения положения светофоров [47], что подчеркивает важность пространственной привязки.

Для систем ADAS уровня L2/L3 идентификация релевантных светофоров исключительно на основе бортовых камер без какой-либо картографической информации представляется чрезмерно сложной задачей для достижения необходимого уровня безопасности и надежности. Использование упрощенных карт, содержащих основную топологию перекрестка и расположение светофоров, в сочетании с данными от систем технического зрения, является оптимальным способом, позволяющим повысить точность и надежность функции распознавания светофоров при сохранении приемлемой стоимости системы. Именно такой подход, основанный на комплексировании данных с видеокамер и информации из упрощенной карты, заложен в основу методики, разрабатываемой в настоящей диссертационной работе.

Выбор конфигурации камер для бортовых систем. При проектировании бортовых систем распознавания светофоров, использующих камеры, ключевым аспектом является выбор их конфигурации и типа. Различные конфигурации обладают своими преимуществами и недостатками, влияющими на дальность обнаружения, точность пеленга и общую надежность системы.

Системы кругового обзора, часто реализуемые с использованием четырех широкоугольных («рыбий глаз») камер, предназначены в первую очередь для задач ближней зоны, таких как помощь при парковке, контроль слепых зон или ассистент движения в пробках [48]. Такие камеры обеспечивают 360-градусный обзор вокруг транспортного средства, однако их разрешающая способность и точность на больших расстояниях, необходимых для заблаговременного распознавания светофоров (например, для трамваев это может быть порядка

100 метров и более), как правило, недостаточны. Таким образом, для задач идентификации светофоров на значительных дистанциях, где требуется преимущественно фронтальный обзор на дальние дистанции, системы кругового обзора не являются оптимальным решением.

Стереокамеры, состоящие из двух камер, разнесенных на известное расстояние (база), позволяют получать информацию о глубине сцены и, соответственно, оценивать расстояние до объектов методом триангуляции [49]. Это свойство полезно для трехмерного позиционирования светофоров. Однако стереосистемы обладают рядом существенных недостатков. Во-первых, они чувствительны к точности калибровки и взаимному расположению камер: даже небольшие смещения или вибрации могут приводить к значительным ошибкам в оценке глубины [50]. Во-вторых, точность оценки глубины стереопарой снижается с увеличением расстояния до объекта [50; 51], что делает их менее эффективными для распознавания удаленных светофоров, особенно на расстояниях свыше 50-70 метров, критичных для трамвайных СПВ. Как отмечается в [51], стереозрение не является надежным на больших дистанциях.

Системы с одной камерой (монокулярные) являются распространенным решением для задачи детекции и распознавания светофоров благодаря своей простоте и относительной дешевизне [9—11; 16; 51; 52]. Современные алгоритмы глубокого обучения позволяют достигать высокой точности в детекции и классификации светофоров с использованием одной камеры. Однако определение точного расстояния до объекта с помощью одной камеры является сложной задачей и часто требует дополнительных эвристик или обучения на больших объемах данных с использованием информации о размерах объектов.

Для преодоления ограничений как стереосистем (в части дальности и калибровки), так и монокулярных систем (в части точной оценки расстояния и работы в широком диапазоне дальностей), в данной диссертационной работе предлагается использовать конфигурацию из двух камер с различными фокусными расстояниями, установленных соосно. Такой подход предполагает использование одной камеры с широкоугольным объективом для обзора ближней

зоны и обнаружения светофоров на небольших расстояниях (например, при подъезде к перекрестку), и второй камеры с телеобъективом (длиннофокусной) для обнаружения и распознавания светофоров на значительных расстояниях. Подобные многокамерные системы с различными полями зрения (FOV - Field of View) для разных дальностей успешно применяются в современных разработках систем восприятия для автономного вождения [46; 51]. Например, в [46] используется комбинация из трех разнофокусных камер, а в [51] – узкоугольная камера для дальнего обзора и широкоугольная для ближнего. Такая конфигурация позволяет эффективно охватывать весь необходимый диапазон расстояний для надежного распознавания светофоров, что особенно актуально для трамваев, требующих заблаговременного получения информации о сигналах для плавного и безопасного управления движением. Отсутствие необходимости в точном стереосопоставлении упрощает калибровку и повышает устойчивость системы к небольшим взаимным смещениям камер. При этом, задача оценки расстояния может решаться либо косвенно (например, по проективному размеру детектированного светофора и знанию его типовых физических размеров), либо с использованием нейросетевых подходов, обученных оценивать расстояние по изображению с одной или обеих камер.

Применение 3D сверточных нейронных сетей для оценки положения и размеров светофоров. Как было отмечено ранее, предлагаемая в работе двухкамерная система с различными фокусными расстояниями не образует классическую стереопару, что затрудняет прямое использование методов триангуляции для точной оценки расстояния до объектов. Однако, информация о трехмерном положении светофоров критически важна для безопасного управления движением. Современные подходы к распознаванию светофоров широко используют сверточные нейронные сети (СНС) для задач детекции и классификации их состояния по двумерным изображениям [3; 17; 53; 54]. Эти сети демонстрируют высокую эффективность анализа визуальной информации и извлечении признаков, необходимых для идентификации светофоров.

В рамках данной диссертационной работы предлагается развитие этого направления путем применения специализированных архитектур 3D сверточных нейронных сетей (3D СНС) для обработки данных с каждой из камер. Новизна заключается в самом факте использования 3D СНС для задачи идентификации светофоров и оценки их трехмерного положения по изображениям, тогда как большинство существующих подходов ограничиваются применением 2D СНС.

Используемая нейронная сеть обучается неявно выводить информацию о глубине и размерах светофора, анализируя пространственные признаки непосредственно из одного изображения, в отличие от методов, требующих явного стереосопоставления или сложных эвристик для монокулярной оценки глубины. Вопросы сопоставления и комплексирования детектированных светофоров с разных камер для формирования единого представления и принятия окончательного решения рассматриваются в последующих главах данной работы. Ожидается, что использование 3D СНС для каждой камеры позволит повысить точность и надежность подсистемы идентификации светофоров [55].

1.2 Постановка задачи и этапы ее решения

Система помощи водителю трамвая состоит из двух камер с объективами 90° и 40° , установленных в кабине водителя трамвая. С каждой камерой связана система координат xyz с центром в точке C . Камеры считаются откалиброванными, с известными внутренними $K_m^{<cam/vr>}$ и внешними $R_m^{<cam/vr>}, T_m^{<cam/vr>}$ параметрами установки относительно системы координат трамвая, где m – номер камеры. Начало системы координат трамвая (далее Vehicle Related – VR), изображенное на рисунке 1.6, совпадает с центром переднего бампера и находится на земле. В системе координат VR производится первичный анализ дорожной сцены. Кроме того, навигационная система трамвая определяет его положение и ориентацию в глобальной метрической системе

координат (Global Metric Coordinate System). GM система координат позволяет определять положение любого объекта в пространстве на основе вектора $s_t^{gm} = [x_t, y_t, z_t, \alpha_t, \beta_t, \theta]$, где x_t, y_t, z_t – координаты объекта, $\alpha_t, \beta_t, \theta$ – углы его ориентации. Очевидно, что для перехода из системы координат Cam в GM необходимо последовательно выполнить переходы из Cam в VR и из VR в GM .



Рисунок 1.6 — Изображение трамвая и системы координат VR

Координаты светофоров в системе координат GM заранее известны. Обозначим их в виде вектора $o_b^{gm} = [x, y, z, \theta]$, где b – порядковый номер светофора, x, y, z – координаты и θ – угол ориентации светофора. На каждом перекрестке светофоры объединены в группы. Под группой светофоров понимается совокупность светофоров (основных и дублирующих), регулирующих движение на одном перекрестке и имеющих общую стоп-линию для остановки транспортного средства. Для уменьшения обозначений дополнительный индекс группы светофоров вводить не будем.

Нейронная сеть распознает светофоры по изображениям, формируемым каждой камерой. Сеть является многоголовой (многозадачной). Каждая из голов

сети имеет свое назначение. Можно выделить: решение задачи сегментации, 3D-детектирования, формирования эмбеддингов и другие. На рисунке 1.7 показан результат работы сети.



Рисунок 1.7 — Визуализация распознавания дорожной сцены (3D детекции – машины (желтые), светофоры (красные), область сегментации – бордюры (красная), трамвайное полотно (бирюзовое))

В каждый момент времени нейронная сеть получает на вход изображение I_t^m . Нейронная сеть находит объекты: транспортные средства (включая легковые и грузовые транспортные средства, общественный транспорт), пешеходов, знаки дорожного движения, светофоры и определяет их пространственное положение, описываемое параллелепипедами (далее 3D-детекциями или просто детекциями).

Детекцию светофора j в момент времени t для камеры m обозначим как $r_j^{t,m}$. Вектор измерений $r_j^{t,m}$ включает в себя:

- координаты (пиксели) $[u, v]$ центра светофора на изображении I_t^m ;
- глубину центра светофора d (depth) в системе координат камеры, m ;
- углы ориентации светофора в пространстве (тангаж (pitch), крен (roll), рыскание (yaw));

- физические размеры объекта: ширина (w), длина (l) и высота (h), м;
- уникальный идентификатор (вектор признаков) области светофора - эмбединг (E).

Все вышеописанные параметры за исключением u, v формируются в системе координат камеры Cam .

Требуется определить соответствие между j -ой детекцией и порядковым номером светофора b , по измерениям $r_j^{t,m}$. Для решения поставленной задачи выполняются следующие этапы (см. рис. 1.8):

1. **Трекинг** – сопоставление детекций светофоров между кадрами для каждой камеры в отдельности, т.е. поиск соответствия между $r_{j1}^{t-1,m}$ и $r_{j2}^{t,m}$;
2. **Объединение** результатов детекций с разных камер для одного и того же светофора в единственное измерение, т.е. поиск соответствия между $r_{j1}^{t,m}$ и $r_{j2}^{t,m+1}$;
3. **Фильтрация** положения светофоров – определение координат светофоров в системе координат GM $s_j^{t,gm}$ по нескольким измерениям;
4. **Сопоставление** полученных координат по результатам работы фильтра со светофорами, имеющимися на карте o_b^{gm} ;
5. **Классификация** – определение класса сигнала светофора $c_j^{t,m}$ по изображению, для принятия решения о возможности проезда.

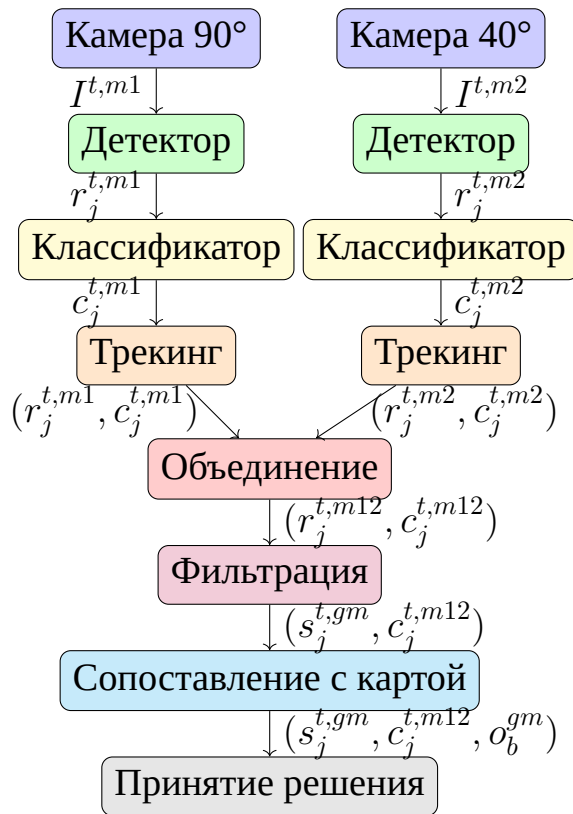


Рисунок 1.8 — Схема алгоритма идентификации светофоров

Выводы

В данной главе был проведен анализ существующих решений в области идентификации светофоров для СПВ и АТС, а также сформулирована задача исследования. Основные выводы заключаются в следующем:

- Распознавание светофоров является критически важной задачей для обеспечения безопасности автономного движения, особенно для рельсового транспорта, такого как трамваи, ввиду их специфических эксплуатационных характеристик.
- Существует множество факторов, усложняющих задачу автоматического распознавания светофоров, включая разнообразные условия окружающей среды (освещение, погода); динамические помехи (перекрытия, другие источники света); различия в стандартах и типах светофорного оборудования; помехи, влияющие на работу навигационных систем;

асинхронная работа камер; а также необходимость работы в реальном времени с ограниченными вычислительными ресурсами.

- Подчеркнута возрастающая роль и требования к надежности систем идентификации светофоров с повышением уровня автоматизации транспортных средств согласно классификации SAE и стандартам функциональной безопасности ISO 26262.
- Проанализированы различные подходы к решению задачи: технологии V2X, системы на основе лидаров и HD-карт, а также бортовые системы на основе видеокамер. Отмечены преимущества и недостатки каждого подхода, обоснован выбор в пользу камерных систем с использованием упрощенных карт для разрабатываемой системы помощи водителю трамвая (уровень L2/L3).
- Обоснована конфигурация сенсорной системы, состоящей из двух соосно расположенных камер с различными фокусными расстояниями (широкоугольная 90° и длиннофокусная 40°), позволяющая эффективно обнаруживать светофоры на разных дистанциях без сложностей, присущих стереосистемам.
- Предложен алгоритм решения поставленной задачи, включающий трекинг детекций на каждой камере, объединение данных с двух камер, фильтрацию для определения координат в глобальной системе, сопоставление с картой и классификацию сигнала светофора.

Глава 2. Методика сопоставления детекций в двухкамерной разнофокусной системе

2.1 Обзор методов трекинга и ассоциации детекций

2.1.1 Межкадровое сопоставление детекций

Межкадровое сопоставление детекций (или отслеживание объектов), часто называемое в англоязычной литературе Multi-Object Tracking (MOT), является критически важным компонентом в системах автономного вождения. Его основная функция заключается в присвоении постоянных уникальных идентификаторов (UID) обнаруженным объектам на последовательности кадров. Этот процесс является фундаментальным для понимания поведения динамических агентов дорожной сцены, таких как транспортные средства, пешеходы и велосипедисты. Надежная система отслеживания (трекинга) позволяет прогнозировать траектории, обнаруживать потенциальные столкновения и обеспечивает стабильное восприятие сцены (мира) с течением времени, что необходимо для безопасного движения [56].

Проблема отслеживания объектов на видеопоследовательностях является областью исследований в компьютерном зрении. Ранние подходы в значительной степени полагались на модели движения для предсказания местоположения объекта на последующих кадрах. Одним из фундаментальных методов в этой области является применение фильтра Калмана [57] для оценки состояния, предсказывая будущее положение объекта на основе его текущего состояния (например, положения и скорости), а затем корректируя это предсказание новым измерением (новой детекцией).

Со временем парадигма, известная как “отслеживание через обнаружение” [58; 59], стала стандартным подходом для MOT [60]. Эта методология

разделяет проблему на два отдельных этапа: обнаружение объектов в каждом кадре независимо, а затем связывание этих обнаружений между кадрами для формирования траекторий.

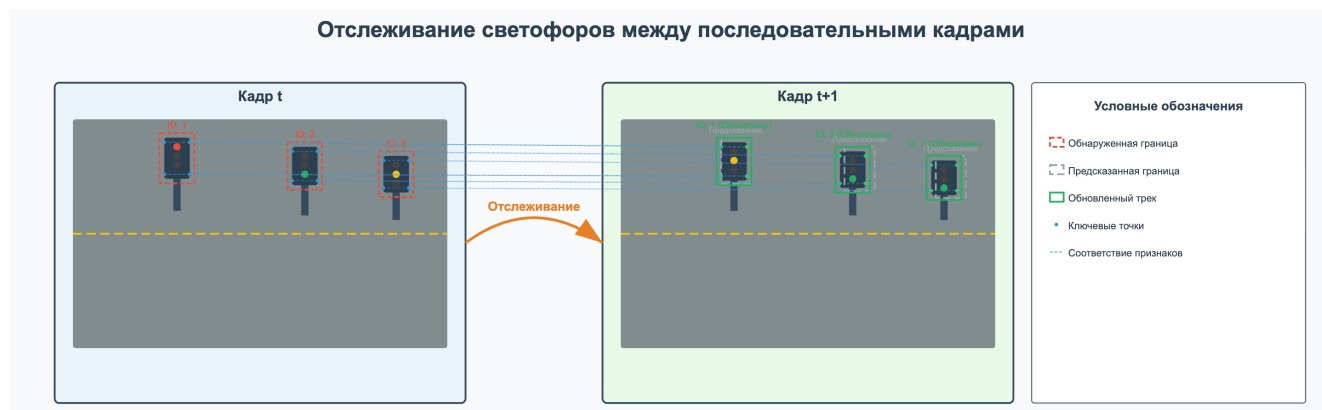


Рисунок 2.1 — Трекинг

Эволюция методов в рамках отслеживания через обнаружение демонстрирует прогресс в этой области. Ярким примером является алгоритм простого онлайн-трекинга в реальном времени (SORT) [61]. SORT оптимизировал процесс отслеживания, уделяя особое внимание эффективности для приложений реального времени. Он использует фильтр Калмана для прогнозирования движения и Венгерский алгоритм [62] для сопоставления обнаружений на основе метрики пространственного перекрытия (Intersection Over Union, IOU) их ограничивающих рамок (bbox'ов). Несмотря на свою эффективность, зависимость SORT исключительно от движения и пространственного перекрытия делает его подверженным ошибкам, особенно во время заслонений или в условиях, когда объекты имеют схожие формы и находятся в непосредственной близости.

Для устранения этих ограничений был представлен DeepSORT, улучшающий SORT за счет включения информации о внешнем виде с использованием модели глубокого обучения [63]. Предварительно обученная сверточная нейронная сеть извлекает признаки внешнего вида для каждой

обнаруженной ограничивающей рамки. Метрика ассоциации включает в себя не только IOU и расстояние на основе движения (расстояние Махаланобиса), но и косинусное расстояние между векторами признаков (эмбеддингами) внешнего вида. Добавление этой метрики делает отслеживание более надежным, позволяя повторно идентифицировать объекты даже после длительных периодов заслонений, что значительно сокращает количество смен идентификаторов.

Важно отметить, что использование векторов признаков не всегда является оптимальным для трекинга светофоров. Основная сложность заключается в том, что при смене сигнала (например, с красного на зеленый) внешний вид светофора меняется и это приводит к тому, что вектор признаков, описывающий объект, может резко измениться, и система трекинга может потерять светофор или принять его за новый объект. Поэтому повышаются требования к качеству обучения нейронной сети. При недостаточном объеме видеопоследовательностей со светофорами в обучающей выборке, эмбеддинги для светофоров будут менее полезны при сопоставлении, и поэтому стоит в большей степени полагаться на геометрические метрики сходства, которые остаются стабильными вне зависимости от активного сигнала светофора.

Несмотря на эти достижения, MOT в сложных дорожных сценах остается проблемой. Основные трудности включают частые и продолжительные заслонения, взаимодействие между похожими объектами, ошибки и несоответствия от детектора объектов, а также ограничения по обработке в реальном времени в автономных системах [64]. Будущие исследования продолжают фокусироваться на разработке более надежных методов ассоциации, улучшении возможностей повторной идентификации и более тесной интеграции отслеживания с обнаружением в рамках сквозного обучения.

2.1.2 Объединение результатов детекций между камерами

Сопоставление объектов, обнаруженных несколькими камерами, является одной из ключевых задач для систем автономного вождения, обеспечивая создание полной картины окружения. В отличие от трекинга в пределах одного видеопотока, где основной проблемой является повторная идентификация объектов после их временного исчезновения (окклюзии), межкамерное сопоставление сталкивается с дополнительными сложностями. К ним относятся различия в ракурсах, освещении, разрешении и частичные перекрытия полей зрения камер [65].

Можно выделить несколько основных подходов к решению этой задачи.

Традиционный подход на основе однокамерного трекинга. Классический метод заключается в двухэтапном процессе. Сначала для каждой камеры независимо выполняется однокамерное отслеживание (SCT). На этом этапе строятся треклеты — траектории движения объектов в поле зрения одной камеры. Затем, на втором этапе, эти треклеты сопоставляются между собой в глобальной системе координат. Для этого используются различные метрики схожести, такие как расстояние между объектами в 3D-пространстве, схожесть их визуальных признаков (цветовых гистограмм, текстур) или временное соответствие. Ярким представителем этого семейства является алгоритм DeepSORT, который был описан ранее. Однако, разделение на локальное отслеживание и глобальное сопоставление может приводить к накоплению ошибок и неоптимальным решениям, особенно в сложных сценариях с большим количеством объектов.

Глобальная модель ассоциаций с предсказанием связей. Более современные подходы стремятся объединить информацию со всех камер в единой модели для принятия совместного решения о сопоставлении. Один из таких методов предлагает использование графовой модели глобальной ассоциации (Global Association Graph) [65]. В этой модели детекции со всех камер представляются в виде вершин графа, а возможные связи между

ними — в виде ребер. Задача сводится к предсказанию наиболее вероятных связей (ребер) в этом графе. Для оценки вероятности связи используются как геометрические данные (положении объектов в 3D), так и признаки внешнего вида, извлекаемые с помощью сетей ре-идентификации (Re-ID). Применение механизмов перекрестного внимания (cross-attention) позволяет модели эффективно взвешивать и объединять информацию из различных источников, улучшая качество сопоставления по сравнению с традиционными методами.

Методы на основе трансформеров. В последние годы архитектура трансформеров, изначально предложенная для задач обработки естественного языка, нашла широкое применение и в компьютерном зрении. Для задачи отслеживания нескольких объектов (MOT) трансформеры позволяют реализовать сквозной (end-to-end) подход, объединяя детектирование и сопоставление в одной модели. Например, в работе [66] используется архитектура энкодер-декодер на основе трансформера для предсказания треков объектов напрямую из изображений. Такой подход избавляет от необходимости в промежуточных эвристиках и вручную настраиваемых параметрах, таких как non-max suppression (NMS) или алгоритмов сопоставления данных. Другие работы, такие как [67; 68], также исследуют применение векторов признаков (эмбеддингов) и механизмов внимания для улучшения качества отслеживания и сопоставления.

Различия между традиционным подходом и глобальной моделью ассоциаций представлены на рисунке 2.2.

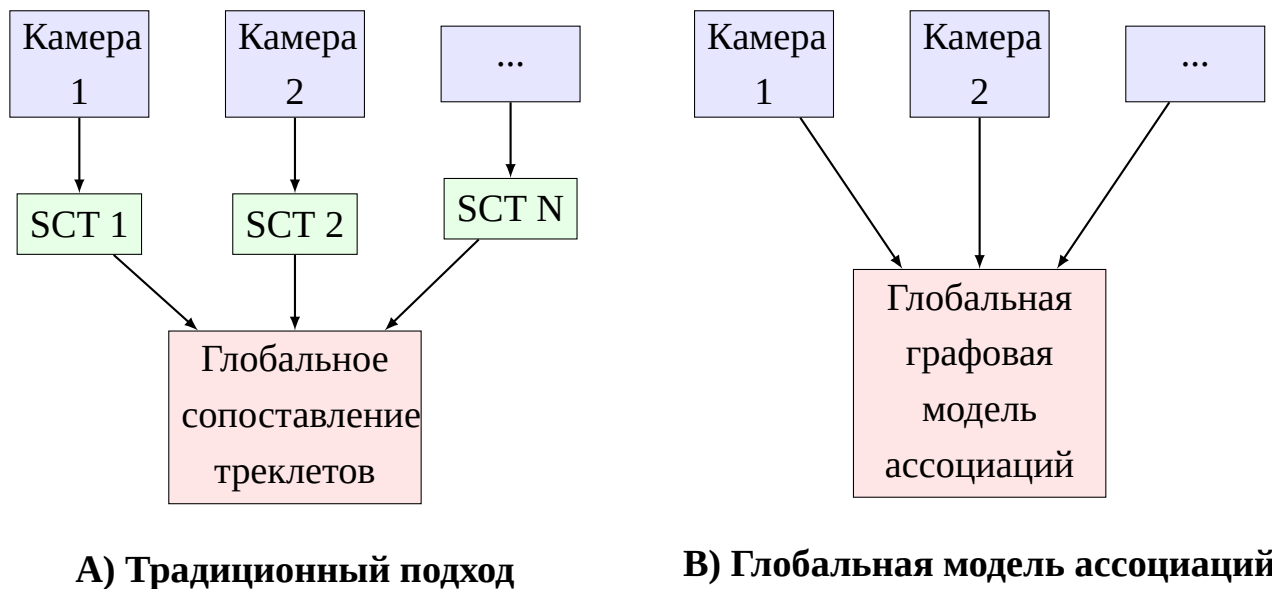


Рисунок 2.2 — Сравнение подходов к межкамерному сопоставлению объектов. А) Традиционный подход с независимым однокамерным трекингом (SCT) и последующим глобальным сопоставлением. В) Современный подход с использованием единой глобальной модели, которая одновременно обрабатывает данные со всех камер

Несмотря на высокий потенциал современных подходов, основанных на глобальных моделях и трансформерах, их применение в реальных СПВ сталкивается со значительными практическими ограничениями. Обучение таких моделей требует обширных и сложных наборов данных, включающих синхронизированные потоки с нескольких камер с точной 3D-разметкой, сбор которых является трудоемким и дорогостоящим процессом. Кроме того, они предъявляют высокие требования к вычислительным ресурсам, которые зачастую превышают возможности бортовых систем. В связи с этим, традиционный двухэтапный подход остается практически целесообразным и распространенным решением в индустрии. Однако, эффективность этого метода критически зависит от качества используемой метрики схожести (или функции стоимости) для сопоставления треклетов. Разработка оптимальной метрики, устойчивой к изменениям ракурса, освещения и перекрытиям, является ключевой задачей при создании надежной системы межкамерного сопоставления.

2.1.3 Метрики схожести для сопоставления треклетов

Эффективность процесса ассоциации детекций между камерами напрямую зависит от выбора метрики схожести или, в более общем смысле, функции стоимости, которая определяет, насколько вероятно, что два треклета (две детекции) принадлежат одному и тому же объекту. Существует множество подходов к определению такой метрики, каждый из которых обладает своими преимуществами и недостатками.

Метрики можно условно разделить на несколько категорий. Наиболее распространенными являются метрики, основанные на визуальной схожести. В этом подходе для каждого обнаруженного объекта извлекается вектор признаков (эмбеддинг), который инкапсулирует его визуальные характеристики, такие как цвет, текстура или форма. Часто для этой цели используются сверточные нейронные сети, обученные на задачах повторной идентификации. Схожесть между двумя треклетами затем вычисляется как расстояние (например, евклидово или косинусное) между их векторами признаков. Современные системы часто агрегируют признаки из нескольких кадров для получения более робастного “якорного” представления объекта [69]. Несмотря на высокую эффективность, данный подход чувствителен к резким изменениям освещения, ракурса камеры и окклюзиям.

Метрики, основанные на пространственно-временных данных, являются другой важной категорией. Эти метрики используют информацию о движении и местоположении объекта. Простейшим вариантом является анализ пересечения треклетов во времени и предсказание траектории на основе фильтра Калмана или других моделей движения. Более сложные методы проецируют координаты объекта из 2D-пространства изображения в 3D-мировые координаты, используя гомографию или данные калибровки камер. Это позволяет вычислять реальное евклидово расстояние между объектами и учитывать физические ограничения на их перемещение между зонами видимости разных камер [70]. Пространственно-

временные метрики хорошо справляются с краткосрочными окклюзиями, но могут давать сбои при непредсказуемых маневрах объекта или при длительном отсутствии в поле зрения камер.

На практике наиболее робастные системы используют гибридный подход, комбинируя различные метрики в единой функции стоимости. Например, система может использовать иерархический подход, где на первом этапе отбрасываются заведомо несовместимые пары треклетов с помощью жестких пространственно-временных ограничений (например, отсутствие временного пересечения для камер с общей зоной видимости), а затем для оставшихся кандидатов вычисляется взвешенная сумма схожести по визуальным признакам и по параметрам движения [69]. Такой подход позволяет сочетать сильные стороны разных методов и нивелировать их недостатки.

Таким образом, не существует универсальной метрики схожести, подходящей для всех сценариев. Выбор конкретной реализации зависит от условий эксплуатации системы (например, плотность объектов, наличие перекрывающихся полей зрения у камер), доступных данных (2D или 3D) и требований к производительности и точности.

2.1.4 Обработка асинхронных измерений

В идеальных условиях все сенсоры в многокамерной системе должны быть строго синхронизированы, чтобы предоставлять измерения, относящиеся к одному и тому же моменту времени. Это позволяет напрямую применять метрики схожести, рассмотренные в предыдущем разделе. Однако в реальных системах помощи водителю достижение идеальной синхронизации часто оказывается невозможным из-за ограничений пропускной способности сети и вычислительных ресурсов бортового компьютера. В результате детекции одного

и того же объекта с разных камер поступают в систему с временными задержками, которые могут достигать десятков или даже сотен миллисекунд.

Временная рассинхронизация (асинхронность) создает фундаментальную проблему для алгоритмов сопоставления. Прямое сравнение пространственных координат или визуальных признаков объектов, зафиксированных в разные моменты времени, может привести к неверным ассоциациям, особенно для динамичных объектов, чье положение и вид могут значительно измениться даже за короткий промежуток времени. Эффективность пространственно-временных метрик, основанных на предсказании траекторий, также снижается, поскольку неопределенность во времени вносит дополнительную ошибку в модель движения.

Для решения этой проблемы существует несколько подходов. Наиболее распространенным является метод компенсации задержек путем предсказания (экстраполяции) состояния объекта. Используя модель движения, такую как фильтр Калмана, система может оценить состояние (положение, скорость) каждого объекта на некий единый, «опорный» момент времени. После этого сопоставление треклетов производится в этом синхронизированном пространстве состояний [71].

Другой, более простой подход заключается в использовании «временного окна» при сопоставлении. Вместо поиска соответствия в тот же самый момент времени, алгоритм ищет кандидатов в определенном интервале, компенсируя возможную задержку. Хотя этот метод вычислительно эффективен, он менее точен, чем подходы, основанные на явном моделировании движения.

В последние годы появляются и более сложные решения, которые пытаются «научиться» обрабатывать асинхронные данные напрямую. Например, современные архитектуры на основе трансформеров могут быть спроектированы так, чтобы учитывать временные метки как один из входных признаков, позволяя сети самостоятельно моделировать и компенсировать временные сдвиги между видами с разных камер [72]. Несмотря на их высокий потенциал, такие методы

требуют значительных вычислительных ресурсов и больших объемов данных для обучения.

Таким образом, обработка асинхронных измерений требует компромисса между точностью, вычислительной сложностью и доступными ресурсами, что делает ее одной из важных практических задач при построении систем восприятия для автономного транспорта.

2.2 Трекинг

2.2.1 Основные этапы трекинга

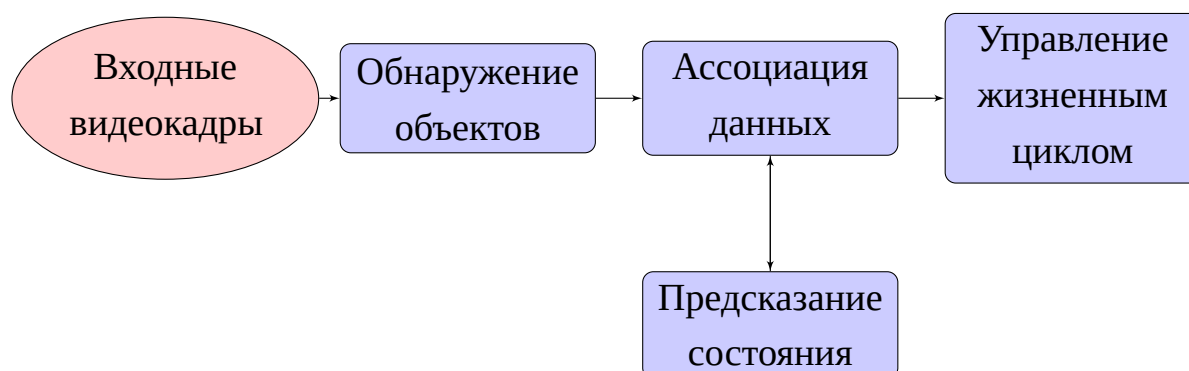


Рисунок 2.3 — Этапы трекинга

Основу алгоритма трекинга составляют три ключевых компонента:

1. **Предсказание состояния:** Оценка будущего местоположения и других параметров объекта на основе его предыдущих состояний.
2. **Ассоциация данных:** Сопоставление предсказанных состояний существующих треков с новыми детекциями на текущем кадре. Это наиболее сложный этап, требующий эффективных метрик для измерения сходства между треками и детекциями. Методы варьируются от простых, как в ByteTrack, где ассоциируется каждая детекция [73], до более

сложных, использующих как пространственную, так и визуальную информацию [74; 75].

3. **Управление жизненным циклом треков:** Механизмы для создания новых треков (когда появляется новый объект), их завершения (когда объект покидает сцену) и обработки пропусков детекций, например, из-за заслонений.

Задача ассоциации данных может решаться путем построения матрицы стоимости, где каждый элемент отражает степень сходства между i -м треком и j -й детекцией. Для оптимального нахождения пар часто применяется Венгерский алгоритм [75] или “жадный” алгоритм. Сходство может оцениваться на основе пространственной близости или визуальных признаков [76].

Далее в этом разделе будет подробно рассмотрен метод трекинга, используемый в рамках данной работы.

2.2.2 Представление состояния и модель движения

Для отслеживания объекта необходимо определить его состояние в каждый момент времени. В задаче трекинга состояние объекта обычно описывается вектором, который включает его кинематические параметры. В данной работе для светофоров используется 4-мерный вектор состояния, который является стандартным для трекинга объектов в задачах автономного вождения:

$$\mathbf{x} = [c_x, c_y, \dot{c}_x, \dot{c}_y]^T, \quad (2.1)$$

где (c_x, c_y) — координаты центра ограничивающей рамки (bounding box), а $(\dot{c}_x, \dot{c}_y)$ — соответствующие скорости изменения этих координат. Такой вектор состояния позволяет моделировать положение и скорость объекта.

Для предсказания состояния объекта на следующем временном шаге используется линейная модель движения с постоянной скоростью и фильтр

Калмана [77]. Архитектура данного фильтра состоит из двух основных этапов: предсказания и обновления.

Этап предсказания. На этом этапе предсказывается состояние объекта $\mathbf{x}_k$ и ковариация ошибки P_k на текущем временном шаге k на основе оценок с предыдущего шага $k - 1$:

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{F}_k \hat{\mathbf{x}}_{k-1|k-1}, \quad (2.2)$$

$$\mathbf{P}_{k|k-1} = \mathbf{F}_k \mathbf{P}_{k-1|k-1} \mathbf{F}_k^T + \mathbf{Q}_k. \quad (2.3)$$

Здесь $\mathbf{F}_k$ — матрица перехода состояний, которая моделирует динамику объекта. Для модели с постоянной скоростью она имеет вид:

$$\mathbf{F}_k = \begin{pmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \quad (2.4)$$

Матрица $\mathbf{Q}_k$ — ковариация шума процесса, которая отражает неопределенность модели движения.

Этап обновления. На этапе обновления предсказанное состояние корректируется с учетом нового измерения $\mathbf{z}_k$. Вектор измерений содержит координаты центра ограничивающей рамки, полученной от детектора:

$$\mathbf{z}_k = [c_x, c_y]^T. \quad (2.5)$$

Матрица измерений $\mathbf{H}_k$ отображает вектор состояния $\mathbf{x}_k$ в пространство измерений $\mathbf{z}_k$:

$$\mathbf{H}_k = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}. \quad (2.6)$$

Процесс обновления включает вычисление коэффициента усиления Калмана $\mathbf{K}_k$ и последующую коррекцию состояния и ковариации ошибки:

$$\mathbf{K}_k = \mathbf{P}_{k|k-1} \mathbf{H}_k^T (\mathbf{H}_k \mathbf{P}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k)^{-1}, \quad (2.7)$$

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k (\mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}), \quad (2.8)$$

$$\mathbf{P}_{k|k} = (\mathbf{I} - \mathbf{K}_k \mathbf{H}_k) \mathbf{P}_{k|k-1}. \quad (2.9)$$

Здесь $\mathbf{R}_k$ — ковариация шума измерений, которая характеризует точность детектора. Матрицы $\mathbf{Q}_k$ и $\mathbf{R}_k$ являются настраиваемыми параметрами фильтра, которые подбираются экспериментально для достижения наилучшего качества отслеживания.

2.2.3 Способ ассоциации данных

Как было отмечено, процесс ассоциации данных является центральным элементом любого алгоритма трекинга по детекциям. В данном разделе подробно рассматривается каскадный подход, который последовательно применяет метрики разной природы для достижения робастного сопоставления треков и детекций.

Ограничения однокомпонентных метрик

Пространственная близость. Метрика IoU, являющаяся основой таких алгоритмов, как SORT [61], оценивает степень перекрытия между предсказанным положением объекта (треклетом) и новой детекцией. Её главное преимущество — вычислительная простота и эффективность при отслеживании объектов с плавной и предсказуемой динамикой. Однако в задаче трекинга светофоров этот подход сталкивается с рядом неразрешимых проблем.

Первая проблема связана с плотным расположением объектов. Светофоры часто располагаются близко друг к другу (например, на одной мачте). При небольшой дистанции до них или при использовании длиннофокусной оптики их проекции на изображении могут значительно перекрываться. В таких случаях IoU становится неоднозначной метрикой, что приводит к частым переключениям идентификаторов.

Вторая проблема заключается в заслонениях. Временное перекрытие отслеживаемого светофора другим объектом (например, грузовиком или автобусом) приводит к разрыву трека. После исчезновения окклюзии восстановить прежний ID только на основе IoU невозможно, так как предсказанное положение может оказаться далеко от реального.

Третья проблема связана со статичностью объектов. Парадоксально, но статичность светофоров усложняет задачу для трекеров, основанных на модели движения. Небольшие ошибки в предсказании положения, вызванные, например, вибрацией камеры, могут быть достаточны, чтобы IoU с реальной детекцией оказалось ниже порога. Современные алгоритмы, такие как OC-SORT [78] и BoT-SORT [79], пытаются решить эти проблемы путем введения дополнительных эвристик, связанных с моделью движения и компенсацией движения камеры, однако они не устраняют фундаментальный недостаток IoU — отсутствие информации об идентичности объекта.

Визуальная схожесть. Использование векторов признаков для повторной идентификации, как в DeepSORT [63], позволяет связать объект с его визуальным представлением. Этот подход более устойчив к временным окклюзиям, так как внешний вид объекта сохраняется. Тем не менее, он также имеет свои слабые стороны.

Во-первых, большинство светофоров имеют стандартный, практически идентичный дизайн. Это делает эмбединги, полученные из нейронной сети, очень похожими друг на друга, что затрудняет различение отдельных объектов в группе.

Во-вторых, смена сигнала значительно меняет визуальное представление одного и того же светофора, что приводит к ”дрейфу” эмбединга и, как следствие, к потере трека.

Алгоритм каскадного сопоставления

Для преодоления описанных ограничений применяется каскадная стратегия сопоставления, которая обрабатывает наиболее надёжную информацию в первую очередь. Процесс разделяется на несколько последовательных этапов, на каждом из которых решается задача о назначениях для подмножества треков и детекций.

Задача о назначениях и Венгерский алгоритм. Пусть есть N треков и M детекций. Необходимо найти такое сопоставление между ними, чтобы суммарная стоимость всех сопоставлений была минимальной. Для этого строится матрица стоимости C размера $N \times M$, где элемент C_{ij} представляет собой стоимость сопоставления i -го трека с j -й детекцией.

Задача формулируется следующим образом: найти бинарную матрицу X размера $N \times M$, минимизирующую целевую функцию:

$$\sum_{i=1}^N \sum_{j=1}^M C_{ij} X_{ij} \quad (2.10)$$

при ограничениях:

$$\sum_{j=1}^M X_{ij} \leq 1 \quad \forall i, \quad \sum_{i=1}^N X_{ij} \leq 1 \quad \forall j, \quad X_{ij} \in \{0, 1\}. \quad (2.11)$$

Эта задача может быть эффективно решена за полиномиальное время с помощью Венгерского алгоритма [62].

Этап 1: Первичное сопоставление по эмбедингам. На первом этапе производится сопоставление активных треков с новыми детекциями на основе визуальной схожести. Эта метрика является наиболее робастной к окклюзиям и ошибкам предсказания движения. Матрица стоимости C^{reid} формируется на

основе косинусного расстояния:

$$d_{cos}(T_i, D_j) = 1 - \frac{\bar{E}_i E_j}{|\bar{E}_i| |E_j|} \quad (2.12)$$

где $\bar{E}_i$ – усреднённый эмбединг треклета, обновляемый по формуле:

$$\bar{E}_i^{(t)} = \beta \cdot \bar{E}_i^{(t-1)} + (1 - \beta) \cdot E_j^{(t)}, \quad \beta = 0.9. \quad (2.13)$$

Сопоставляются только пары, у которых стоимость $C_{ij}^{reid} < \tau_1$, где τ_1 — пороговое значение.

Этап 2: Вторичная проверка по геометрическому перекрытию. Треки и детекции, оставшиеся несопоставленными после первого этапа, проходят вторую стадию сопоставления. На этот раз используется метрика IoU, которая эффективна для краткосрочного трекинга при отсутствии окклюзий:

$$IoU(B_i, B_j) = \frac{B_i \cap B_j}{B_i \cup B_j}, \quad (2.14)$$

где B_i, B_j – ограничивающие рамки. Сопоставление происходит при $IoU \geq \tau_2$.

Этап 3: Инициализация новых треков. Детекции, оставшиеся без пары после этапов 1-2, иницииируют новые треклеты с начальным эмбедингом $\bar{E}_i(0) = E_j$.

Такой каскадный подход, сочетающий различные метрики, позволяет создать гибкую и эффективную систему трекинга, способную адаптироваться к сложным и динамичным сценариям дорожного движения.

2.3 Объединение результатов детекций между камерами

2.3.1 Анализ существующих подходов

Для надежности решения задачи распознавания светофоров необходимо одновременно распознавать как основные, так и дублирующие светофоры.

Кроме того, необходимо распознавать сигналы светофора заблаговременно, чтобы при необходимости обеспечить своевременное торможение транспортного средства. С этой целью в проекте используется две камеры, с разными фокусными расстояниями, отличающимися примерно в два раза. Одна из камер настроена на работу с дальними объектами, вторая с ближними. Изображения с каждой камеры независимо друг от друга обрабатываются нейронной сетью, поэтому результаты распознавания объектов и светофоров полученные с двух камер необходимо объединять, т.е. найти соответствие между ними. Поскольку под трекингом чаще всего понимается процедура межкадрового слежения, а не межкамерного, чтобы избежать путаницы, используется термин «объединение».

В последние годы был достигнут значительный прогресс в области MOT, что подробно рассматривается в обзорах [56; 60]. Однако стандартные методы MOT не всегда напрямую применимы к задаче объединения светофоров между разнофокусными камерами. Основные сложности включают:

- **Схожесть внешнего вида.** Светофоры одного типа практически неотличимы друг от друга, что делает методы, основанные на визуальных признаках, менее надежными.
- **Изменение состояния.** Постоянное переключение сигналов (красный, желтый, зеленый) изменяет вектор признаков объекта, что может привести к потере трека.
- **Разные типы светофоров.** Наличие пешеходных, велосипедных и транспортных светофоров в одной сцене требует их точной классификации перед сопоставлением.
- **Непересекающиеся поля зрения.** В некоторых конфигурациях поля зрения камер могут не пересекаться или пересекаться лишь частично, что требует более сложных методов для ассоциации объектов [80].

Для решения этих проблем разрабатываются специализированные подходы. Например, в работе [51] для сопоставления светофоров между камерами используется высокоточная карта (HD-карта) в качестве общего источника данных. Вместо прямого объединения детекций между камерами, известные

3D-координаты светофоров с карты проецируются в 2D-пространство каждого изображения. Затем, для каждой камеры отдельно, выполняется сопоставление реальных детекций и спроецированных боксов с помощью Венгерского алгоритма. Детекции с разных камер считаются принадлежащими одному и тому же объекту, если они ассоциируются с одним и тем же объектом в HD-карте. Тем не менее, такой подход имеет существенные ограничения. Проецирование 3D-координат из карты на 2D-изображение и последующая попытка ассоциировать детекции с объектами карты не всегда являются достаточными. В реальных условиях на одном столбе часто располагается несколько светофоров, также они могут быть разных типов (например, для разных полос движения или транспортные и пешеходные). При проецировании на изображение их боксы могут оказаться очень близко друг к другу, что создает неоднозначность процесса сопоставления.

Современные подходы к объединению объектов между камерами часто опираются на сложные архитектуры нейронных сетей, которые, несмотря на высокую точность, предъявляют значительные требования к вычислительным ресурсам. Рассмотрим несколько таких подходов и их применимость в контексте СПВ.

Один из передовых методов использует графовые трансформерные сети для построения глобальной модели отслеживаемых объектов со всех камер одновременно [65]. В этой модели каждый объект представлен как узел в графе, содержащий информацию о его местоположении и визуальные признаки. Трансформерная архитектура с механизмами внимания используется для обновления состояния объектов и предсказания их движения. Такой подход позволяет избежать каскадных ошибок, возникающих при независимом отслеживании на каждой камере с последующим объединением. Однако его главным недостатком является высокая вычислительная сложность. Трансформеры, особенно с перекрестным вниманием, требуют значительных ресурсов, что делает их малоприспособленными для встраиваемых СПВ с ограниченной

производительностью. Кроме того, для обучения таких моделей требуются обширные и тщательно размеченные наборы данных, такие как nuScenes [81].

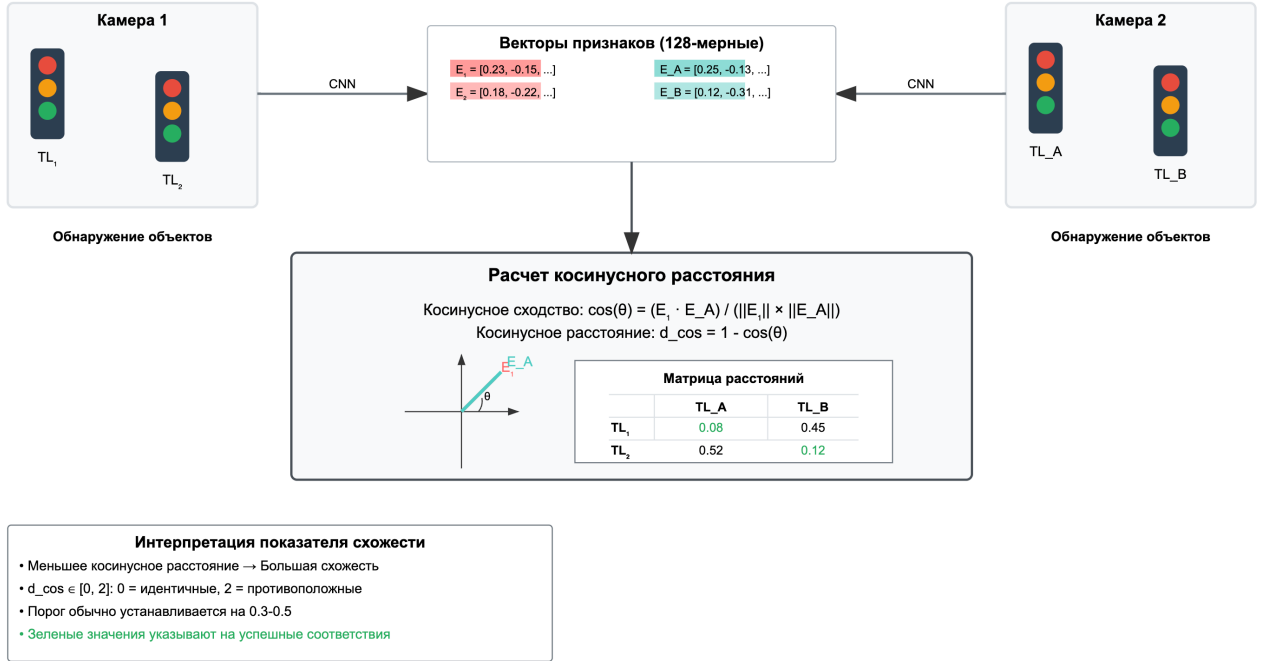
Другой популярный подход заключается в объединении обнаруженных 3D-объектов в общем пространстве (например, на проекции «вид сверху» с англ. Bird's-Eye-View, BEV) еще до этапа сопоставления [82]. Детекции со всех камер преобразуются в единую систему координат и объединяются в более полное представление объекта. После этого для временной ассоциации используется стандартный алгоритм трекинга. Эта стратегия позволяет эффективно работать с частично видимыми объектами. Критическим аспектом здесь является точность калибровки камер. Любые ошибки в матрицах внутренних или внешних параметров приводят к неверному проецированию и, как следствие, к ошибкам слияния. Генерация *точных* 3D-детекций по монокулярным изображениям сама по себе является ресурсоемкой задачей, а необходимость в идеальной калибровке, которая может нарушаться в процессе эксплуатации транспортного средства, снижает надежность метода для бюджетных систем.

2.3.2 Предлагаемый подход

Учитывая ограничения описанных выше методов, необходимо рассмотреть более простые и вычислительно эффективные подходы к объединению детекций светофоров. В данной работе исследуется возможность использования базовых метрик схожести, которые могут быть реализованы в системах реального времени без существенных вычислительных затрат. Задача объединения данных не является тривиальной и может быть решена с использованием различных функций расстояния, которые анализируют схожесть объектов в разных доменах: на изображении (2D) и в трехмерной сцене (3D).

Косинусное расстояние между эмбедингами (2.12). Данная техника широко распространена на практике и не зависит от параметров настроек

камеры. Однако эмбединги могут совпасть у похожих объектов, которые при этом находятся в разных частях изображения и эмбединг может меняться от кадра к кадру (напр. из-за мигания светофора или смены его сигнала).



Меньшее косинусное расстояние указывает на большую вероятность соответствия объекта между кадрами

Рисунок 2.4 — Косинусное расстояние между эмбедингами

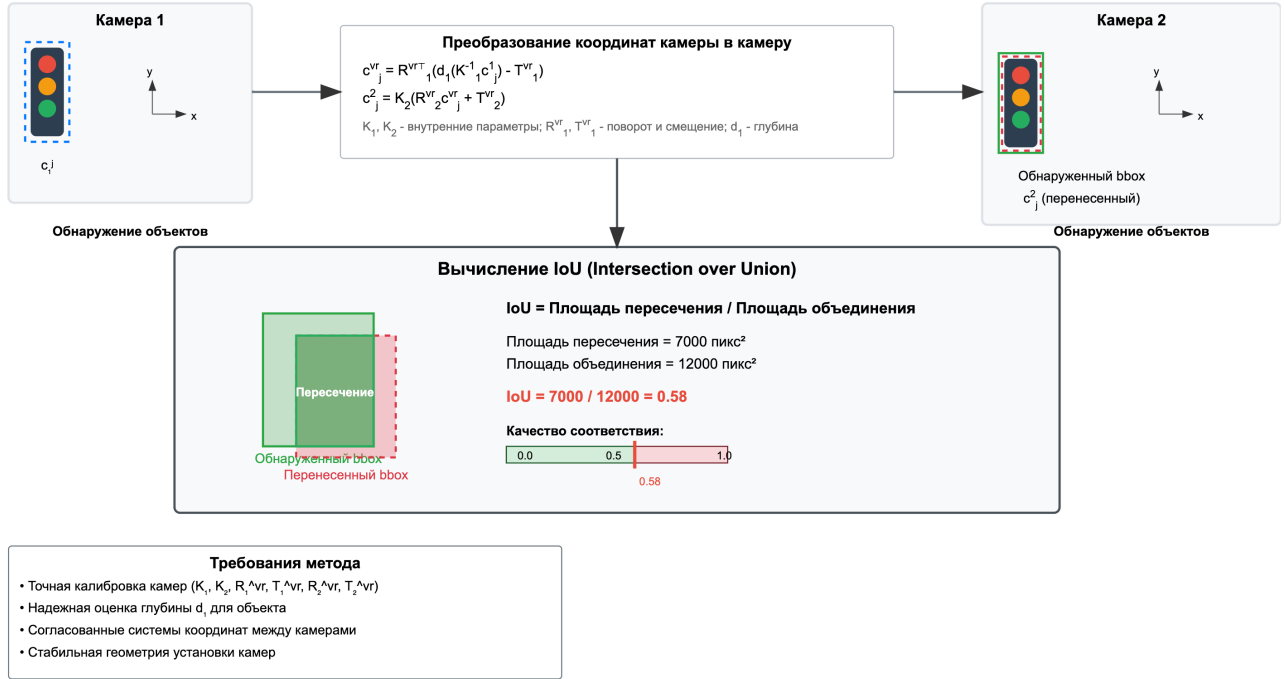
Мера пересечения двух боксов путем переноса результатов распознавания с одного изображения на другое. Для переноса центра bbox с изображения первой камеры c_j^1 на изображение второй камеры c_j^2 используется цепочка преобразований:

$$c_j^{vr} = R_1^{vr\top} (d_1(K_1^{-1}c_j^1) - T_1^{vr}) \quad (2.15)$$

$$c_j^2 = K_2(R_2^{vr}c_j^{vr} + T_2^{vr}) \quad (2.16)$$

где K_i – внутренние параметры камеры i , R_i^{vr} и T_i^{vr} – матрицы поворота и смещения камеры i относительно системы координат транспортного

средства, d_1 – глубина центра bbox в первой камере. После переноса центров и масштабирования размеров bbox вычисляется стандартный IoU через уравнение (2.14) между перенесенным и детектированным bbox. Данный подход требует точной калибровки камер и надежной оценки глубины.



Высокий IoU указывает на успешное геометрическое соответствие между камерами

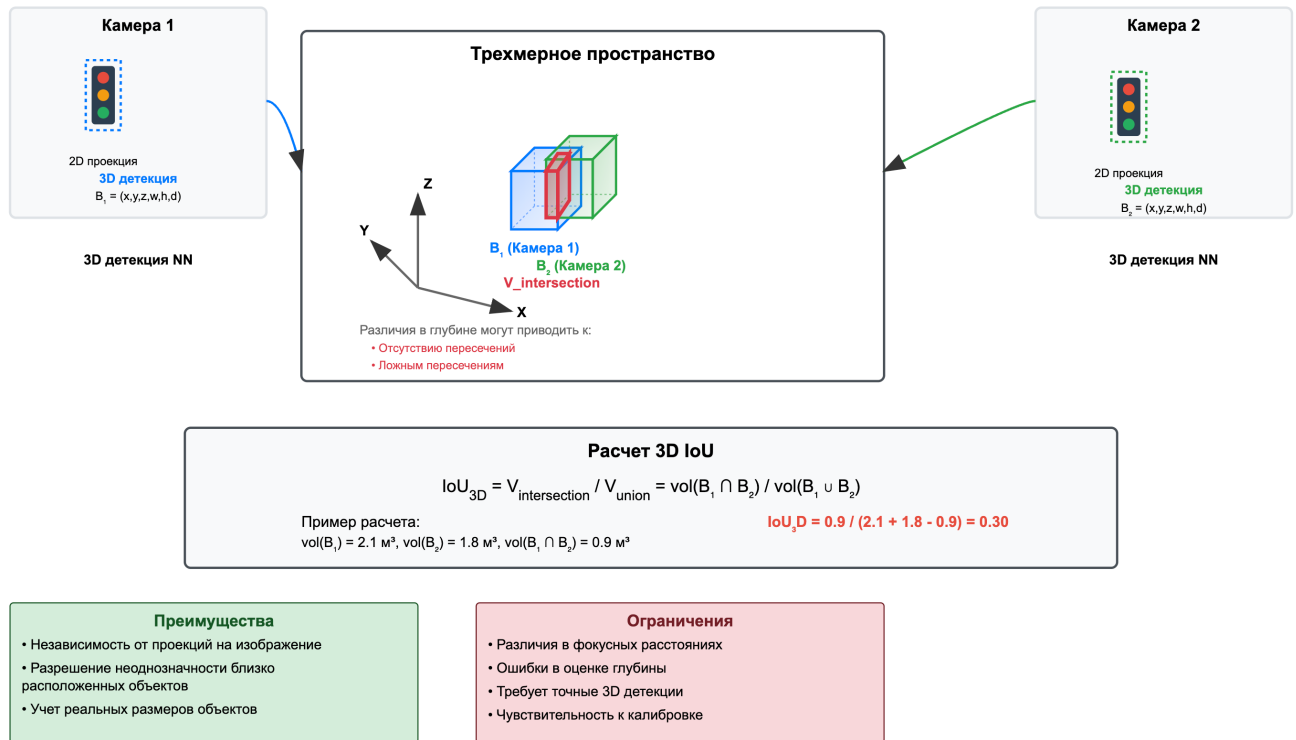
Рисунок 2.5 — Мера пересечения двух боксов

Расчет IoU в трехмерном пространстве. Поскольку нейронная сеть производит 3D детекции, эти данные можно применять для определения пересечений объектов независимо от их проекций на изображение согласно:

$$IoU_{3D} = \frac{V_{intersection}}{V_{union}} = \frac{\text{vol}(B_1 \cap B_2)}{\text{vol}(B_1 \cup B_2)} \quad (2.17)$$

где B_1 и B_2 – кубоиды (3D боксы) с первой и второй камеры соответственно, vol – объем. Такой подход помогает разрешить неоднозначность в ситуациях, когда объекты находятся близко друг к другу в кадре. Тем не менее, существенная разница в фокусных расстояниях камер может приводить к значительным

расхождению в оценке глубины объектов, что либо исключает пересечение кубоидов в пространстве, либо создает ложные пересечения.



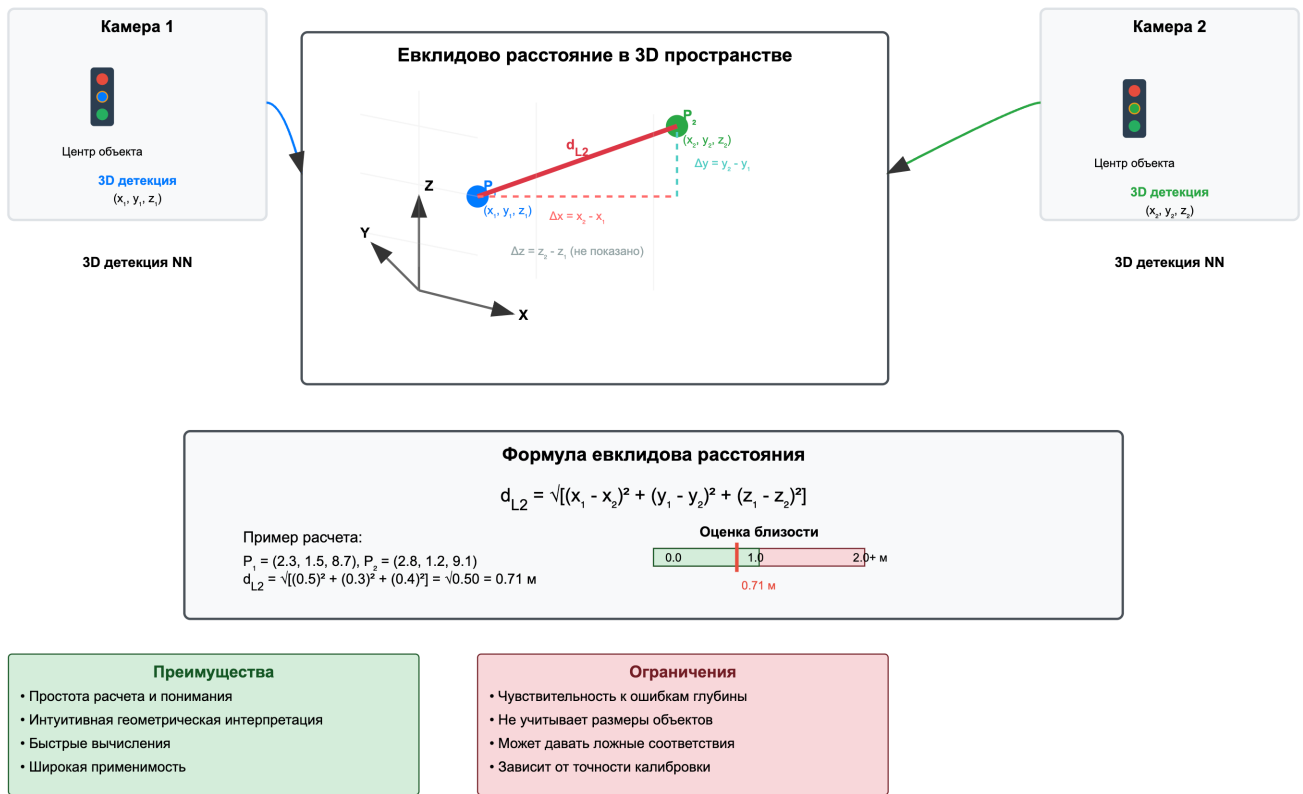
3D IoU обеспечивает геометрически корректное сравнение объектов в мировых координатах

Рисунок 2.6 — Мера пересечения двух кубоидов

Евклидово расстояние (L2) между центрами объектов (в пространстве) рассчитывается как:

$$d_{L2} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \quad (2.18)$$

где (x_1, y_1, z_1) и (x_2, y_2, z_2) – координаты центров объектов с первой и второй камеры соответственно. Понятный и популярный подход, однако из-за ошибок в предсказанной глубине объекты также могут не совпадать.



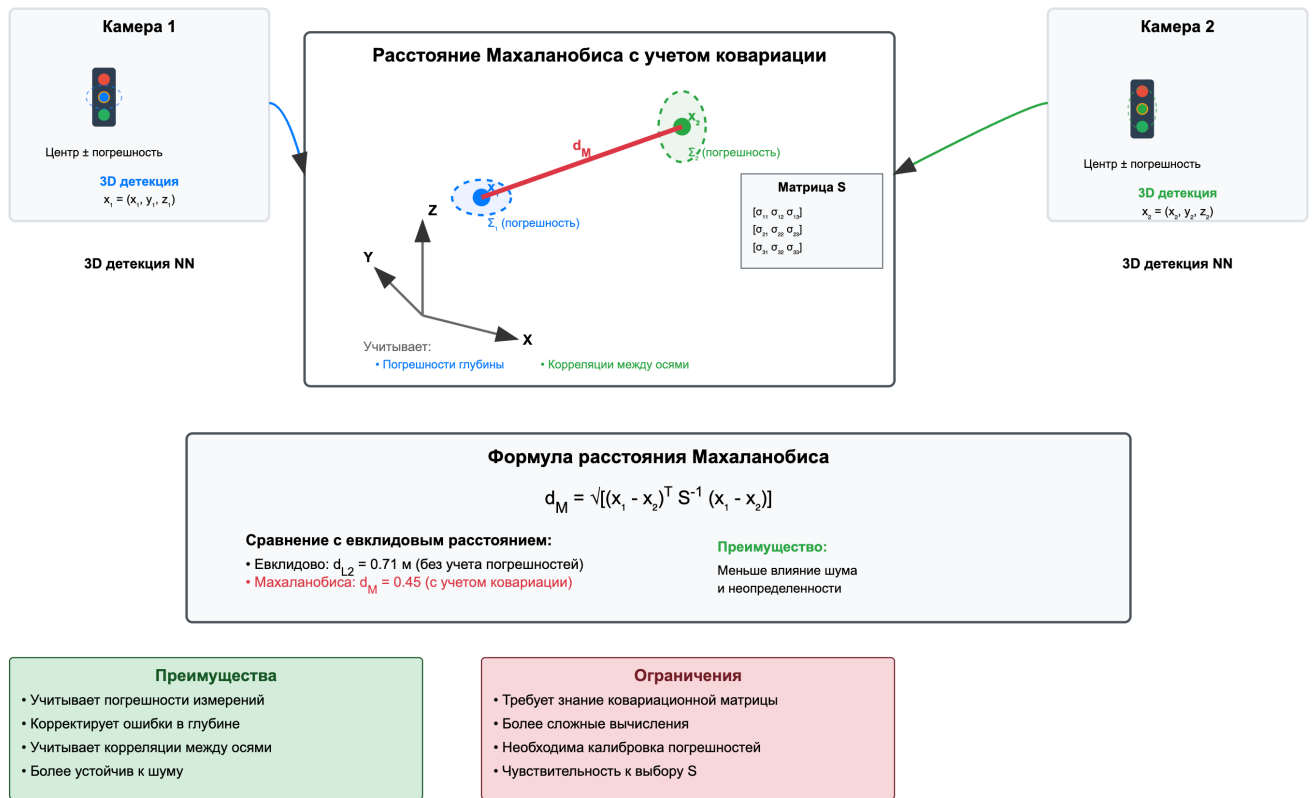
Евклидово расстояние обеспечивает простую меру близости центров объектов в 3D пространстве

Рисунок 2.7 — Евклидово расстояние между центрами объектов

Расстояние Махаланобиса, близкое по своей сути к Евклидову расстоянию, но с возможностью корректировки ошибок в определении расстояния до объектов, полученных каждой камерой. Этот метод также учитывает погрешности (шум) в измерениях глубины, получаемых от нейронной сети:

$$d_M = \sqrt{(x_1 - x_2)^T S^{-1} (x_1 - x_2)} \quad (2.19)$$

где x_1 и x_2 — векторы координат центров объектов с первой и второй камеры соответственно, S — ковариационная матрица, учитывающая погрешности измерений.



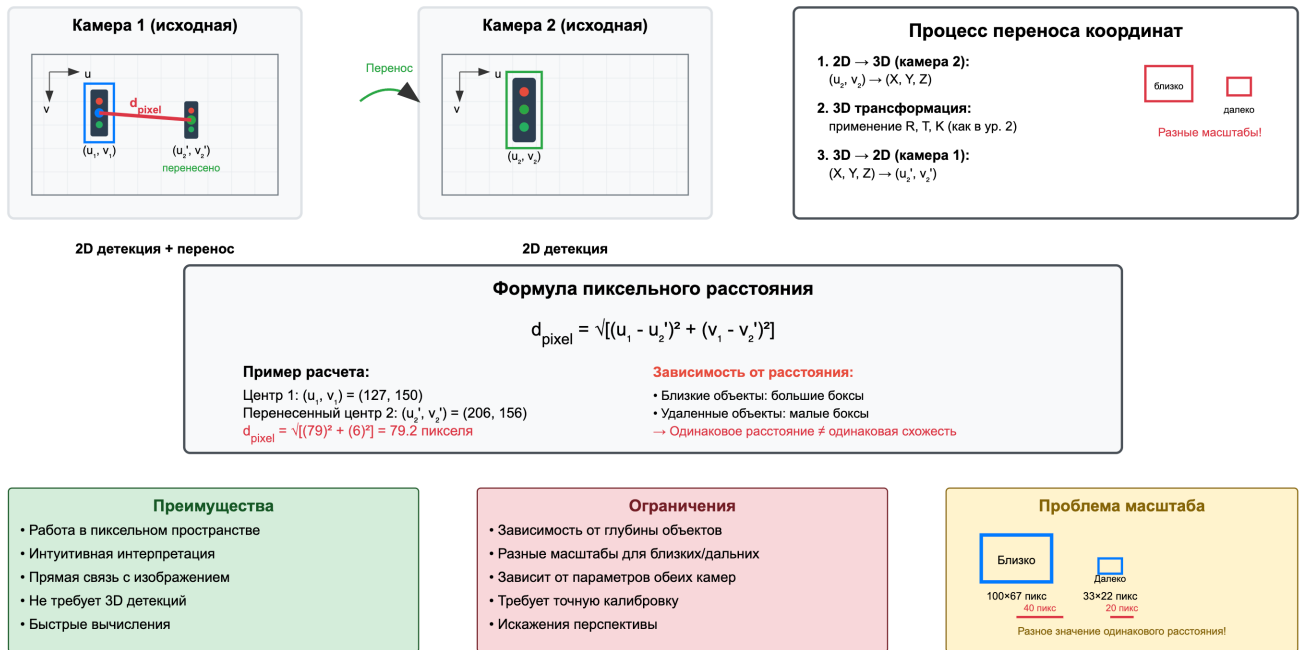
Расстояние Махаланобиса учитывает неопределенности и корреляции в измерениях глубины

Рисунок 2.8 — Расстояние Махаланобиса

Расстояние между центрами 2D детекций путем переноса координат центров выделенных областей на одно изображение и подсчета расстояния между ними согласно:

$$d_{pixel} = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2} \quad (2.20)$$

где (u_1, v_1) – координаты центра детекции на первой камере, (u_2, v_2) – координаты центра после переноса со второй камеры на первую согласно уравнению (2.15). В ряде случаев сложно учитывать разницу в пикселях для удаленных объектов, поскольку размеры боксов сильно изменяются на одном изображении относительно второго. Кроме того метрика зависит от параметров обеих камер.



Пиксельное расстояние зависит от глубины объектов и параметров камер

Рисунок 2.9 — Расстояние между центрами 2D детекций

Как видно из анализа, способов “объединения” измерений достаточно много. Для выбора оптимальной стратегии совмещения измерений или их комбинаций необходимо исходить из следующих факторов: метрики должны выдавать нормализованные значения в диапазоне от 0 до 1, а при использовании нескольких метрик необходимо определить коэффициенты их влияния.

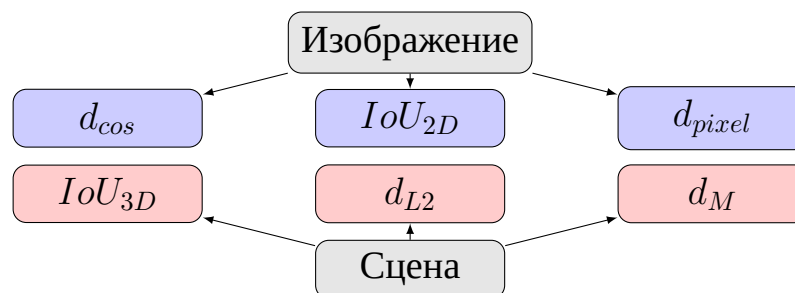


Рисунок 2.10 — Классификация метрик по доменам: изображение (2D) и сцена (3D)

Следует отметить, что метрики, основанные на трехмерной сцене (см. рисунок 2.10), можно считать производными от данных изображений, поскольку 3D-координаты получаются путем преобразования 2D-данных. Поэтому для построения итоговой метрики было решено использовать только оригинальные, работающие непосредственно с изображениями. На основании проведенных экспериментов была создана итоговая метрика, которая объединяет данные о взаимном расположении центров 2D-объектов, эмбединги и IoU 2D-объектов. С последними двумя компонентами вводятся веса, позволяющие адаптировать их вклад в итоговую метрику. Итоговая метрика соответствия M вычисляется следующим образом:

$$M = \begin{cases} w_{emb} \cdot d_{cos} + w_{iou} \cdot (1 - \text{IoU}), & \text{если } d_{pixel} < d_{thresh} \\ \infty, & \text{иначе} \end{cases} \quad (2.21)$$

где d_{cos} – косинусное расстояние между эмбедингами, IoU – мера пересечения 2D-боксов, w_{emb} и w_{iou} – весовые коэффициенты ($w_{emb} + w_{iou} = 1$), d_{pixel} – расстояние между центрами 2D-боксов в пикселях, d_{thresh} – пороговое значение. Чем меньше значение M , тем “ближе” объекты.

Для проведения экспериментов и подбора оптимальных весовых коэффициентов специально подготовлен набор данных. Он состоит из большого количества пар изображений, полученных с двух камер. Набор данных охватывает широкий спектр реальных дорожных ситуаций. В него включены изображения снятые в различных погодных условиях и в разное время суток (день и ночь). Кроме того, в датасете представлены различные виды светофоров, что обеспечивает его разнообразие и репрезентативность.

Ключевой особенностью набора данных является наличие детальной ручной разметки. Для каждого светофора на изображениях не только указана его ограничивающая рамка, но и установлено соответствие между его детекциями на обеих камерах. На рисунке 2.11 представлены примеры изображений из набора

данных.



Рисунок 2.11 — Пример изображений из набора данных для сопоставления светофоров.

Задача оптимизации весов формулировалась как задача минимизации функции ошибки E :

$$E(w_{emb}, w_{iou}) = \sum_{i=1}^N \sum_{j=1}^K L(y_{ij}, \hat{y}_{ij}(w_{emb}, w_{iou})) \quad (2.22)$$

где:

- N – количество кадров в наборе данных;
- K – количество пар детекций в кадре;
- y_{ij} – эталонное соответствие (1 – детекции соответствуют одному светофору, 0 – разным);
- $\hat{y}_{ij}$ – предсказанное соответствие на основе метрики M ;
- L – функция потерь в виде бинарной кросс-энтропии

$$L(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}) \quad (2.23)$$

При этом накладывались ограничения

$$\begin{cases} w_{emb} + w_{iou} = 1 \\ w_{emb}, w_{iou} \geq 0 \end{cases} \quad (2.24)$$

В результате многочисленных экспериментов было установлено, что веса находятся в диапазоне: $w_{emb} \in [0.6, 0.7]$ для эмбеддингов и $w_{iou} \in [0.3, 0.4]$ для IoU. На рисунке 2.12 представлены результаты оптимизации весовых коэффициентов.

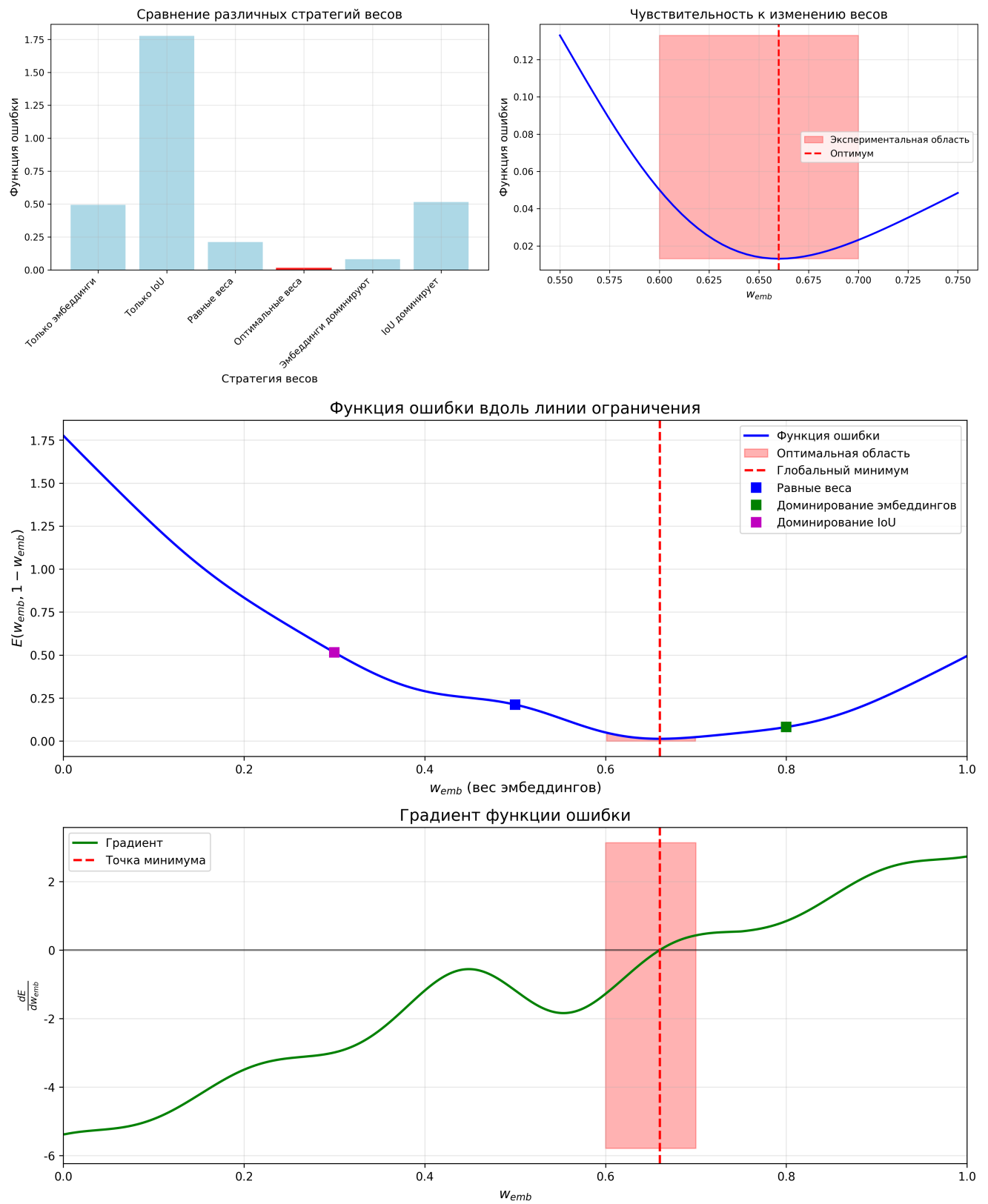


Рисунок 2.12 — Анализ процесса оптимизации весовых коэффициентов

Итоговое сопоставление светофоров между камерами осуществляется с помощью венгерского алгоритма (2.10), (2.11).

Полученные экспериментальные значения хорошо коррелируют с практическими данными, когда наблюдение одного и того же светофора с разных камер в силу их асинхронности и смены ракурса может приводить к изменениям эмбедингов детектируемых областей.

Выводы

В данной главе была решена задача сопоставления детекций светофоров в двухкамерной разнофокусной системе. Проанализированы существующие подходы и предложена собственная методика, адаптированная для применения в СПВ. Основные результаты и выводы, сделанные в главе, можно сформулировать следующим образом:

- Проведен детальный анализ существующих методов отслеживания объектов и межкамерной ассоциации, выявлены их сильные и слабые стороны применительно к задаче распознавания светофоров. Было показано, что стандартные подходы не всегда эффективны в условиях высокой визуальной схожести объектов и динамического изменения их состояния.
- Выделены ключевые практические проблемы: высокая плотность и схожесть светофоров, частые окклюзии, изменение внешнего вида из-за смены сигнала, а также строгие вычислительные ограничения бортовых СПВ, не позволяющие применять ресурсоемкие алгоритмы на основе трансформеров или слияния данных в BEV-пространстве.
- Для однокамерного отслеживания предложена каскадная стратегия сопоставления, которая повышает надежность трекинга. Она отдает приоритет сопоставлению по визуальным признакам для

повторной идентификации объектов после окклюзий и использует пространственное перекрытие для краткосрочного отслеживания.

- Продемонстрирована нецелесообразность применения сложных подходов межкамерного объединения детекций, основанных на HD-картах или сложных нейросетевых архитектурах, для СПВ из-за их высокой вычислительной стоимости и чувствительности к точности калибровки и качеству данных.
- Разработана и предложена новая гибридная метрика для сопоставления детекций между разнофокусными камерами. Метрика объединяет информацию о визуальном сходстве (косинусное расстояние), пространственном перекрытии после репроекции (IoU 2D) и геометрической близости в пространстве изображения.
- Путем экспериментов на размеченном наборе данных были определены оптимальные весовые коэффициенты для компонент гибридной метрики. Установлено, что для надежного сопоставления визуальные признаки ($w_{emb} \in [0.6, 0.7]$) являются более значимыми, чем информация о пространственном перекрытии ($w_{iou} \in [0.3, 0.4]$).
- Итоговое сопоставление объектов как в рамках однокамерного трекинга, так и при межкамерном объединении, решается как задача о назначениях с помощью Венгерского алгоритма, что гарантирует нахождение оптимального соответствия на каждом шаге.
- Предложенный в главе подход представляет собой прагматичное и вычислительно эффективное решение, сочетающее классические алгоритмы с тщательно подобранными метриками, что делает его хорошо подходящим для реального применения во встраиваемых СПВ.

Глава 3. Методика фильтрации пространственных координат светофоров

3.1 Обзор методов фильтрации и оценки координат

3.1.1 Методы оптимальной фильтрации

В СПВ и АТС оценка точного трехмерного положения объектов в пространстве является фундаментальной задачей. Данные, получаемые от сенсоров, таких как камеры и лидары, неизбежно содержат шум. Для получения стабильной и надежной оценки координат светофоров по последовательности зашумленных измерений необходимо применять методы оптимальной фильтрации. Эти методы позволяют не только сгладить шум, но и предсказать состояние объекта между измерениями, что особенно важно когда детекции отсутствуют на протяжении нескольких кадров из-за окклюзий или других помех.

Основой для большинства современных алгоритмов фильтрации является Байесовская фильтрация [83]. В рамках этого подхода задача оценки состояния сводится к вычислению апостериорной плотности вероятности вектора состояния на основе всех доступных измерений. Рекурсивное вычисление этой плотности описывается двумя шагами: предсказанием и коррекцией.

На шаге предсказания априорная плотность вероятности состояния на момент времени k вычисляется на основе модели динамики системы и апостериорной плотности с предыдущего шага:

$$p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k | \mathbf{x}_{k-1}) p(\mathbf{x}_{k-1} | \mathbf{z}_{1:k-1}) d\mathbf{x}_{k-1} \quad (3.1)$$

где $\mathbf{x}_k$ — вектор состояния в момент времени k , $\mathbf{z}_{1:k-1}$ — совокупность всех измерений до момента $k - 1$, $p(\mathbf{x}_k | \mathbf{x}_{k-1})$ — модель перехода состояния.

На шаге коррекции априорная плотность уточняется с использованием текущего измерения $\mathbf{z}_k$ по правилу Байеса

$$p(\mathbf{x}_k | \mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{z}_{1:k-1})}{p(\mathbf{z}_k | \mathbf{z}_{1:k-1})} \quad (3.2)$$

где $p(\mathbf{z}_k | \mathbf{x}_k)$ — функция правдоподобия измерения, а знаменатель представляет собой нормализующую константу

$$p(\mathbf{z}_k | \mathbf{z}_{1:k-1}) = \int p(\mathbf{z}_k | \mathbf{x}_k) p(\mathbf{x}_k | \mathbf{z}_{1:k-1}) d\mathbf{x}_k \quad (3.3)$$

Результатом является апостериорная плотность вероятности $p(\mathbf{x}_k | \mathbf{z}_{1:k})$, содержащая всю доступную информацию о состоянии системы на момент времени k .

Для линейных динамических систем с Гауссовым шумом оптимальным решением этой задачи является фильтр Калмана (Kalman Filter, KF) [57]. Он обеспечивает минимальную среднеквадратичную ошибку оценки и широко используется в различных инженерных приложениях, включая навигацию и трекинг. Как отмечалось в 2.2 при анализе методов трекинга, фильтр Калмана служит базовым инструментом для предсказания положения объектов между последовательными кадрами.

Расширенный фильтр Калмана. Несмотря на то, что классический фильтр Калмана широко применяется в инженерных задачах, в реальных условиях восприятия и управления автономных транспортных средств модели, описывающие динамику и измерения, зачастую оказываются существенно нелинейными. Для работы с нелинейными системами был разработан ряд модификаций фильтра Калмана. Наиболее известным является расширенный фильтр Калмана (Extended Kalman Filter, EKF) [84]. Идея EKF заключается в линеаризации нелинейной модели в окрестности текущей оценки состояния с помощью разложения в ряд Тейлора первого порядка.

Пусть изменение вектора состояния $\mathbf{x}_k$ и модель измерения $\mathbf{z}_k$ описываются нелинейными функциями

$$\mathbf{x}_k = f(\mathbf{x}_{k-1}, \mathbf{u}_{k-1}) + \mathbf{w}_{k-1}, \quad \mathbf{z}_k = h(\mathbf{x}_k) + \mathbf{v}_k, \quad (3.4)$$

где $\mathbf{w}_{k-1} \sim \mathcal{N}(\mathbf{0}, \mathbf{Q}_{k-1})$ — шум процесса, а $\mathbf{v}_k \sim \mathcal{N}(\mathbf{0}, \mathbf{R}_k)$ — шум измерения.

ЕКФ линеаризует f и h в окрестности текущей оценки $\hat{\mathbf{x}}_{k-1|k-1}$ при помощи первых производных (Якобианов):

$$\mathbf{F}_{k-1} = \left. \frac{\partial f}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_{k-1|k-1}}, \quad \mathbf{H}_k = \left. \frac{\partial h}{\partial \mathbf{x}} \right|_{\hat{\mathbf{x}}_{k|k-1}} \quad (3.5)$$

Тогда этапы предсказания и коррекции выполняются по формулам классического фильтра Калмана (см. уравнения 2.2, 2.3, 2.7, 2.8, 2.9).

Линеаризация обеспечивает только аппроксимацию первого порядка и требует вычисления Якобианов, аналитически или численно, что затрудняет использование ЕКФ при работе с сильно нелинейными моделями.

Сигма-точечный фильтр Калмана. Альтернативой ЕКФ, позволяющей избежать проблем с линеаризацией, является сигма-точечный фильтр Калмана (Unscented Kalman Filter, UKF) [85; 86]. Вместо аналитической линеаризации, UKF использует так называемое сигма-точечное преобразование (unscented transform) для аппроксимации апостериорного распределения вероятностей набора точек (сигма-точек) χ_i на множестве $2n + 1$, которые детерминированно выбираются «вокруг» текущей оценки (см. рис. 3.1). Эти точки пропускаются через нелинейную функцию, после чего вычисляется новое математическое ожидание и ковариация.

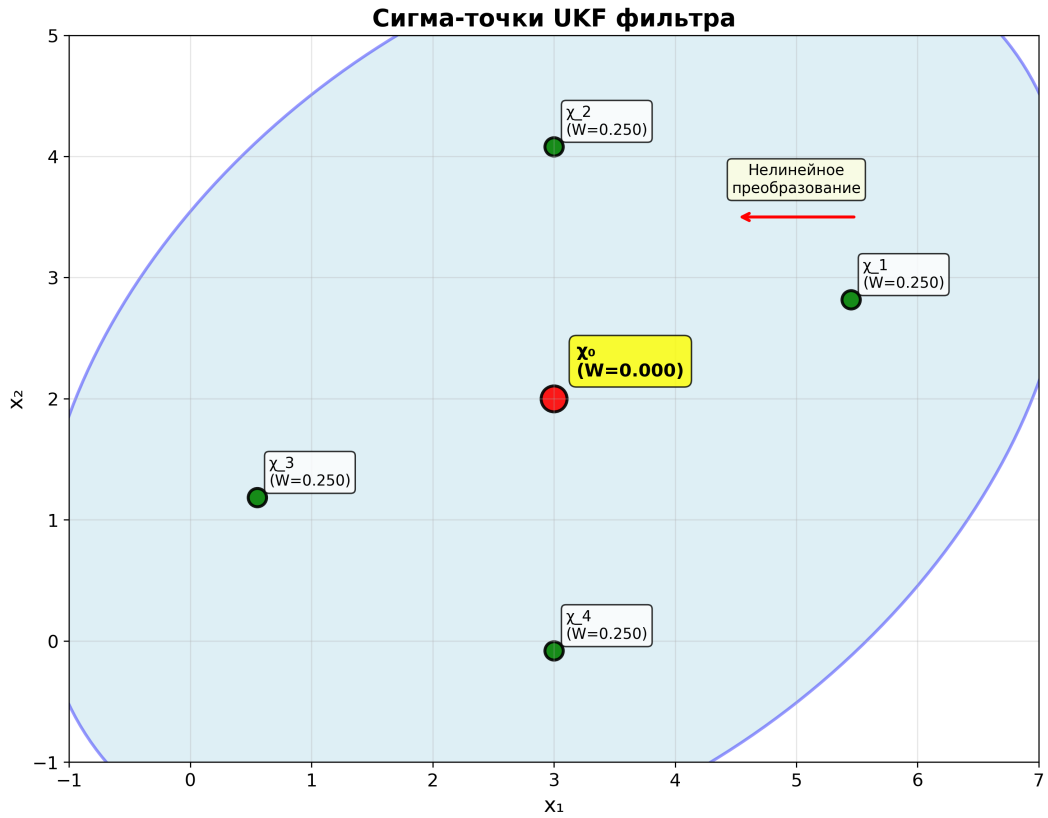


Рисунок 3.1 — Сигма-точки UKF

Пусть $n = \dim \mathbf{x}$,

$$\chi_{k-1}^{(0)} = \hat{\mathbf{x}}_{k-1|k-1}, \quad (3.6)$$

$$\chi_{k-1}^{(i)} = \hat{\mathbf{x}}_{k-1|k-1} + \left[\sqrt{(n + \lambda) \mathbf{P}_{k-1|k-1}} \right]_i, \quad i = 1, \dots, n, \quad (3.7)$$

$$\chi_{k-1}^{(i+n)} = \hat{\mathbf{x}}_{k-1|k-1} - \left[\sqrt{(n + \lambda) \mathbf{P}_{k-1|k-1}} \right]_i, \quad (3.8)$$

где $\lambda = \alpha^2(n + \kappa) - \alpha$ — параметр, определяющий «разброс» сигма-точек вокруг оценки (обычно выбирается малым, например, 10^{-3}), а κ — дополнительный параметр, влияющий на распределение сигма-точек (часто $\kappa = 0$ или $\kappa = 3 - n$). После нелинейного преобразования получаем $\chi_{k|k-1}^{(i)} = f(\chi_{k-1}^{(i)}, \mathbf{u}_{k-1})$, и тогда этап предсказания выполняется по формулам:

$$\hat{\mathbf{x}}_{k|k-1} = \sum_{i=0}^{2n} w_i^{(m)} \chi_{k|k-1}^{(i)}, \quad (3.9)$$

$$\mathbf{P}_{k|k-1} = \sum_{i=0}^{2n} w_i^{(c)} (\chi_{k|k-1}^{(i)} - \hat{\mathbf{x}}_{k|k-1}) (\cdot)^\top + \mathbf{Q}_{k-1}. \quad (3.10)$$

Здесь $w_i^{(m)}$ и $w_i^{(c)}$ — веса для вычисления среднего и ковариации соответственно, которые определяются параметрами α , κ и дополнительным параметром β (характеризует априорные знания о распределении, для гауссовского случая обычно $\beta = 2$). Веса вычисляются следующим образом:

$$w_0^{(m)} = \frac{\lambda}{n + \lambda}, \quad (3.11)$$

$$w_0^{(c)} = \frac{\lambda}{n + \lambda} + (1 - \alpha^2 + \beta), \quad (3.12)$$

$$w_i^{(m)} = w_i^{(c)} = \frac{1}{2(n + \lambda)}, \quad i = 1, \dots, 2n. \quad (3.13)$$

Аналогично этапу предсказания, для шага обновления (коррекции) в UKF формируется «измерительное сигма-облако». Все сигма-точки, полученные на предыдущем шаге предсказания, пропускаются через нелинейную функцию измерения $h(\cdot)$:

$$\mathbf{z}_{k|k-1}^{(i)} = h\left(\chi_{k|k-1}^{(i)}\right), \quad i = 0, \dots, 2n. \quad (3.14)$$

На основе преобразованных сигма-точек вычисляются аппроксимированные предсказанные измерения и ковариация измерения:

$$\hat{\mathbf{z}}_{k|k-1} = \sum_{i=0}^{2n} w_i^{(m)} \mathbf{z}_{k|k-1}^{(i)}, \quad (3.15)$$

$$\mathbf{S}_k = \sum_{i=0}^{2n} w_i^{(c)} \left(\mathbf{z}_{k|k-1}^{(i)} - \hat{\mathbf{z}}_{k|k-1} \right) (\cdot)^\top + \mathbf{R}_k, \quad (3.16)$$

где $\mathbf{S}_k$ — аппроксимированная ковариация измерения, а $\mathbf{R}_k$ — ковариация шума измерения.

Далее вычисляется кросс-ковариация между состоянием и измерением:

$$\mathbf{C}_{xz} = \sum_{i=0}^{2n} w_i^{(c)} \left(\mathbf{x}_{k|k-1}^{(i)} - \hat{\mathbf{x}}_{k|k-1} \right) \left(\mathbf{z}_{k|k-1}^{(i)} - \hat{\mathbf{z}}_{k|k-1} \right)^\top. \quad (3.17)$$

На основании этих величин определяется усиление Калмана:

$$\mathbf{K}_k = \mathbf{C}_{xz} \mathbf{S}_k^{-1}. \quad (3.18)$$

И, наконец, выполняется обновление оценки состояния и ковариации:

$$\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \left(\mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1} \right), \quad (3.19)$$

$$\mathbf{P}_{k|k} = \mathbf{P}_{k|k-1} - \mathbf{K}_k \mathbf{S}_k \mathbf{K}_k^\top. \quad (3.20)$$

Таким образом, UKF позволяет корректно учитывать нелинейные зависимости как в модели перехода (динамической модели), так и в модели измерения, обеспечивая более точную оценку состояния по сравнению с EKF, особенно при существенной нелинейности функций $f(\cdot)$ и $h(\cdot)$.

Другие методы. Существуют и другие подходы, например, Гауссовские суммирующие фильтры (Gaussian-Sum Filter, GSF), которые аппроксимируют негауссовы распределения взвешенной суммой нескольких Гауссовых распределений [87]. Каждый компонент этой суммы обрабатывается отдельным фильтром Калмана. Этот метод является компромиссом между точностью и вычислительной сложностью. В случаях, когда распределение вероятностей сильно отличается от Гауссова, более правильным подходом являются методы Монте-Карло для последовательных данных, также известные как фильтры частиц [88; 89]. Эти методы аппроксимируют апостериорную плотность вероятности набором случайных выборок (частиц) с соответствующими весами. Фильтры частиц способны работать с любыми нелинейностями и негауссовыми шумами, однако их вычислительная сложность значительно выше, чем у UKF, что делает их применение в системах реального времени затруднительным.

Именно поэтому в рассматриваемой задаче использование фильтров частиц будет неоптимальным.

Сравнительный анализ EKF и UKF. Как видно из таблицы 2, UKF обеспечивает более высокую точность без необходимости численного дифференцирования, что критично для сложных моделей в системах распознавания. Более того, многочисленные экспериментальные исследования [90—92] подтверждают, что при сопоставимых настройках UKF демонстрирует меньшую среднеквадратичную ошибку и лучшую сходимость.

Таблица 2 — Ключевые характеристики расширенного и сигма-точечного фильтров Калмана

Критерий	EKF	UKF
Порядок точности аппроксимации	Первый (линейный)	Второй (или выше)
Требуется ли Якобиан	Да	Нет
Численная устойчивость	Зависит от F_k, H_k ; возможны расходимости	Выше, так как нет ошибки линеаризации
Вычислительная сложность	$\mathcal{O}(n^3)$ (обновление ковариаций) + расчёт Якобиана	$\mathcal{O}(n^3)$ при $2n + 1$ сигма-точках, но без производных
Параметры настройки	Q, R	$Q, R, \alpha, \beta, \kappa$

3.1.2 Тесносвязанные схемы комплексирования

Методы оптимальной фильтрации позволяют получать единую оценку состояния объекта путем объединения данных от различных датчиков. При использовании классического, слабосвязанного подхода вектор состояния и вектор измерений представляются в одной системе координат. Например, когда положение и скорость объекта определяются навигационной системой, предполагается, что датчики также предоставляют измерения в этих же

физических величинах. Такое решение упрощает математическую модель, однако требует предварительной обработки «сырых» данных. Этот промежуточный этап преобразования неизбежно ведет к потере части информации и накоплению коррелированных погрешностей.

В противоположность этому, тесносвязанный подход, изначально разработанный для навигационных систем GNSS/INS, предполагает работу фильтра непосредственно с «сырыми» измерениями — псевдодальностями и доплеровскими скоростями, — а не с вычисленными на их основе координатами приёмника. Такая архитектура позволяет значительно повысить устойчивость к выбросам и обеспечить непрерывную работу даже при потере сигналов от нескольких спутников [93; 94]. В современных реализациях применяются адаптивные алгоритмы EKF/UKF и их SVD-устойчивые модификации, которые повышают точность навигации в условиях плотной городской застройки на 30–50 % [95; 96].

Идеи тесносвязанного комплексирования находят применение и в задачах технического зрения. Например, в системах визуально-инерциальной одометрии совместная обработка данных с лидара, камеры и IMU в рамках единого факторного графа позволяет напрямую связать координаты особых точек и инерциальные измерения с состоянием платформы, как это реализовано в LVI-SAM [97]. Работы последнего десятилетия показывают, что переход от слабосвязанных («каскадных») архитектур к тесносвязанным обеспечивает прирост точности и отказоустойчивости, особенно при наличии ошибок взаимной калибровки датчиков.

Тем не менее большинство известных работ, описывающих тесносвязанные системы, исходят из предположения об однородности пространства состояний и наблюдений. Как правило, измерения либо выражаются в навигационных координатах (GNSS/INS), либо принадлежат пространству трёхмерной карты (LiDAR-SLAM). В задачах компьютерного зрения ситуация иная: нейронная сеть формирует детекции в двумерных координатах изображения (u, v) и оценивает относительную глубину d , в то время как состояние светофора описывается

его трёхмерной позицией в мировой системе координат. Предлагаемая в работе модель (см. разд. 3.2) сохраняет эту гетерогенность, выполняя комплексирование данных непосредственно в фильтре. Такой подход позволяет избежать промежуточных преобразований и потерь, максимально используя информацию, содержащуюся в исходных данных.

Важно отметить, что подобная постановка задачи встречается в литературе лишь фрагментарно и не является систематизированной. Хотя авторы работ [95; 98] упоминают возможность использования гибридного пространства измерений, полноценных моделей, объединяющих пространственный вектор состояния с двумерными детекциями на изображении без предварительной триангуляции, предложено не было. Таким образом, область тесно связанного комплексирования применительно к системам распознавания светофоров на основе видеоданных остаётся малоизученной, что подчёркивает новизну и актуальность предлагаемого исследования.

3.2 Оптимальная фильтрация координат местоположения светофоров

3.2.1 Описание алгоритмов фильтрации

Как отмечалось ранее, выходные данные детектора содержат шум, который необходимо отфильтровать для получения стабильной оценки положения объекта. В ходе исследования были разработаны и сопоставлены два алгоритма на основе расширенного и сигма-точечного фильтров Калмана.

Подход 1. Расширенный фильтр Калмана. Первый (наивный) подход основан на ЕKF с упрощенной моделью состояния, предполагающей, что положение светофора статично. Вектор состояния s_t^{gm} и модель движения описываются следующим образом:

$$\mathbf{s}_t^{gm} = [x_{gm}, y_{gm}]^\top, \quad \mathbf{s}_{k+1}^{gm} = \mathbf{s}_k^{gm} + \mathbf{w}_k, \quad (3.21)$$

где $[x_{gm}, y_{gm}]$ — координаты центра светофора в глобальной системе координат (GM), а $\mathbf{w}_k \sim \mathcal{N}(0, \mathbf{Q}_k)$ — шум модели, который в данном контексте отражает не движение самого светофора, который считается строго статичным, а накопленные ошибки навигационной системы трамвая, выход которой используется для перевода координат светофора в глобальную систему координат. Ковариационная матрица шума процесса $\mathbf{Q}_k$ задаётся диагональной: $\mathbf{Q}_k = \text{diag}(\sigma_x^2, \sigma_y^2)$, где σ_x и σ_y — малые величины, характеризующие уровень неопределённости в оценке положения, обусловленный ошибками одометрии и навигации ТС.

Вектор измерений $\mathbf{z}_t^{vr}$ представляет собой координаты светофора, измеренные относительно транспортного средства, в полярной СК:

$$\mathbf{z}_t^{vr} = [r, \varphi]^\top = h(\mathbf{s}_t^{gm}) + \mathbf{v}_t, \quad (3.22)$$

где r — расстояние до светофора, φ — угол пеленга, а $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R}_t)$ — шум измерений. Выбор полярных координат обусловлен тем, что при обработке изображений с камеры угловая точность выше, чем точность оценки расстояния. Это отражено в ковариационной матрице шума измерений: $\mathbf{R}_t = \text{diag}(\sigma_r^2, \sigma_\varphi^2)$, где $\sigma_r^2 \gg \sigma_\varphi^2$.

Нелинейная функция преобразования $h(\cdot)$ из глобальных декартовых координат в полярные координаты, связанные с автомобилем, имеет вид:

$$h(\mathbf{s}_t^{gm}) = \begin{bmatrix} \sqrt{(x_{gm})^2 + (y_{gm})^2} \\ \arctan(y_{gm}/x_{gm}) \end{bmatrix}. \quad (3.23)$$

Соответствующий якобиан $\mathbf{H}$ для данной функции вычисляется как:

$$\mathbf{H} = \frac{\partial h}{\partial \mathbf{s}^{gm}} = \begin{bmatrix} \frac{x_{gm}}{r} & \frac{y_{gm}}{r} \\ -\frac{y_{gm}}{r^2} & \frac{x_{gm}}{r^2} \end{bmatrix}, \quad \text{где } r = \sqrt{x_{gm}^2 + y_{gm}^2}. \quad (3.24)$$

Процедура фильтрации выполняется в соответствии со стандартными этапами предсказания и коррекции ЕKF, как описано в разделе 3.1. Итоговое

состояние $\hat{\mathbf{s}}_t^{gm}$ представляет собой “оптимальную” оценку положения светофора. Однако применение данного фильтра не дало желаемых результатов вследствие ограниченности модели состояния и неточностей, вносимых линеаризацией.

Подход 2. Сигма-точечный фильтр Калмана с расширенной моделью.

Второй подход использует UKF и значительно более полную модель состояния, что позволяет точнее описывать геометрию и положение светофора. Уравнение состояния сохраняет предположение о статичности (3.21), но вектор состояния $\mathbf{s}_t^{gm}$ расширен:

$$\mathbf{s}_t^{gm} = [x_{gm}, y_{gm}, z_{gm}, w, l, h, \theta_{gm}]^\top, \quad (3.25)$$

где $[x_{gm}, y_{gm}, z_{gm}]$ — 3D-координаты центра светофора в системе GM, $[w, l, h]$ — его ширина, длина и высота в метрах, а θ_{gm} — угол ориентации (рысканье). Включение высоты z_{gm} и ориентации θ_{gm} критически важно для корректного проецирования 3D-модели на плоскость изображения. Ковариационная матрица шума процесса $\mathbf{Q}_k$ для этого вектора имеет размерность 7×7 и является диагональной, отражая независимые малые флуктуации для каждого из параметров.

Ключевой особенностью этого подхода является тесно связанная схема комплексирования, в которой вектор измерений $\mathbf{z}_t^{vr}$ напрямую использует выходные данные нейронной сети:

$$\mathbf{z}_t^{vr} = [u_{cam}, v_{cam}, d_{cam}, \theta_{vr}, w_{det}, l_{det}, h_{det}]^\top. \quad (3.26)$$

Здесь $[u_{cam}, v_{cam}]$ — координаты центра описывающего прямоугольника (детекции) светофора в пикселях, d_{cam} — оцененная глубина (расстояние до светофора), θ_{vr} — угол ориентации относительно автомобиля, а $[w_{det}, l_{det}, h_{det}]$ — оцененные детектором габариты. Ковариационная матрица шума измерений $\mathbf{R}_k$ также диагональная (7×7), причем ее диагональные элементы $(\sigma_u^2, \sigma_v^2, \sigma_d^2, \dots)$ настроены в соответствии с наблюдаемой точностью нейросетевой модели для каждого из параметров.

Функция преобразования $h(\cdot)$ из вектора состояния в вектор измерений выполняется последовательно через цепочку координатных преобразований: из

глобальной системы (GM) в систему автомобиля (VR), а затем в систему камеры (CAM) с последующим проецированием на плоскость изображения.

$$\mathbf{p}_{vr} = \mathbf{R}_m^{gm \rightarrow vr} \cdot \mathbf{p}_{gm} + \mathbf{T}_m^{gm \rightarrow vr} \quad (3.27)$$

$$\mathbf{p}_{cam} = \mathbf{R}_m^{vr \rightarrow cam} \cdot \mathbf{p}_{vr} + \mathbf{T}_m^{vr \rightarrow cam} \quad (3.28)$$

$$\begin{bmatrix} u_{cam} \\ v_{cam} \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \end{bmatrix} \begin{bmatrix} x_{cam}/z_{cam} \\ y_{cam}/z_{cam} \\ 1 \end{bmatrix} \quad (3.29)$$

$$d_{cam} = z_{cam} \quad (3.30)$$

$$\theta_{vr} = \theta_{gm} + \Delta\theta_{gm/vr} \quad (3.31)$$

где $\{f_x, f_y, c_x, c_y\}$ — внутренние параметры камеры, а $\Delta\theta_{gm/vr}$ — разница в углах рысканья между системами координат GM и VR. Размеры объекта $[w, l, h]$ передаются в вектор измерений напрямую.

Алгоритм UKF, подробно описанный в разделе 3.1, применяется для выполнения этапов предсказания и коррекции без необходимости вычисления якобианов, что является значительным преимуществом при такой сложной функции $h(\cdot)$.

Таким образом, использование расширенного вектора состояния в UKF позволяет отслеживать все ключевые параметры светофора, включая его ориентацию и габариты. Тесно связанная архитектура обеспечивает более точную и устойчивую фильтрацию за счет прямого использования измерений от нейронной сети без промежуточных преобразований, которые могут накапливать ошибку. Применение UKF вместо EKF устранило проблемы, связанные с линеаризацией, и повысило общую робастность и точность оценки состояния.

3.2.2 Оценка точности фильтрации

Для оценки и сопоставления эффективности предложенных подходов к фильтрации было проведено комплексное моделирование. В ходе эксперимента использовались синтетические данные, параметры которых были подобраны таким образом, чтобы максимально точно отражать реальные дорожные условия. Была сгенерирована протяженная и сложная траектория движения трамвая, включающая как прямолинейные участки, так и повороты различной кривизны. Модели шумов навигационной системы и нейросетевого детектора были откалиброваны в соответствии с характеристиками, наблюдаемыми у реальных сенсоров и алгоритмов.

Генерация синтетических данных для моделирования. Траектория движения трамвая была сгенерирована с помощью В-сплайнов, обеспечивающая плавность и наличие поворотов. Истинное положение трамвая в момент времени t описывается вектором $\mathbf{p}_t = [x_t, y_t, z_t, \psi_t]^\top$, включающим 3D-координаты и угол рысканья. Выходные данные навигационной системы моделировались путем добавления аддитивного белого гауссовского шума к истинной траектории:

$$\hat{\mathbf{p}}_t = \mathbf{p}_t + \mathbf{n}_p, \quad \mathbf{n}_p \sim \mathcal{N}(0, \mathbf{Q}_{nav}), \quad (3.32)$$

где $\mathbf{Q}_{nav}$ — диагональная ковариационная матрица, определяющая уровень шума для координат и ориентации. Светофоры размещались вдоль сгенерированной траектории на реалистичных боковых смещениях и высотах.

Процесс измерений, выполняемый нейросетевым детектором, моделировался путем проецирования истинного 3D-положения светофора в систему координат камеры с последующим добавлением шума. Ключевой особенностью модели шума является симуляция неидеальной оценки расстояния, свойственной монокулярным системам. Измеренное расстояние $\hat{d}$ до объекта систематически завышается с увеличением истинного расстояния d . Эта зависимость

моделируется следующим образом:

$$\hat{d} = d \cdot \left(1 + f_{bias} \left(1 - e^{-d/d_{0,bias}} \right) \right) + \varepsilon_d, \quad (3.33)$$

где f_{bias} — максимальный коэффициент систематической ошибки, а $d_{0,bias}$ — характерная дистанция, определяющая скорость роста ошибки. Компонент ε_d представляет собой дополнительный случайный шум, моделируемый экспоненциальным распределением, масштаб которого также зависит от дистанции, определяясь параметрами f_{stoch} и $d_{0,stoch}$. Остальные компоненты вектора измерений, такие как пиксельные координаты центра детекции, ориентация и габариты, моделировались путем добавления к их истинным значениям аддитивного гауссовского шума с нулевым средним и заранее определенными дисперсиями. Численные значения стандартных отклонений и параметров модели шума дистанции для каждого сценария приведены в таблице 3.

Таблица 3 — Параметры шумов измерений для различных сценариев моделирования

Обозначение	Ед. изм.	Слабый	Средний	Сильный
σ_{uv}	пикс.	2.0	5.0	10.0
σ_θ	рад	0.1	0.2	0.3
σ_{dim}	м	0.05	0.1	0.2
f_{bias}	—	0.08	0.15	0.25
$d_{0,bias}$	м	30.0	35.0	40.0
f_{stoch}	—	0.05	0.10	0.18
$d_{0,stoch}$	м	40.0	45.0	50.0
σ_r	м	1.0	2.0	3.0
σ_φ	рад	0.05	0.1	0.15

Результаты моделирования. Сравнительный анализ проводился в трех сценариях с различными уровнями шума измерений: низким, средним и высоким. Для демонстрации робастности и предельных возможностей алгоритмов далее приводятся результаты, полученные в сценарии с высоким уровнем шума, который является наиболее сложным и показательным.

На рисунке 3.2 представлен анализ сходимости обоих фильтров для одного из светофоров. Графики показывают, что UKF демонстрирует значительно более быструю сходимость к истинному положению по сравнению с EKF. Ошибка положения для UKF быстро уменьшается и стабилизируется на низком уровне, в то время как оценка EKF остается смещенной и содержит значительные флуктуации на протяжении всего времени наблюдения. Распределение ошибок также подтверждает, что UKF обеспечивает более точную и стабильную оценку.

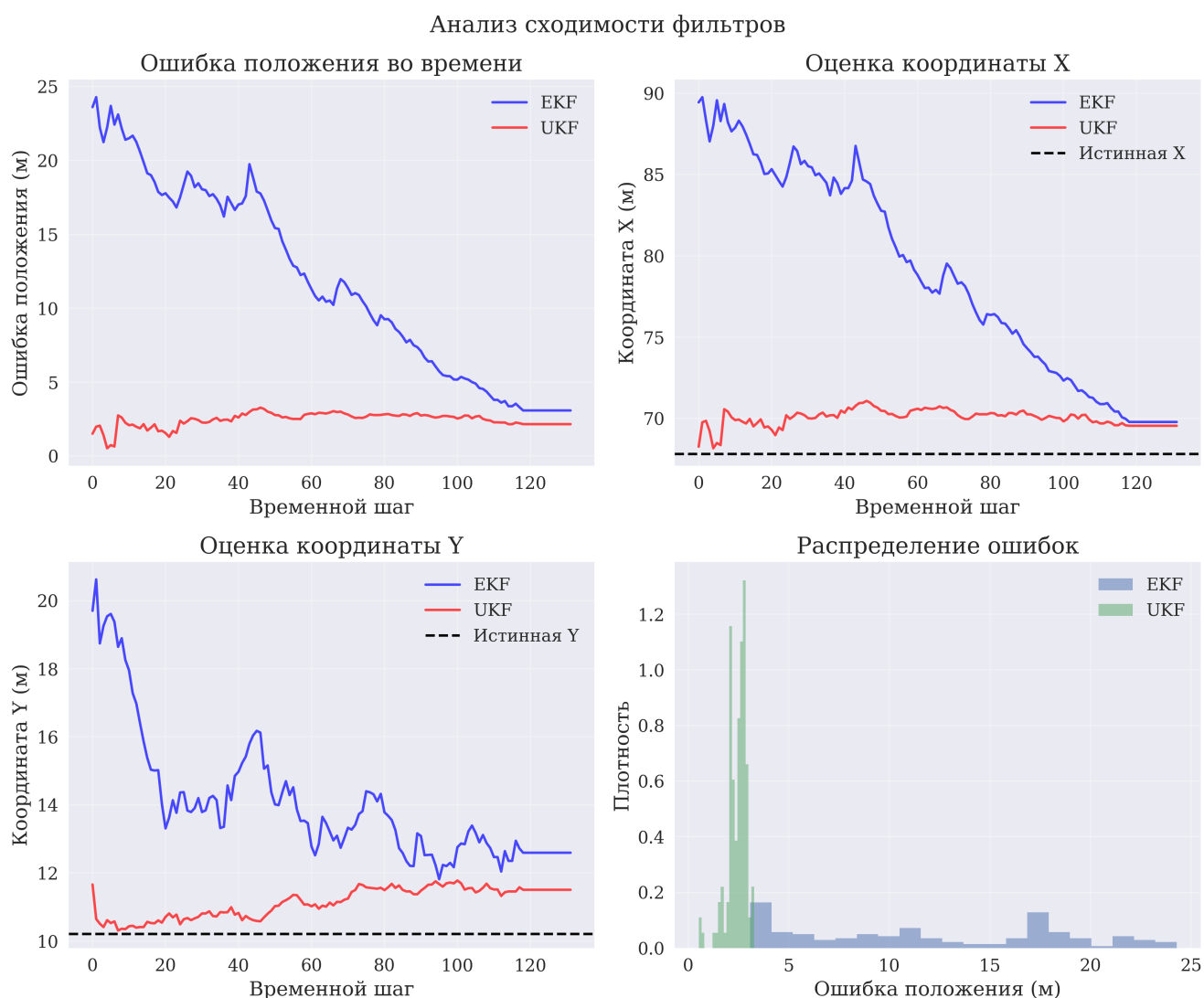


Рисунок 3.2 — Анализ сходимости фильтров в условиях сильного шума

На рисунке 3.3 показаны результаты оценки местоположения для трех случайно выбранных светофоров. Оценки, полученные с помощью

UKF, формируют плотное облако точек вокруг истинного положения, что свидетельствует о высокой точности и устойчивости алгоритма. В отличие от этого, оценки ЕKF сильно рассеяны и часто значительно отклоняются от истинных координат, что указывает на его неспособность эффективно справляться с нелинейностями и высоким уровнем шума.

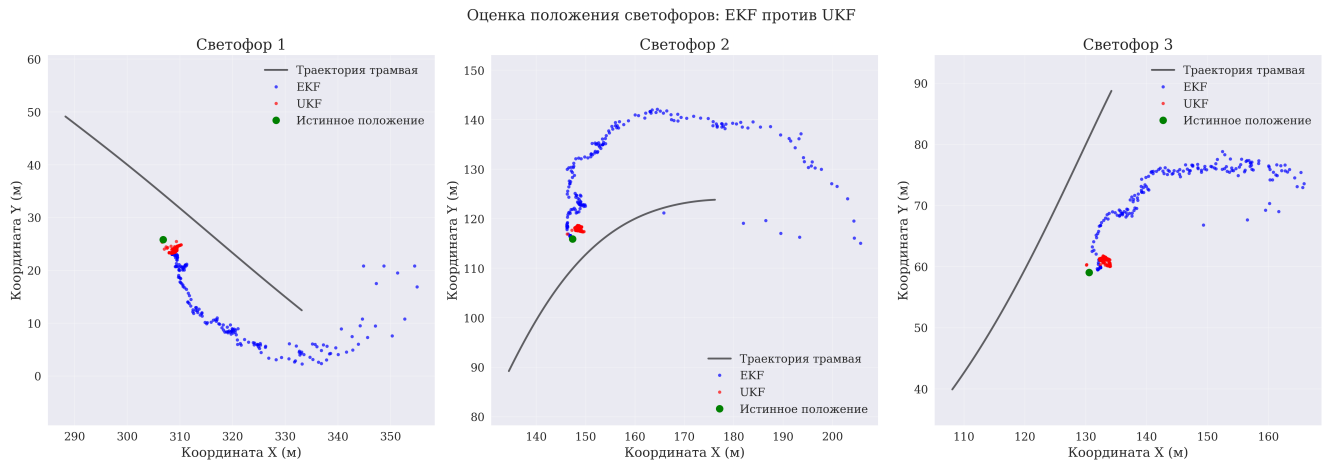


Рисунок 3.3 — Сравнение оценок положения светофоров для ЕKF и UKF

Количественное сравнение среднеквадратичной ошибки (СКО) представлено на рисунке 3.4. Диаграмма наглядно иллюстрирует превосходство UKF над ЕKF. Для каждого из трех рассматриваемых светофоров СКО у UKF оказывается в несколько раз ниже, что подтверждает выводы, сделанные на основе анализа траекторий оценок.

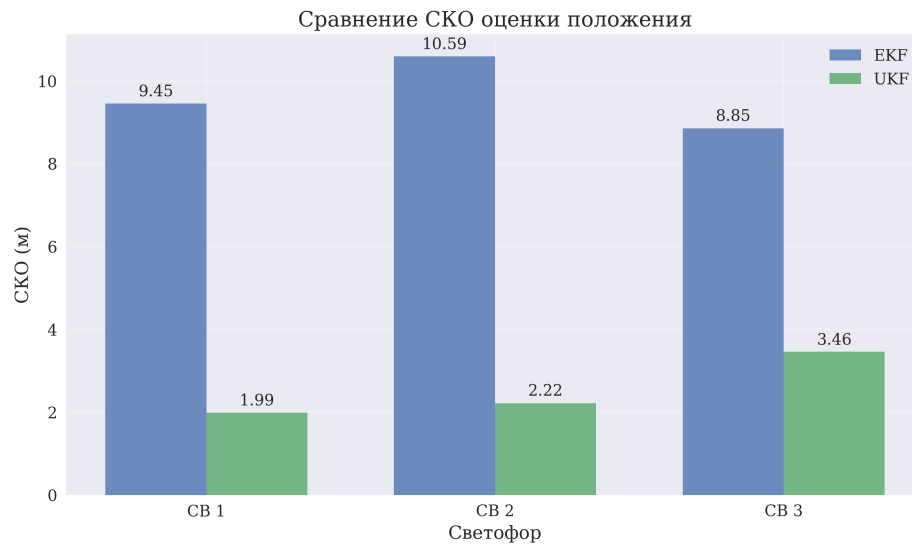


Рисунок 3.4 — Сравнение СКО оценки положения для EKF и UKF

Обобщенные результаты моделирования для всех трех сценариев сведены в таблице 4. Данные, усредненные по всем светофорам в сцене, показывают, что UKF стабильно превосходит EKF по всем ключевым показателям во всех сценариях. Улучшение по СКО составляет от 78% до 85% в пользу UKF. С ростом уровня шума ошибка обоих фильтров закономерно возрастает, однако UKF демонстрирует значительно большую робастность: в сценарии с сильным шумом его средняя СКО (1.83 м) и средняя абсолютная ошибка (САО) (1.48 м) остаются на уровне, сопоставимом с показателями EKF в сценарии со слабым шумом (3.01 м и 1.44 м).

Таблица 4 — Сводные метрики оценки положения светофоров разных фильтров

Метрика	Слабый шум		Средний шум		Сильный шум	
	EKF	UKF	EKF	UKF	EKF	UKF
Средняя СКО, м	3.01	0.41	5.92	0.69	10.42	1.83
Средняя САО, м	1.44	0.33	2.63	0.56	4.42	1.48
Улучшение по СКО	84.2%		85.3%		78.1%	

Таким образом, результаты экспериментального моделирования однозначно свидетельствуют о существенных преимуществах подхода на основе сигматочечного фильтра Калмана с расширенной моделью состояния. Его способность

адекватно обрабатывать нелинейные зависимости без линеаризации в сочетании с более полной моделью, описывающей как положение, так и геометрию объекта, обеспечивает значительный прирост точности, робастности и скорости сходимости по сравнению с наивным подходом на базе расширенного фильтра Калмана. Полученные данные служат основанием для выбора UKF в качестве основного алгоритма фильтрации в разрабатываемой системе.

3.3 Сопоставление результатов с картой

После идентификации и отслеживания светофоров с помощью бортовой системы технического зрения, осуществляется их сопоставление с априорной информацией, хранящейся в цифровой карте. Этот процесс позволяет не только верифицировать обнаруженные объекты, но и связать их с контекстом дорожной сети, что необходимо для принятия правильных и безопасных решений. В отличие от многих современных систем автономного вождения, которые полагаются на дорогостоящие и трудоемкие в создании HD-карты, в данной работе используется более гибкий и экономически эффективный подход, основанный на семантических «упрощенных» цифровых картах.

3.3.1 Семантическая цифровая карта

Основой для навигации служит специализированная семантическая цифровая карта, которая создается без использования дорогостоящих мобильных лабораторий с лидарами и не претендует на сантиметровую точность, свойственную HD-картам. Вместо этого, карта представляет собой 2D-схему, обогащенную ключевой семантической информацией, достаточной для безопасного движения трамвая. Такой подход позволяет значительно снизить

затраты на создание и поддержку картографической основы. В работе [99], также исследуют использование упрощенных «топологических» карт вместо полноценных HD-карт, подтверждая перспективность данного направления.

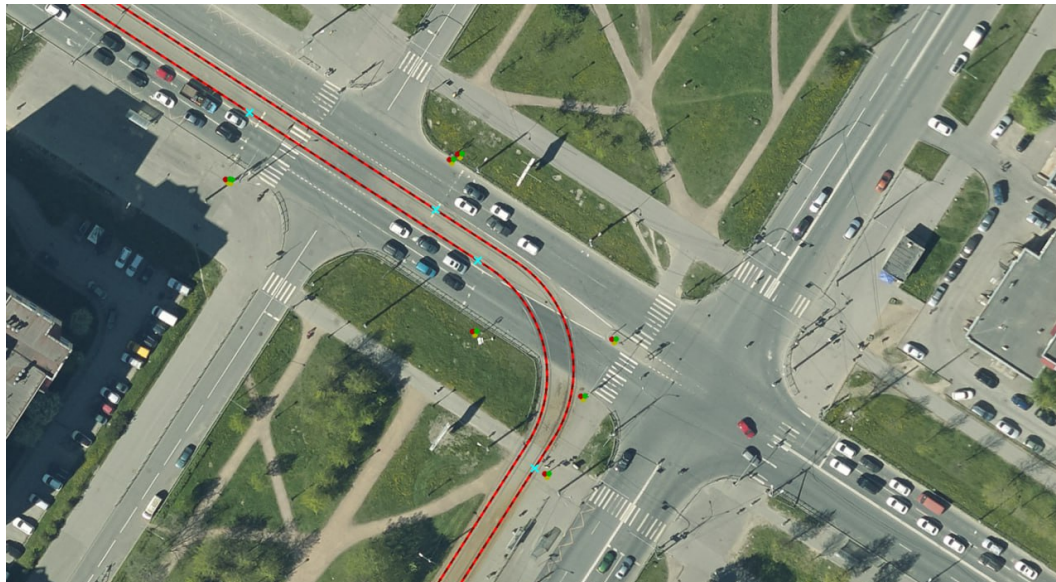


Рисунок 3.5 — Цифровая карта светофоров и путей

Процесс создания карты начинается с использования спутниковых снимков высокого разрешения, как показано на рисунке 3.5. Оператор вручную определяет и наносит на карту координаты светофоров. В случаях, когда спутниковый снимок не позволяет однозначно определить количество или тип светофоров на одном столбе, используются панорамные снимки улиц из общедоступных картографических сервисов (например, Яндекс.Карты). Такой комбинированный подход, использующий спутниковые и уличные изображения, также находит отражение в современных исследованиях по созданию карт для автономных систем, например, в работе [100].

Карта содержит следующие семантические слои:

- **Координаты светофоров.** 2D-координаты (широта и долгота) каждого светофора.
- **Тип светофора.** Классификация по назначению (автомобильный, пешеходный, трамвайный).

- **Привязка сигналов к маневрам.** Для сложных перекрестков карта содержит информацию о том, какая группа сигналов светофора (например, дополнительная секция) соответствует какому маневру (движение прямо, поворот направо). Это позволяет транспортному средству отслеживать только релевантные для его маршрута сигналы, что перекликается с идеями автоматического определения семантики перекрестков [45].
- **Маршрутная сеть.** Векторное представление трамвайных путей и разрешенных траекторий движения.

Ключевым преимуществом данного подхода является возможность быстрого и дешевого обновления карты. Поскольку трамваи, оборудованные камерами, постоянно курсируют по городу, система получает непрерывный поток актуальных данных о дорожной обстановке. Эти данные используются для обнаружения изменений: появления новых светофоров, демонтажа старых или изменения их конфигурации. Такой подход, по сути, является формой «краудсорсинга» картографических данных. Возможности обновления HD-карт с помощью данных, собираемых парком транспортных средств, активно исследуются, и методы обнаружения изменений в реальном времени, описанные в [101], подтверждают состоятельность этой концепции. Обширный обзор методов создания и обновления HD-карт, включая краудсорсинговые подходы, представлен в [102]. Таким образом, создается динамическая картографическая система, способная адаптироваться к изменениям городской среды без привлечения специализированных и дорогостоящих картографических служб.

3.3.2 Алгоритм сопоставления с картой

После получения вектора координат отслеженных светофоров от системы технического зрения, выполняется их сопоставление с объектами

из семантической карты. Процедура сопоставления аналогична той, что описана в [47], где детекции объектов на изображениях связываются с геопространственными данными для создания карты.



Рисунок 3.6 — Сопоставление результатов с картой

Сопоставление координат местоположений светофоров с рассчитанными значениями осуществляется с использованием расстояния Махаланобиса, которое учитывает ковариацию ошибок определения координат. Это позволяет сделать метрику более робастной к неопределенностям как в данных детектора, так и в координатах объектов на карте. Расстояние определяется по следующей формуле

$$d_{ib} = \sqrt{(s_i^{gm} - o_b^{gm})^T P^{-1} (s_i^{gm} - o_b^{gm})} \quad (3.34)$$

где o_b^{gm} – вектор координат для b -го светофора из карты, s_i^{gm} – вектор координат i -го детектированного светофора, P^{-1} – обратная матрица ковариации ошибок определения координат светофора.

При формировании матрицы «стоимости» для последующего решения задачи о назначениях вводится строгое ограничение по типу объектов: сопоставление возможно только между светофорами одного и того же типа. Этот подход гарантирует, что, например, автомобильный светофор, обнаруженный системой технического зрения, не будет ошибочно ассоциирован с пешеходным светофором с карты. Таким образом, для пар «измерение–объект» с совпадающими типами, значение в матрице стоимости обратно пропорционально расстоянию Махаланобиса, а для пар с разными типами

«стоимость» устанавливается бесконечно большой. Далее с использованием Венгерского алгоритма (2.10) определяются оптимальные пары, что позволяет однозначно ассоциировать каждый отслеженный светофор с его представлением в цифровой карте.

3.4 Классификация сигналов светофоров

3.4.1 Классификационная нейронная сеть

Решение задачи классификации осуществляется с использованием глубокой нейронной сети. На вход сети подается фрагмент изображения светофора $\hat{I}_j^{t,m}$, сформированный детектором, а на выходе сети формируется класс сигнала светофора $c_j^{t,m}$. Классификатор сигналов светофора содержит 25 классов. Архитектура сети состоит из двух последовательных блоков: извлечения признаков и предсказания классов. Для извлечения признаков используется нейронная сеть архитектуры RegNet, показанная на рисунке 3.7.

Архитектура RegNet была выбрана благодаря ее высокой производительности и эффективности по сравнению с другими современными архитектурами. В отличие от подходов, направленных на разработку единственной, узкоспециализированной сети, RegNet является результатом проектирования пространства проектных решений (design space) для сетей. Такой подход позволяет создавать простые, регулярные и высокоэффективные модели, которые хорошо обобщаются в различных вычислительных режимах. Ключевая идея заключается в том, что ширина и глубина эффективных сетей могут быть описаны квантованной линейной функцией. Это приводит к созданию моделей, которые не только точны, но и до 5 раз быстрее на ГПУ по сравнению с аналогичными

моделями, такими как EfficientNet [103; 104].

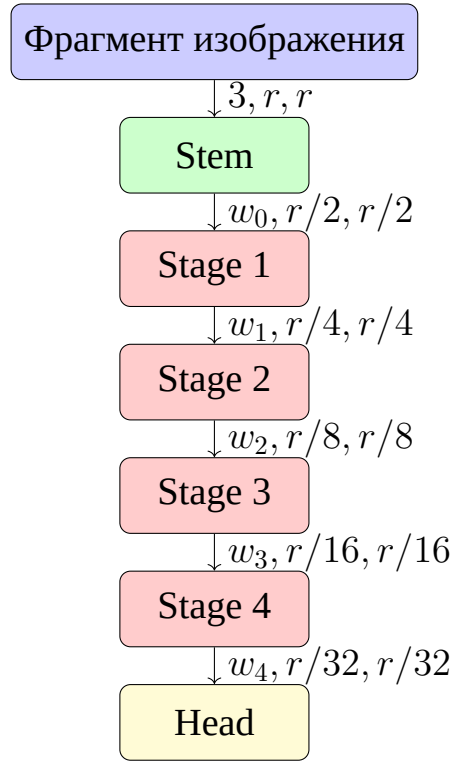


Рисунок 3.7 — Архитектура RegNet

Поскольку распределение классов в собранном датасете сильно несбалансированно, вместо стандартной функции кросс-энтропии при обучении используется Focal Loss. Для многоклассовой классификации кросс-энтропия определяется как

$$CE(p_t) = - \sum_{c=1}^C y_c \log(p_c), \quad (3.35)$$

где p_c – предсказанная вероятность для класса c , а y_c – “one-hot encoding” вектор, определяющий принадлежность к классу c . Focal Loss добавляет масштабирующий множитель для автоматического перевзвешивания классов

$$FL(p_t) = - \sum_{c=1}^C \alpha_c y_c (1 - p_c)^\gamma \log(p_c), \quad (3.36)$$

где γ – фокусирующий параметр ($\gamma \geq 0$), а α_c – балансирующий коэффициент для класса c . Масштабный коэффициент $(1 - p_c)^\gamma$ уменьшается до

нуля по мере роста уверенности нейронной сети в правильности предсказанного класса, а α_c позволяет явно задать вес для каждого класса.

Разметка фрагментов изображений, содержащих светофоры, использует систему классов учитывающую: тип светофора (автомобильный, трамвайный, пешеходный), ориентацию (направление), количество секций (5 секционный, 3 секционный, 2 секционный), а также специальные символы и их комбинации (прямо, налево, прямо и налево и т. д.).

Примеры светофоров разных классов показаны на рисунке 3.8.



Рисунок 3.8 — Примеры светофоров разных классов

3.4.2 Обеспечение временной стабильности сигналов

Для обеспечения стабильности распознавания сигналов светофора во времени применяется механизм временной фильтрации. Этот механизм решает проблему «дрожания» предсказаний, когда нейронная сеть может выдавать разные классы для одного и того же светофора на последовательных кадрах, например, из-за кратковременных помех или ошибок сети.

На каждом кадре для каждого обнаруженного светофора выполняется поиск соответствующего ему светофора на предыдущем кадре. Связывание осуществляется за счет трекинга (см. разд. 2.2) и совпадения типа светофора (автомобильный, пешеходный или трамвайный). Фильтрация по типу не позволяет связать светофоры, предназначенные для разных участников движения.

Вместо простого подсчета кадров, фильтрация основана на времени, в течение которого сигнал непрерывно наблюдается. Для повышения робастности и предотвращения ситуаций, когда процесс подтверждения постоянно прерывается случайными шумами, применяется механизм, основанный на параллельной проверке гипотез. Вместо одного класса-кандидата, система одновременно отслеживает несколько гипотез о возможном состоянии светофора, позволяя им конкурировать между собой.

Для каждого отслеживаемого светофора j система поддерживает набор активных гипотез $\mathcal{P}_j$. Каждая гипотеза $p \in \mathcal{P}_j$ характеризуется кортежем (c_p, A_p, L_p) , где c_p — класс сигнала, A_p — накопленное время подтверждения, а L_p — временная метка последнего наблюдения.

При поступлении нового результата классификации C_j^t с временной меткой τ_t , алгоритм выполняет следующие шаги:

1. **Обновление существующей гипотезы:** Если для класса C_j^t уже существует гипотеза p ($C_j^t = c_p$), ее накопленное время увеличивается: $A_p \leftarrow A_p + (\tau_t - \tau_{t-1})$, а время последнего наблюдения обновляется: $L_p \leftarrow \tau_t$.
2. **Создание новой гипотезы:** Если класс C_j^t не совпадает ни с одной из активных гипотез и отличается от текущего стабильного состояния S_j , создается новая гипотеза $p_{\text{new}} = (C_j^t, \tau_t - \tau_{t-1}, \tau_t)$, которая добавляется в набор $\mathcal{P}_j$.

Переключение стабильного состояния S_j происходит, как только накопленное время A_p для какой-либо гипотезы p превысит заданный порог T_{thresh} :

$$S_j^t = \begin{cases} c_p, & \text{если } \exists p \in \mathcal{P}_j : A_p \geq T_{\text{thresh}} \\ S_j^{t-1}, & \text{в противном случае} \end{cases} \quad (3.37)$$

После успешного переключения все активные гипотезы из набора $\mathcal{P}_j$ сбрасываются.

Для предотвращения накопления устаревших данных вводится также порог T_{timeout} . На каждом шаге система проверяет все гипотезы $p \in \mathcal{P}_j$. Если с момента последнего наблюдения гипотезы прошло времени больше, чем T_{timeout} (т.е. $\tau_t - L_p > T_{\text{timeout}}$), она считается неактуальной и удаляется из набора $\mathcal{P}_j$. Это гарантирует, что единичные шумные распознавания со временем отфильтровываются и не влияют на принятие решений.

Выводы

В данной главе была представлена методика точной и робастной оценки пространственного положения светофоров и классификации их сигналов. Были рассмотрены и сопоставлены различные подходы к фильтрации зашумленных данных, получаемых от системы технического зрения. Разработанный подход, основанный на тесно связанной архитектуре с использованием сигма-точечного фильтра Калмана, продемонстрировал свою эффективность. Кроме того, были описаны алгоритмы для сопоставления полученных данных с цифровой картой и для надежной классификации сигналов светофоров с обеспечением временной стабильности. Основные результаты и выводы, можно сформулировать следующим образом:

- Проведен систематический анализ методов оптимальной фильтрации, на основе которого обоснован выбор сигма-точечного фильтра Калмана как наиболее подходящего для решения нелинейной задачи оценки координат

светофоров, превосходящего по своим характеристикам расширенный фильтр Калмана.

- Разработана новая тесносвязанная модель фильтрации координат светофоров, которая, в отличие от существующих подходов, напрямую интегрирует трехмерный вектор состояния объекта в глобальной системе координат и двумерные детекции (пиксельные координаты и глубина) с выхода нейронной сети.
- Реализован алгоритм на основе UKF с расширенным семимерным вектором состояния, включающим 3D-координаты, физические размеры и ориентацию светофора, что обеспечивает более точное геометрическое моделирование по сравнению с упрощенными подходами.
- Показано преимущество предложенного подхода на базе UKF с полной моделью состояния над наивным подходом на базе EKF с упрощенной моделью, что подтверждает эффективность отказа от линеаризации и использования более полного описания состояния.
- Разработана методика сопоставления отфильтрованных координат светофоров со специализированной семантической цифровой картой, представляющей собой экономически эффективную альтернативу дорогостоящим HD-картам.
- Алгоритм сопоставления с картой использует расстояние Махаланобиса для учета неопределенностей оценок и венгерский алгоритм для оптимального назначения, что обеспечивает робастную привязку детектированных объектов к объектам на карте.
- Для классификации сигналов светофоров применена эффективная нейронная сеть на базе архитектуры RegNet. Для борьбы со значительным дисбалансом классов в обучающей выборке использована функция потерь Focal Loss, что повысило качество распознавания редких классов сигналов.
- Внедрен механизм временной стабилизации результатов классификации на основе конечного автомата, который устраняет «дрожание»

предсказаний и обеспечивает смену состояния сигнала только после его уверенного распознавания на протяжении нескольких последовательных кадров.

Глава 4. Методика оценки точности алгоритма идентификации светофоров

4.1 Анализ существующих подходов оценки точности работы подсистемы распознавания светофоров

Оценка точности работы систем распознавания сигналов светофора является многогранной задачей, которая охватывает как точность обнаружения и классификации на отдельных изображениях, так и стабильность отслеживания объектов в видеопотоке. В данном разделе рассматриваются ключевые подходы к оценке, включая используемые наборы данных и метрики, которые стали стандартом в данной области.

4.1.1 Наборы данных для оценки точности

Для обучения и тестирования систем распознавания светофоров необходимо использовать большие и разнообразные наборы данных, которые отражают реальные дорожные условия. Современные исследования в области автономного вождения привели к созданию нескольких эталонных наборов данных, которые широко используются для оценки производительности систем компьютерного зрения:

- **KITTI** [105] является одним из первых и наиболее известных наборов данных для автономного вождения. Он содержит часы видеоданных, снятых в различных условиях в немецком городе Карлсруэ. Хотя KITTI более известен своими задачами по оценке одометрии и обнаружению 3D-объектов, он также включает аннотации для светофоров.
- **Cityscapes** [106] представляет собой крупномасштабный набор данных, ориентированный на семантическое понимание городских уличных сцен.

Он содержит видеопоследовательности из 50 городов и предлагает высококачественные пиксельные аннотации для 30 классов объектов, включая светофоры.

- **BDD100K** [107] — это один из самых крупных и разнообразных наборов данных для автономного вождения, содержащий более 100 000 видеороликов. Он включает аннотации для различных задач, в том числе для обнаружения светофоров, и охватывает разнообразные погодные условия, время суток и сценарии вождения.
- **DTLD (Deep Traffic Light Dataset)** [108] — это набор данных, специально созданный для задачи детального распознавания светофоров. Он содержит более 50 000 изображений с более чем 250 000 аннотаций светофоров, включая их точное состояние (красный, желтый, зеленый, выключен), тип и ориентацию. Данные были собраны в Дании в различных погодных и световых условиях.
- **Waymo Open Dataset** [109] — это еще один крупномасштабный набор данных, выпущенный компанией Waymo. Он содержит данные с различных сенсоров, включая камеры и лидары, и предоставляет высококачественные 2D и 3D-аннотации для различных объектов, включая светофоры и их состояния.

Несмотря на наличие наборов данных, их использование сопряжено с определенными трудностями. Во-первых, процесс сбора и аннотирования данных требует значительных ресурсов [110]. Во-вторых, даже самые большие наборы данных могут быть несбалансированными, так как редкие, но критически важные сценарии (например, нестандартные сигналы светофора или сложные погодные условия) могут быть представлены в недостаточном объеме.

Важно отметить, что для оценки системы распознавания светофоров в контексте трамвайного движения перечисленные наборы данных не подходят. Они сфокусированы на автомобильных и пешеходных светофорах и не содержат аннотированных изображений трамвайных светофоров, которые значительно отличаются по форме и набору сигналов. Это фундаментальное ограничение

делает невозможным их применение для рассматриваемой задачи. В связи с этим для обучения и тестирования системы был собран и размечен собственный специализированный набор данных, который адекватно отражает условия эксплуатации трамвая.

4.1.2 Метрики оценки точности

Первый уровень оценки точности — это изображение. Нейронная сеть генерирует прямоугольную область вероятного расположения светофора и определяет класс сигнала светофора (движение, остановка, внимание). Правильным считается совпадение областей, сгенерированных сетью, с областями разметки светофора и полное совпадение классов. Благодаря наличию разметки можно сформировать полную группу событий (включая ошибки первого и второго рода) и рассчитать основные и производные параметры точности.

На уровне отдельных изображений основными метриками являются:

- **Точность (Precision)** показывает, какая доля объектов, обнаруженных системой как светофоры, действительно является светофорами. Она рассчитывается как:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (4.1)$$

где TP (True Positives) — количество правильно обнаруженных светофоров, а FP (False Positives) — количество ложных обнаружений.

- **Полнота (Recall)** показывает, какую долю реальных светофоров система смогла обнаружить. Она рассчитывается как:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (4.2)$$

где FN (False Negatives) — количество пропущенных светофоров.

- **F1-мера (F1-Score)** является гармоническим средним между точностью и полнотой и предоставляет сбалансированную оценку

$$\text{F1-Score} = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad (4.3)$$

На втором уровне процедуры распознавания могут использоваться целые видеопоследовательности, что помогает справляться с засветами, погодными артефактами и прочими визуальными искажениями. Обычно для оценки точности работы таких алгоритмов используется метрика MOTA, которая учитывает три типа ошибок: ложные срабатывания, пропуски и ошибки смены идентификатора.

Метрика MOTA рассчитывается по следующей формуле:

$$\text{MOTA} = 1 - \frac{\sum_t (FN_t + FP_t + \text{IDSW}_t)}{\sum_t GT_t} \quad (4.4)$$

где:

- t — индекс кадра в видеопоследовательности;
- FN_t — количество пропущенных объектов (false negatives) в кадре t ;
- FP_t — количество ложных обнаружений (false positives) в кадре t ;
- IDSW_t — количество переключений идентификаторов (ID switches) в кадре t ;
- GT_t — количество реальных объектов (ground truth) в кадре t .

MOTA может принимать значения от $-\infty$ до 1. Значение, близкое к 1, указывает на высокую точность отслеживания. Эта метрика стала де-

факто стандартом для оценки производительности алгоритмов отслеживания нескольких объектов.

Хотя Precision, Recall и MOTA, являются незаменимыми для оценки качества модуля восприятия (обнаружения и отслеживания), они принципиально недостаточны для оценки производительности всего контура принятия решений. Основная цель системы — не просто «видеть» светофоры, а обеспечивать выполнение трамваем правильного действия на перекрестке. Система может иметь отличный показатель MOTA, но не выдать своевременную команду на остановку, что делает ее неэффективной или даже опасной. Последствия ошибок, такие как проезд на красный свет, не отражаются адекватно в этих метриках, которые склонны рассматривать все ошибки как равнозначные.

Таким образом, требуется более высокий уровень оценки, который анализирует правильность управляющих решений системы. Это особенно актуально для автономных систем, где производительность напрямую влияет на безопасность и эффективность. В таких случаях оценка основывается на конечном результате: остановилось ли транспортное средство правильно на запрещающий сигнал и продолжило ли движение на разрешающий?

Для СПВ, рассматриваемой в данной работе, прямая оценка усложняется наличием человека-водителя, который остается главным действующим лицом. Действия водителя часто упреждают решения системы. Для преодоления этой проблемы оценка должна проводиться в «теневом режиме». В этом режиме система работает параллельно с водителем, принимая собственные решения, но не воздействуя на органы управления. Путем логгирования предполагаемых действий системы и их сравнения с действиями водителя и реальным состоянием светофора можно провести более детальный анализ [111].

4.2 Действия и сценарии работы системы

Комплексную оценку подсистемы можно определить, используя целевые и фактические действия в зависимости от сигнала светофора:

- ОСТАНОВКА при запрещающем сигнале;
- ДВИЖЕНИЕ при разрешающем сигнале.

Под целевыми понимается истинное (правильное) действие, задаваемое водителем. В контексте проезда перекрестка наиболее критичным действием является ОСТАНОВКА. Данное событие взято как базовое («позитивное» с точки зрения матрицы несоответствий). На основе предпринятых системой действий и фактических результатов можно сформировать следующую матрицу соответствия действий, показанную в таблице 5.

Таблица 5 — Матрица соответствия действий

Событие	Целевое действие	Фактическое действие системы
TN	ДВИЖЕНИЕ	ДВИЖЕНИЕ (A_N)
TP	ОСТАНОВКА	ОСТАНОВКА (A_P)
FN	ОСТАНОВКА	ДВИЖЕНИЕ (A_N)
FP	ДВИЖЕНИЕ	ОСТАНОВКА (A_P)

Исходя из таблицы 5 формируются события:

- TP – ОСТАНОВКА на **запрещающий** сигнал;
- TN – ДВИЖЕНИЕ на **разрешающий** сигнал;
- FP – ОСТАНОВКА на **разрешающий** сигнал;
- FN – ДВИЖЕНИЕ на **запрещающий** сигнал. Самая критичная ошибка с точки зрения безопасности движения.

Работа системы описывается двумя основными сценариями, которые напрямую связаны с положением трамвая относительно перекрестка.

Сценарий 1: Подъезд к перекрестку (S_1) — активируется, когда система распознает светофор на пути следования трамвая. Основная задача на этом этапе — безопасно подъехать к стоп-линии и принять решение о дальнейших действиях на основе сигнала светофора. Если сигнал запрещающий, система должна инициировать остановку. Если сигнал разрешающий, трамвай может продолжить движение через перекресток.

Сценарий 2: Трогание с перекрестка (S_2) — этот сценарий является логическим продолжением первого в случае, если трамвай совершил остановку на запрещающий сигнал. Система ожидает смены сигнала на разрешающий, после чего позволяет водителю начать движение.

Для реализации описанных сценариев система оперирует набором состояний. Каждое состояние соответствует определенному этапу процесса проезда перекрестка. Переходы между состояниями инициируются событиями, такими как распознавание светофора или изменение его сигнала.

Основные состояния системы:

- **Неактивна** (Q_0) – система находится в режиме ожидания, видеопоток с камер анализируется, но светофор в зоне видимости отсутствует.
- **Подъезд** (Q_1) – система обнаружила светофор и сопровождает его. В этом состоянии происходит анализ сигнала и принимается решение: ОСТАНОВКА или ДВИЖЕНИЕ.
- **Ожидание** (Q_2) – трамвай остановился у стоп-линии на запрещающий сигнал. Система ожидает смены сигнала на разрешающий.
- **Проезд** (Q_3) – трамвай движется через перекресток на разрешающий сигнал. Система продолжает контролировать сигнал светофора на случай его изменения.

Диаграмма состояний системы представлена на рисунке 4.1.

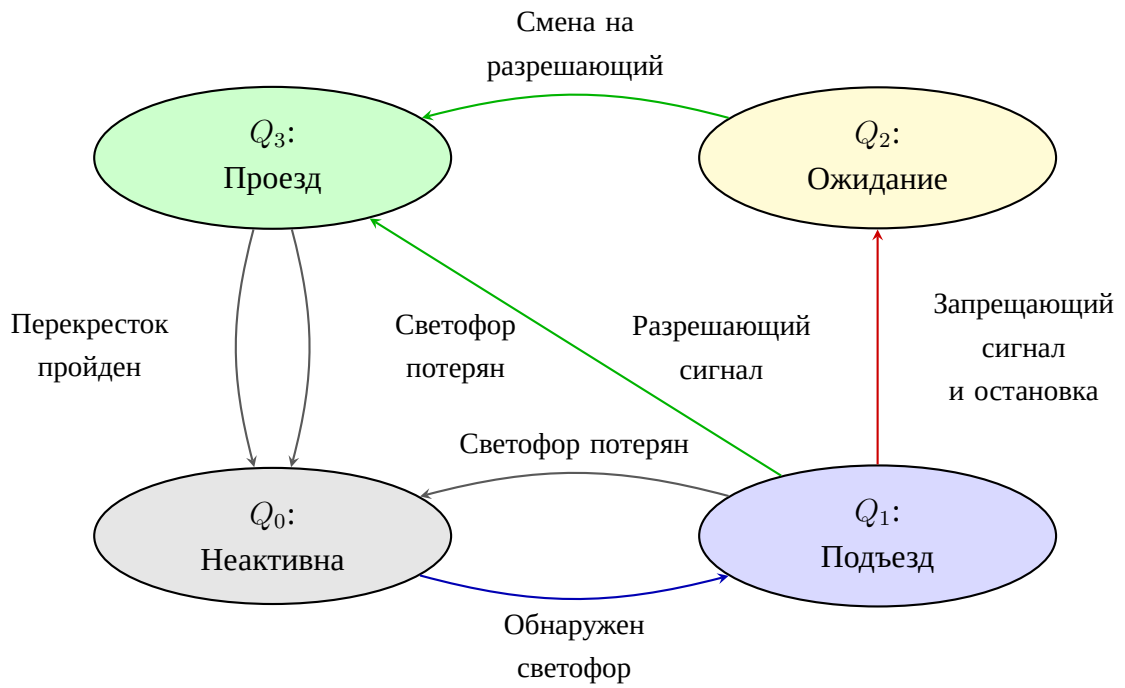


Рисунок 4.1 — Диаграмма состояний системы управления проездом перекрестка

Переходы между состояниями, показанными на диаграмме, представлены в таблице 6.

Таблица 6 — Основные внешние события и соответствующие переходы состояний системы

Событие	Условие/Описание	Переход состояния
Обнаружен светофор	Камера зафиксировала и распознала объект «светофор»	$Q_0 \rightarrow Q_1$
Светофор потерян	Ранее обнаруженный светофор больше не находится в поле зрения камер (например, трамвай проехал перекресток)	Q_1 или $Q_3 \rightarrow Q_0$
Анализ сигнала светофора: запрещающий сигнал	В состоянии Q_1 обнаружен запрещающий сигнал, трамвай останавливается	$Q_1 \rightarrow Q_2$
Анализ сигнала светофора: разрешающий сигнал	В состоянии Q_1 обнаружен разрешающий сигнал	$Q_1 \rightarrow Q_3$
Смена сигнала на разрешающий	В состоянии Q_2 система ожидает смены сигнала для продолжения движения	$Q_2 \rightarrow Q_3$
Перекресток пройден	После успешного проезда перекрестка (стоп-линия пройдена)	$Q_3 \rightarrow Q_0$

4.3 Входная информация и критерии формирования событий

Для оценки эффективности разработанной подсистемы помощи водителю необходимо определить набор входных данных и разработать систему критериев, позволяющих объективно классифицировать её поведение.

4.3.1 Входная информация и эталонное поведение

Основным источником данных для оценки служат регистрируемые параметры движения и управления транспортным средством. В качестве эталонного поведения, с которым сравниваются действия автоматической подсистемы, принимаются действия опытного водителя.

В качестве входной информации выступают:

- D_i — положение ручки контроллера хода, задаваемое водителем;
- v_i — текущая скорость движения трамвая;
- H_i — положение ручки контроллера хода, устанавливаемое подсистемой;
- s_i — целевая скорость для торможения, формируемая подсистемой.

Значения ручки контроллера хода для разных моделей трамваев могут различаться, но имеют симметричную шкалу $[-b, b]$. Для простоты запишем их в диапазоне $H_i \in [-b, b]$ и $D_i \in [-b, b]$. При этом значения ручки хода в диапазоне от $[-b, 0]$ считаются «торможением», а $(0, b]$ – «разгоном».

4.3.2 Критерии формирования событий

Для объективной оценки работы системы необходимо формализовать критерии, по которым её действия будут классифицироваться как корректные или некорректные в рамках заданных сценариев S_1 (принятие решения об остановке перед перекрестком) и S_2 (принятие решения о начале движения). Для этого вводится система событий, основанная на сравнении действий подсистемы и эталонных действий водителя. Критерии для классификации событий представлены в таблице 7 и таблице 8. Эти критерии помогают

разделить все исходы на четыре категории: истинно положительные (TP), истинно отрицательные (TN), ложноположительные (FP) и ложноотрицательные (FN).

Таблица 7 — Критерии формирования группы событий для сценария S_1

Событие	Действие	Критерий ручки хода	Критерий скорости
TP	A_P	$H_i \leq 0, D_i \leq 0$	$s_i \leq v_i, v_i \rightarrow 0$
TN	A_N	$H_i \geq D_i$	$v_i > 0$
FP	A_P	$H_i < D_i, H_i < 0$	$s_i < v_i$
FN	A_N	$H_i \geq 0, D_i \leq 0$	$s_i > v_i, v_i \rightarrow 0$

Описание критериев для сценария S_1 :

1. **TP:** Система корректно принимает решение об остановке. Критерий $H_i \leq 0, D_i \leq 0$ показывает, что, и система, и водитель замедляют вагон. Критерий $s_i \leq v_i, v_i \rightarrow 0$ указывает на снижение скорости до нуля.
2. **TN:** Система корректно принимает решение о продолжении движения. $H_i \geq D_i$ показывает, что система не затормаживает вагон, а $v_i > 0$ подтверждает отсутствие необходимости в торможении.
3. **FP:** Система некорректно принимает решение об остановке. Критерий $H_i < 0$ демонстрирует, что система замедляет вагон, хотя критерии $H_i < D_i$ и $s_i < v_i$ показывают, что для этого нет необходимости.
4. **FN:** Система некорректно принимает решение о продолжении движения. Критерии $H_i \geq 0$ и $s_i > v_i$ показывают, что система не затормаживает вагон, а $D_i \leq 0$ и $v_i \rightarrow 0$ указывают на то, что водитель тормозит на перекрестке.

Таблица 8 — Критерии формирования группы событий для сценария S_2

Событие	Действие	Критерий ручки хода	Критерий скорости
TP	A_P	$H_i \leq 0, D_i \leq 0$	$v_i = 0, s_i = 0$
TN	A_N	$H_i > D_i, D_i > 0$	$s_i \geq v_i, v_i > 0$
FP	A_P	$H_i \leq 0, D_i > 0$	$s_i = 0, v_i > 0$
FN	A_N	$H_i > 0, D_i \leq 0$	$s_i > 0, v_i = 0$

Описание критериев для сценария S_2 :

1. **ТР:** Система корректно принимает решение о продолжении стоянки. Критерий $H_i \leq 0$, $D_i \leq 0$ показывает, что, и система, и водитель применяют тормозные позиции. Критерий $v_i = 0$, $s_i = 0$, показывает, что вагон стоит на месте.
2. **TN:** Система корректно принимает решение о начале движения. $H_i > D_i$, $D_i > 0$ показывает, что, и система, и водитель применяют ходовые позиции (т.е. разгоняют вагон), а $s_i \geq v_i$, $v_i > 0$ подтверждает корректность начала движения.
3. **FP:** Система некорректно принимает решение о продолжении стоянки. Критерии $H_i \leq 0$, $s_i = 0$ демонстрируют, что система держит вагон на месте, хотя критерии $D_i > 0$, $v_i > 0$ показывают, что необходимо начать движение.
4. **FN:** Система некорректно принимает решение о начале движения. Критерии $H_i > 0$, $s_i > 0$ показывают, что система начинает движение, а $D_i \leq 0$, $v_i = 0$ указывают на то, что водитель все еще стоит на перекрестке.

4.4 Результаты оценки точности системы

Для проведения оценки разработанной системы в реальных городских условиях был собран обширный набор данных. Сбор данных производился с нескольких трамваев, курсирующих по различным маршрутам. Такой подход позволил сформировать репрезентативную выборку, охватывающую широкий спектр дорожных ситуаций, с которыми сталкивается автономное транспортное средство в городской среде.

Навигационные данные обеспечивали точную информацию о местоположении и динамике движения, видеопоток служил основным источником информации

для системы компьютерного зрения, а действия водителя использовались для формирования эталонной разметки о реальных состояниях светофоров.

Суммарно, тестовая выборка включает 437 событий проезда перекрестков, полученных из более чем 100 часов видеозаписей, собранных на 8 различных трамвайных маршрутах города Санкт-Петербург. С целью обеспечения разнообразия условий, данные были сбалансированы по времени суток и сезону. Около 70% (306 событий) были записаны в дневное время, а оставшиеся 30% (131 событие) – в ночное, что позволило оценить работу системы при различном освещении. Сезонное распределение включает 60% данных, собранных в летний период (262 события), и 40% – в зимний (175 событий), что позволило учесть такие факторы, как снег, блики от мокрого асфальта и низкое солнце. В выборку также включены как стандартные, так и сложные сценарии проезда перекрестков, с автомобильными и бело-лунными трамвайными светофорами, что является ключевым фактором для адекватной оценки робастности и надежности предлагаемой системы.



Рисунок 4.2 — Примеры данных из тестовой выборки

Для проведения сравнительного анализа были определены три конфигурации системы. Базовая конфигурация (K_1) включает в себя использование EKF и одной камеры; конфигурация (K_2) использует UKF и одну камеру; третья конфигурация (K_3) также использует UKF, но в сочетании с данными от двух камер. Ниже представлены результаты, полученные при использовании предлагаемых подходов для оценки точности трех конфигураций исходной системы:

Таблица 9 — Оценка точности представленных алгоритмов

	Смешанные перекрестки (в том числе пешеходные светофоры)		
События	K_1	K_2	K_3
TP	160	181	192
TN	167	183	180
FP	83	57	60
FN	27	16	5
Precision	65.84	76.05	76.19
Recall	85.56	91.88	97.46

В таблице 9 рассчитываются как базовые, так и производные значения параметров, при этом из анализа видно, что для однокамерного сетапа качество работы второго алгоритма лучше в силу использования «глубокой» схемы комплексирования измерений на уровне изображения, что позволяет повысить точность предсказания на 10%, а полноту на 6%. С другой стороны, при использовании двух камер качество работы системы повышается именно в части полноты на 6%, что объясняется лучшей возможностью детектирования дальних «дублирующих» светофоров. На рисунке 4.3 представлены фрагменты офлайн-стендов детектирования светофоров на перекрестке. Кубоидами показаны все объекты, распознанные нейронной сетью, цветом показаны разные типы объектов (желтый – автомобиль, красный или зеленый – светофор, синими точками – положение объектов на карте, спроецированное в кадр). Видно, что детектированные нейросетью объекты имеют связь с точками на земле, что

говорит об успешном решении задачи сопоставления.

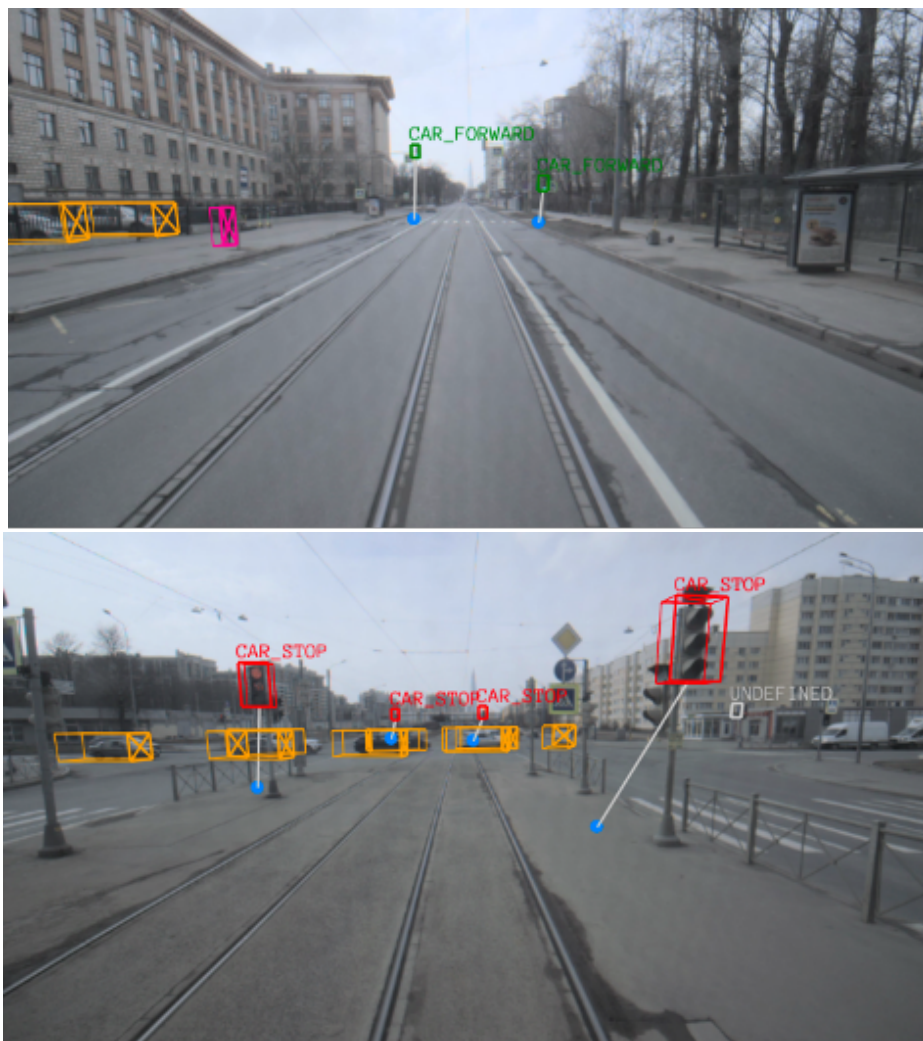


Рисунок 4.3 — Визуализация результатов идентификации светофоров на перекрестках города Санкт-Петербург

Более детальный анализ ошибок показывает, что в контексте управления автономным трамваем ошибки второго рода (FN, ложно-отрицательные срабатывания) представляют наибольшую опасность. Пропуск существующего сигнала светофора, особенно запрещающего, может привести к аварийной ситуации. Результаты показывают, что предложенная конфигурация с двумя камерами (K_3) демонстрирует крайне высокую полноту (97.46%), минимизируя число таких опасных пропусков до 5 случаев на всю тестовую выборку. Это

свидетельствует о высокой надежности системы в части обнаружения светофоров на перекрестках города.

С другой стороны, ошибки первого рода (FP, ложно-положительные срабатывания), которые снижают точность (precision), менее критичны для безопасности. Ложное определение сигнала как запрещающего может привести к излишнему снижению скорости или необоснованной остановке, что влияет на эффективность и плавность движения, но не создает угрозы безопасности. Значение точности в 76.19% для конфигурации K_3 указывает на наличие таких ошибок, что является приемлемым компромиссом в пользу максимизации полноты и, следовательно, безопасности.

Для контекстуализации полученных результатов проведем сравнение с альтернативными подходами, представленными в современных исследованиях. В недавней работе [15] приводится всесторонний анализ производительности современных архитектур детектирования, таких как YOLOv8, на различных публичных наборах данных. Результаты в этой работе сильно варьируются в зависимости от сложности набора данных. Например, на наборе данных HDTLR, содержащем относительно крупные и четкие изображения светофоров, достигаются значения точности 96% и полноты 98%. Однако на более сложном и разнообразном наборе данных DTLD, который включает множество мелких объектов и нерелевантных светофоров, те же модели показывают точность около 75%, а полноту всего 54%.

На этом фоне результаты, полученные алгоритмом K_3 , демонстрируют высокую конкурентоспособность. Достигнутое значение полноты в **97.46%** находится на уровне лучших показателей, полученных на более простых наборах данных, что подчеркивает исключительную надежность системы в обнаружении всех релевантных светофоров и минимизации опасных пропусков (ошибок второго рода). В то же время, точность в **76.19%** сопоставима с результатами на сложных наборах данных, таких как DTLD. Это свидетельствует о том, что собственный набор данных, собранный в реальных условиях эксплуатации трамваев, обладает высокой сложностью и вариативностью. Таким образом,

предложенная система обеспечивает баланс между безопасностью (максимальная полнота) и эффективностью (сопоставимая точность) при работе в сложных городских условиях.

Выводы

В рамках данной главы была разработана и апробирована комплексная методика оценки точности работы СПВ трамвая при проезде регулируемых перекрестков. Были решены следующие задачи:

- Проведен анализ существующих подходов к оценке систем распознавания светофоров, который выявил их неприменимость для специфических задач трамвайного движения. Это обосновало необходимость разработки собственной методики, ориентированной на оценку конечного результата работы системы.
- Предложена новая оценка, основанная не на метриках качества детектирования (таких как МОТА), а на анализе правильности управляющих действий системы («ОСТАНОВКА»/«ДВИЖЕНИЕ»), что позволяет напрямую судить о ее безопасности и эффективности.
- Определены два основных сценария работы системы: S_1 (подъезд к перекрестку) и S_2 (трогание с перекрестка), для которых были разработаны детальные критерии формирования событий.
- Сформированы четкие критерии для классификации событий (ТР, ТН, ФР, FN), основанные на сопоставлении действий системы с действиями водителя, выступающими в роли эталона, что обеспечивает объективность оценки в «теневом режиме».
- Для проведения экспериментов был собран уникальный набор данных, отражающий реальные условия эксплуатации трамваев в Санкт-

Петербурге, включая различные погодные условия, время суток и типы светофоров.

- Проведена оценка трех различных комбинаций сенсоров и алгоритмов, по результатам которой наилучшие показатели продемонстрировала конфигурация с двумя камерами и модифицированным алгоритмом K_3 , достигнув полноты **97.46%** и точности **76.19%**.
- Подтверждена высокая надежность разработанной системы в части минимизации критических ошибок второго рода (проезд на запрещающий сигнал). Высокое значение полноты свидетельствует о способности системы обнаруживать практически все релевантные сигналы светофоров.
- Проведено сравнение полученных результатов с современными исследованиями, которое показало конкурентоспособность предложенного решения. Система обеспечивает баланс между безопасностью (максимальная полнота) и эффективностью (сопоставимая точность) при работе в сложных городских условиях.

Заключение

В настоящей диссертационной работе было проведено исследование, направленное на решение комплексной задачи надежной идентификации светофоров для беспилотных транспортных средств с использованием многокамерных систем. Актуальность данной работы обусловлена растущей потребностью в создании надежных и точных систем восприятия для современных СПВ и АТС. Основное внимание в исследовании было уделено разработке новых алгоритмов и методик, обеспечивающих объединение данных от различных сенсоров, сопровождение светофоров в динамическом окружении и их точную локализацию относительно картографических данных с целью определения релевантности. Проведенное исследование позволило создать комплексное решение, повышающее точность идентификации светофоров.

Основные результаты работы заключаются в следующем:

1. Проведен анализ существующих моделей и алгоритмов решения задачи идентификации светофоров в различных условиях обстановки и разном наборе сенсоров. Обоснован наиболее рациональный облик системы, исходные данные для формирования алгоритмов.
2. Проведен анализ существующих моделей и алгоритмов детектирования и классификации объектов сцены, обоснованы преимущества использования 3D-измерений нейронных сетей.
3. Сформирована общая постановка задачи идентификации светофоров в мультикамерных системах и проведена ее декомпозиция на этапы выполнения.
4. По каждому предложенному этапу: межкадровое сопоставление, межкамерное сопоставление, фильтрация измерений, сопоставление с картой, предложены новые алгоритмы решения.
5. Разработан алгоритм объединения результатов детекции каждой камеры в отдельности (межкадровый трекинг) и совместно на основе

“взвешенных” комбинированных метрик схожести, учитывающий близость эмбедингов и пересечения областей детекций на изображении. На основе имеющегося датасета произведена оценка весовых коэффициентов новой функции схожести, обеспечивающей надежное сопоставление детекций даже в условиях несинхронизированных измерений с камер, когда области светофоров на изображении могут иметь отличные сигналы в силу эффекта моргания.

6. Разработан алгоритм оценивания (фильтрации) положения светофоров в пространстве на основе UKF фильтра Калмана с учетом детекций на изображении, более точно учитывающий нелинейность уравнений связи и обеспечивающий более высокую точность позиционирования, что на практике позволяет повысить точность идентификации на 10% и полноты на 6% для однокамерного сетапа и точность идентификации на 10% и полноты на 12% для двух камер.
7. Предложен алгоритм сопоставления полученных координат детекций с координатами светофоров из карты на основе метрики Махаланобиса и допустимых классов нейронных сетей, что позволяет на практике уменьшить требования к разметке светофоров, находящихся в геоинформационной базе данных.
8. Разработана методика оценки точности результатов работы алгоритма идентификации, учитывающая действия водителя и не требующая проведения процедур разметки данных, что существенно ускоряет процесс оценки качества работы подсистемы идентификации светофоров.

Разработанные алгоритмы и методики вносят весомый вклад в развитие области автономного вождения, предлагая практичное и эффективное решение одной из ключевых задач восприятия. Предложенные подходы могут быть интегрированы в существующие СПВ для расширения их функциональности и повышения уровня безопасности. Дальнейшие направления исследований включают исследование возможностей применения новых типов сенсоров

и моделей машинного обучения для достижения еще большей точности и надежности.

Список сокращений и условных обозначений

3D	трёхмерный (three-dimensional)
АТС	Автономное транспортное средство
ИТС	Интеллектуальные транспортные системы
ПВ	Плотность вероятности
САО	Средняя абсолютная ошибка
СКО	Среднеквадратичное отклонение
СНС	Сверточная нейронная сеть
СПВ	Система помощи водителю
ТС	Транспортное средство
ADAS	Advanced Driver-Assistance Systems — системы помощи водителю
ASIL	Automotive Safety Integrity Level — уровень полноты безопасности автомобиля
BEV	Bird's-Eye-View — вид сверху
C-V2X	Cellular Vehicle-to-Everything — сотовая технология связи «транспортное средство-ко-всему»
DSRC	Dedicated Short-Range Communications — беспроводная связь малого радиуса действия
EKF	Extended Kalman Filter — расширенный фильтр Калмана
F1	F1-Score — гармоническое среднее точности и полноты
FN	False Negative — ложноотрицательное срабатывание
FOV	Field of View — поле зрения
FP	False Positive — ложноположительное срабатывание
GAN	Generative Adversarial Network — генеративно-сопоставительная сеть
GM	Global Metric — глобальная метрическая система координат

GNSS	Global Navigation Satellite System — глобальная навигационная спутниковая система
GPS	Global Positioning System — глобальная система позиционирования
GSF	Gaussian-Sum Filter — гауссовский суммирующий фильтр
GT	Ground Truth — эталонная разметка
HD	High Definition — высокое разрешение
IMU	Inertial Measurement Unit — инерциальный измерительный модуль
INS	Inertial Navigation System — инерциальная навигационная система
IOU	Intersection over Union — метрика «пересечение над объединением»
KF	Kalman Filter — фильтр Калмана
LVI-SAM	LiDAR-Visual-Inertial SLAM — метод совместной локализации и картирования
MOT	Multi-Object Tracking — отслеживание нескольких объектов
MOTA	Multiple Object Tracking Accuracy — комплексная метрика точности трекинга
NMS	Non-Maximum Suppression — подавление немаксимумов
RGB	Red-Green-Blue — цветовая модель
RegNet	Regularized Network — семейство сверточных сетей
SAE	Society of Automotive Engineers — Общество автомобильных инженеров
SCT	Single-Camera Tracking — однокамерный трекинг
SLAM	Simultaneous Localization and Mapping — одновременная локализация и построение карты
SPaT	Signal Phase and Timing — данные о фазе и времени сигнала светофора

SSD	Single Shot Detector — нейросетевая архитектура для детекции объектов
SVD	Singular Value Decomposition — сингулярное разложение матрицы
TN	True Negative — истинно отрицательное срабатывание
TP	True Positive — истинно положительное срабатывание
UID	Unique Identifier — уникальный идентификатор
UKF	Unscented Kalman Filter — сигма-точечный фильтр Калмана
V2I	Vehicle-to-Infrastructure — взаимодействие «транспортное средство-инфраструктура»
V2N	Vehicle-to-Network — взаимодействие «транспортное средство-сеть»
V2P	Vehicle-to-Pedestrian — взаимодействие «транспортное средство-пешеход»
V2V	Vehicle-to-Vehicle — взаимодействие «транспортное средство-транспортное средство»
V2X	Vehicle-to-Everything — взаимодействие «транспортное средство-ко-всему»
VIO	Visual-Inertial Odometry — визуально-инерциальная одометрия
VR	Vehicle Related — система координат, связанная с транспортным средством
YOLO	You Only Look Once — семейство моделей детекции объектов
mAP	mean Average Precision — средняя точность детекции

Словарь терминов

3D-детекция : Задача компьютерного зрения по обнаружению объектов на изображении или в 3D-сцене и определению их положения в трехмерном пространстве, как правило, в виде ограничивающего параллелепипеда.

Байесовская оценка : Статистический метод, основанный на теореме Байеса, который позволяет рекурсивно обновлять оценку состояния системы по мере поступления новых измерений.

Венгерский алгоритм : Алгоритм комбинаторной оптимизации, который решает задачу о назначениях (например, сопоставление треков и детекций) за полиномиальное время.

Вид сверху (Bird's-Eye-View, BEV) : Проекция трехмерной сцены на горизонтальную плоскость (вид сверху). Часто используется в системах автономного вождения для объединения данных с разных сенсоров и планирования траектории.

Визуально-инерциальная одометрия (VIO) : Технология определения положения и ориентации объекта в пространстве путем совместной обработки данных с камер и инерциальных датчиков (IMU).

Гомография : В компьютерном зрении — проективное преобразование, описывающее, как точки на одной плоскости отображаются на другую плоскость. Используется для сопоставления изображений с разных ракурсов.

Дрейф эмбединга : Проблема в трекинге на основе визуальных признаков, когда векторное представление (эмбединг) объекта постепенно изменяется из-за смены освещения, ракурса или других факторов, что может привести к потере отслеживаемого объекта.

Задача о назначениях : Классическая задача комбинаторной оптимизации, заключающаяся в поиске оптимального сопоставления (с минимальной суммарной стоимостью) между элементами двух множеств.

ИТС (Интеллектуальные транспортные системы) : Комплекс взаимосвязанных автоматизированных систем, предназначенных для управления транспортными потоками и обеспечения безопасности дорожного движения.

Каскадное сопоставление : Стратегия ассоциации данных в трекинге, которая последовательно применяет несколько этапов сопоставления с использованием разных метрик для повышения надежности.

Классификация : В контексте данной работы — процесс определения класса сигнала светофора (например, красный, зеленый, желтый).

Ковариация шума : В фильтре Калмана — матрица, характеризующая неопределенность (шум). Различают ковариацию шума процесса (модели движения) и ковариацию шума измерений (сенсора).

Компьютерное зрение : Теория и технология создания машин, которые могут производить обнаружение, отслеживание и классификацию объектов.

Конечный автомат : Математическая модель, описывающая систему с конечным числом состояний и переходами между ними. Используется для управления логикой, например, для стабильной классификации сигналов светофора.

Косинусное расстояние : Мера, используемая для определения схожести между двумя ненулевыми векторами в многомерном пространстве. Вычисляется как 1 минус косинусное сходство.

Краудсорсинг : Привлечение широкого круга лиц для решения каких-либо проблем, в данном контексте — использование данных, собранных парком транспортных средств, для обновления цифровых карт.

Кубоид : Трехмерный параллелепипед (3D bounding box), используемый для описания положения, размеров и ориентации объекта в 3D-пространстве.

Матрица несоответствий (Confusion Matrix) : Таблица, позволяющая визуализировать производительность алгоритма классификации. Каждая строка представляет экземпляры в прогнозируемом классе, а каждый столбец — экземпляры в фактическом классе.

Матрица Якоби (Якобиан) : Матрица, составленная из всех частных производных первого порядка вектор-функции. Используется в расширенном фильтре Калмана для линеаризации нелинейных моделей.

Модель движения : Математическая модель, описывающая, как состояние объекта (например, положение и скорость) изменяется с течением времени.

Ошибка I рода (Ложноположительное срабатывание, FP) : Ошибка, при которой гипотеза ошибочно принимается. В контексте данной работы — система принимает решение об остановке, когда это не требуется (например, на разрешающий сигнал светофора).

Ошибка II рода (Ложноотрицательное срабатывание, FN) : Ошибка, при которой гипотеза ошибочно отвергается. В контексте данной работы — система принимает решение о продолжении движения, когда требуется остановка (например, на запрещающий сигнал). Является наиболее критичной ошибкой с точки зрения безопасности.

Плотность вероятности (ПВ) : Функция, описывающая относительную вероятность того, что случайная величина примет определенное значение. В Байесовской фильтрации различают априорную (до измерения) и апостериорную (после измерения) ПВ.

Полнота (Recall) : Метрика, показывающая долю реальных объектов заданного класса, которые система смогла правильно обнаружить. Вычисляется как отношение истинно положительных срабатываний (TP) к сумме истинно положительных и ложноотрицательных (FN).

Радар (Radio Detection and Ranging) : Радиолокатор — система для обнаружения воздушных, морских и наземных объектов, а также для определения их дальности и скорости.

Расстояние Махаланобиса : Мера расстояния между точкой и распределением вероятностей, которая учитывает корреляции между переменными. Используется для сопоставления объектов с учетом неопределенности их состояний.

Расширенный фильтр Калмана (EKF) : Модификация фильтра Калмана для работы с нелинейными системами, которая использует линеаризацию модели с помощью разложения в ряд Тейлора первого порядка.

Ре-идентификация (Re-ID) : Задача компьютерного зрения, заключающаяся в повторной идентификации одного и того же объекта на изображениях, полученных с разных, непересекающихся камер или в разное время.

Робастность : Способность алгоритма или системы сохранять свою эффективность при изменении входных данных или условий эксплуатации, таких как погодные условия, освещение или наличие помех.

Семантическая карта : Цифровая карта, обогащенная смысловой информацией об объектах (например, тип светофора, разрешенные маневры), а не только их точными геометрическими координатами.

Сигма-точечный фильтр Калмана (UKF) : Нелинейная модификация фильтра Калмана, которая аппроксимирует распределение вероятностей с помощью детерминированно выбранного набора точек (сигма-точек), что позволяет избежать аналитической линеаризации и повысить точность.

СНС (Сверточная нейронная сеть) : Специальная архитектура искусственных нейронных сетей, нацеленная на эффективное распознавание образов, ставшая де-факто стандартом в различных задачах компьютерного зрения.

Сопоставление (Ассоциация данных) : Процесс установления соответствия между различными наборами данных. В контексте трекинга — сопоставление существующих треков с новыми детекциями. В контексте многокамерной системы — сопоставление детекций одного объекта с разных камер.

”Сырые”измерения : Первичные данные, получаемые непосредственно с сенсора без какой-либо предварительной обработки (например, пиксельные координаты с камеры или псевдодальности со спутникового приемника).

Тесносвязанная схема комплексирования : Подход к объединению данных от нескольких сенсоров, при котором фильтр оптимальной оценки

работает непосредственно с ”сырыми” измерениями от всех датчиков, в отличие от слабосвязанной схемы, где данные предварительно обрабатываются независимо.

Теневой режим (Shadow Mode) : Режим работы системы помощи водителю, при котором она функционирует параллельно с водителем, принимая и регистрируя собственные решения, но не вмешиваясь в управление транспортным средством. Используется для безопасной оценки и отладки.

Точность (Precision) : Метрика, показывающая, какая доля объектов, обнаруженных системой как принадлежащие к заданному классу, действительно к нему относится. Вычисляется как отношение истинно положительных срабатываний (TP) к сумме истинно положительных и ложноположительных (FP).

Трекинг (Отслеживание объектов, MOT) : Процесс анализа видеопоследовательности для идентификации и отслеживания движущихся объектов во времени. Включает в себя присвоение уникальных идентификаторов объектам и построение их траекторий.

Трекинг по детекциям (Tracking-by-Detection) : Доминирующая парадигма в отслеживании объектов, разделяющая задачу на два этапа: обнаружение объектов в каждом кадре и их последующее связывание во времени.

Треклет : Частичная траектория объекта, построенная в результате отслеживания на коротком отрезке времени.

ТС (Транспортное средство) : Устройство, предназначенное для перевозки по дорогам людей, грузов или оборудования, установленного на нём.

Управление жизненным циклом треков : Набор правил в алгоритмах трекинга, отвечающих за инициализацию новых треков, их удаление и обработку пропусков детекций.

Фактор-граф : Тип вероятностной графической модели, используемый для представления факторизации функции. Применяется в задачах одновременной локализации и построения карты (SLAM).

Фильтр Калмана : Рекурсивный алгоритм, который оценивает состояние динамической системы по серии неполных и зашумленных измерений. Широко

используется для предсказания движения и сглаживания траекторий в задачах трекинга.

Фильтр частиц (PF) : Метод последовательного Монте-Карло, который аппроксимирует плотность вероятности набором взвешенных случайных выборок (частиц). Способен работать со сложными нелинейными и негауссовыми системами.

Фильтрация : В контексте данной работы — процесс рекурсивной оценки состояния (например, координат) объекта по нескольким последовательным измерениям для сглаживания шумов и получения более точной оценки.

Эмбединг : Векторное представление объекта (например, области изображения со светофором) в пространстве признаков, полученное с помощью нейронной сети. Используется для идентификации и сравнения объектов.

Эталонная разметка (Ground Truth) : Набор данных, считающийся абсолютно истинным и используемый для обучения и оценки моделей машинного обучения. Включает в себя правильные классы и координаты объектов.

ADAS (Advanced Driver-Assistance Systems) : Системы помощи водителю — электронные системы, помогающие водителю в процессе управления транспортным средством.

ASIL (Automotive Safety Integrity Level) : Уровень полноты безопасности автомобиля — классификация рисков, определённая в стандарте ISO 26262 для функциональной безопасности автомобильных систем.

C-V2X (Cellular Vehicle-to-Everything) : Технология сотовой связи для взаимодействия транспортного средства со всем окружением, использующая мобильные сети.

DeepSORT : Расширение алгоритма отслеживания SORT, которое дополнительно использует признаки внешнего вида, извлеченные глубокой нейронной сетью (эмбединги), для более надежного сопоставления объектов после окклюзий.

DSRC (Dedicated Short-Range Communications) : Технология беспроводной связи малого радиуса действия, предназначенная для автомобильных приложений.

F1-мера (F1-Score) : Гармоническое среднее между точностью и полнотой, предоставляющее сбалансированную оценку производительности модели, особенно в случаях несбалансированных классов.

Focal Loss : Модифицированная функция потерь для задачи классификации, которая позволяет фокусировать обучение на "сложных" примерах, уменьшая вес "легких правильно классифицируемых примеров. Эффективна при сильном дисбалансе классов.

FOV (Field of View) : Поле зрения — угловое пространство, видимое камерой или другим сенсором.

GAN (Generative Adversarial Network) : Генеративно-сопоставительная сеть — алгоритм машинного обучения без учителя, построенный на комбинации двух нейронных сетей.

GLOSA (Green Light Optimized Speed Advisory) : Система информирования об оптимальной скорости для проезда перекрестка на зеленый сигнал светофора.

GM (Global Metric) : Глобальная метрическая система координат, используемая для определения абсолютного положения объектов в пространстве.

GNSS/INS : Интегрированная навигационная система, объединяющая данные от Глобальной навигационной спутниковой системы (GNSS) и инерциального измерительного блока (INS) для обеспечения точного и непрерывного определения местоположения и ориентации.

GPS (Global Positioning System) : Глобальная система позиционирования — спутниковая система навигации, обеспечивающая измерение расстояния, времени и определяющая местоположение.

HD-карта (High-Definition Map) : Высокоточная цифровая карта — детализированное многослойное представление дорожной среды с сантиметровой точностью.

ID Switch : В задачах отслеживания объектов — ошибка, при которой одному и тому же объекту присваиваются разные идентификаторы в разные моменты времени.

IMU (Inertial Measurement Unit) : Инерциальный измерительный модуль — устройство для измерения линейного ускорения и угловой скорости транспортного средства.

IoU (Intersection over Union) : Метрика ”пересечение над объединением” — метрика для оценки точности детекции объектов, вычисляющая степень пересечения предсказанной и истинной ограничивающих рамок.

IoU 3D : Метрика ”пересечение над объединением” для трехмерных ограничивающих рамок (кубоидов), вычисляющая отношение объема их пересечения к объему их объединения.

ISO 26262 : Международный стандарт функциональной безопасности электрических и электронных систем, устанавливаемых в серийные легковые автомобили.

LED (Light-Emitting Diode) : Светодиод — полупроводниковый прибор, создающий оптическое излучение при пропускании через него электрического тока.

Lidar (Light Detection and Ranging) : Лидар — технология получения и обработки информации об удаленных объектах с помощью активных оптических систем, использующих явления отражения света.

mAP (mean Average Precision) : Средняя точность — метрика оценки качества работы моделей детектирования объектов, усредненная по всем классам и уровням отсечки.

MOTA (Multiple Object Tracking Accuracy) : Комплексная метрика для оценки качества отслеживания нескольких объектов, которая учитывает три типа ошибок: ложноположительные срабатывания (FP), ложноотрицательные срабатывания (FN) и ошибки смены идентификатора (ID Switch).

NMS (Non-Maximum Suppression) : Подавление немаксимумов — алгоритм постобработки в задачах детектирования для устранения избыточных,

сильно перекрывающихся ограничивающих рамок, относящихся к одному и тому же объекту.

RegNet : Архитектура сверточных нейронных сетей, разработанная на основе систематического анализа пространства проектных решений, что обеспечивает высокую эффективность и масштабируемость.

RGB-камера : Камера, формирующая цветное изображение на основе трех основных цветов: красного (Red), зеленого (Green) и синего (Blue).

SAE J3016 : Стандарт, разработанный Обществом автомобильных инженеров (SAE), определяющий шесть уровней автоматизации вождения (от 0 до 5).

Sim2Real (Simulation to Reality) : Подход в робототехнике и машинном обучении, заключающийся в обучении модели в симулированной среде с последующим переносом и адаптацией для работы в реальном мире.

SPaT (Signal Phase and Timing) : Сообщения о фазе и времени сигнала — данные, передаваемые от светофорной инфраструктуры, содержащие информацию о текущей и будущей фазах светофорных сигналов.

SSD (Single Shot Detector) : Нейросетевая архитектура для детектирования объектов, которая выполняет предсказания за один проход.

V2I (Vehicle-to-Infrastructure) : Взаимодействие транспортного средства с дорожной инфраструктурой.

V2N (Vehicle-to-Network) : Взаимодействие транспортного средства с сетью (например, облачными сервисами).

V2P (Vehicle-to-Pedestrian) : Взаимодействие транспортного средства с пешеходами.

V2V (Vehicle-to-Vehicle) : Взаимодействие транспортного средства с другим транспортным средством.

V2X (Vehicle-to-Everything) : Комплексная технология, позволяющая транспортным средствам обмениваться информацией с любыми другими объектами, которые могут на них повлиять.

VR (Vehicle Related) : Система координат, связанная с транспортным средством (в данном случае — с трамваем), с началом отсчета в определенной точке на корпусе.

YOLO (You Only Look Once) : Семейство нейросетевых моделей для детектирования объектов в реальном времени, выполняющих предсказания за один проход по изображению.

Список литературы

1. *World Health Organization*. Global Status Report on Road Safety 2018 [Текст] / World Health Organization ; World Health Organization. — Geneva, 2018. — URL: <https://www.who.int/publications/i/item/9789241565684> ; WHO Press.
2. *Gautam, S.* Image-based automatic traffic lights detection system for autonomous cars: a review [Текст] / S. Gautam, A. Kumar // *Multimedia Tools and Applications*. — 2023. — DOI: [10.1007/s11042-023-14340-1](https://doi.org/10.1007/s11042-023-14340-1).
3. *Yang, L.* Deep Traffic Light Perception with Spatiotemporal Analysis for Autonomous Driving [Текст] / L. Yang, X. Wang, J. Wang // 2020 European Control Conference (ECC). — 2020. — DOI: [10.23919/ecc51009.2020.9143756](https://doi.org/10.23919/ecc51009.2020.9143756).
4. *Montero, D.* Utah officials celebrate 100th anniversary of traffic signal [Текст] / D. Montero // *The Salt Lake Tribune*. — 2012. — Окт. — URL: <https://archive.slttrib.com/story.php?ref=/slttrib/politics/55027680-90/1912-green-invention-lake.html.csp> ; Transportation: Salt Lake City police officer gets credit — but not patent — for his 1912 creation.
5. *United Nations Economic Commission for Europe*. Convention on Road Signs and Signals (Vienna Convention) [Текст] / United Nations Economic Commission for Europe. — 1968. — URL: https://treaties.un.org/Pages/ViewDetailsIII.aspx?src=TREATY&mtdsg_no=XI-B-20&chapter=11 ; Done at Vienna on 8 November 1968. United Nations Treaty Collection.
6. *Zuñiga, J. M.* International Effort Toward Uniformity on Road Traffic Signs, Signals, and Markings [Текст] / J. M. Zuñiga // *Highway Research Record*. — 1969. — Т. 299. — С. 1—23. — Analysis of international differences in traffic light standards and Vienna Convention implementation.
7. *Kulkarni, R.* Traffic Light Detection and Recognition for Self Driving Cars Using Deep Learning [Текст] / R. Kulkarni, S. Dhavalikar, S. Bangar // 2018

- Fourth International Conference on Computing Communication Control and Automation (ICCUBEA). — 2018. — DOI: [10.1109/ICCUBEA.2018.8697819](https://doi.org/10.1109/ICCUBEA.2018.8697819).
8. *Vitas, D.* Traffic Light Detection in Autonomous Driving Systems [Текст] / D. Vitas, M. Tomic, M. Burul // IEEE Consumer Electronics Magazine. — 2020. — DOI: [10.1109/MCE.2020.2969156](https://doi.org/10.1109/MCE.2020.2969156).
 9. Traffic Light Detection by Integrating Feature Fusion and Attention Mechanism [Текст] / С.-Н. Chuang [и др.] // Electronics. — 2023. — Т. 12, № 17. — С. 3727. — DOI: [10.3390/electronics12173727](https://doi.org/10.3390/electronics12173727).
 10. YOLO-LLTS: Real-Time Low-Light Traffic Sign Detection via Prior-Guided Enhancement and Multi-Branch Feature Interaction [Текст] / Z. Lin [и др.]. — 2025. — DOI: [10.48550/arXiv.2503.13883](https://doi.org/10.48550/arXiv.2503.13883). — arXiv: [2503.13883](https://arxiv.org/abs/2503.13883) [cs.cv]. — URL: <https://arxiv.org/abs/2503.13883>.
 11. Video-based Traffic Light Recognition by Rockchip RV1126 for Autonomous Driving [Текст] / M. Fan [и др.] //. — 2025. — DOI: [10.48550/arXiv.2503.23965](https://doi.org/10.48550/arXiv.2503.23965). — arXiv: [2503.23965](https://arxiv.org/abs/2503.23965) [cs.cv]. — URL: <https://arxiv.org/abs/2503.23965>.
 12. SqueezeDet-Based Nighttime Traffic Light Detection with Filtering Rules [Текст] / Y. Huo [и др.] // 2019 2nd China Symposium on Cognitive Computing and Hybrid Intelligence (CCHI). — 2019. — DOI: [10.1109/CCHI.2019.8901919](https://doi.org/10.1109/CCHI.2019.8901919).
 13. *Hassan, E. T.* Semantic Consistency: The Key to Improve Traffic Light Detection with Data Augmentation [Текст] / E. T. Hassan, N. Li, L. Ren // 2020 IEEE Intelligent Vehicles Symposium (IV). — 2020. — DOI: [10.1109/IV47402.2020.9304653](https://doi.org/10.1109/IV47402.2020.9304653).
 14. Traffic Light Recognition Using Deep Learning and Prior Maps for Autonomous Cars [Текст] / L. C. Possatti [и др.] // 2019 International Joint Conference on Neural Networks (IJCNN). — 2019. — DOI: [10.1109/IJCNN.2019.8851927](https://doi.org/10.1109/IJCNN.2019.8851927).

15. TLD-READY: Traffic Light Detection - Relevance Estimation and Deployment Analysis [Текст] / N. Polley [и др.] // ArXiv. — 2024. — DOI: [10.48550/arXiv.2409.07284](https://doi.org/10.48550/arXiv.2409.07284).
16. *De Guia, J.* Development of Traffic Light and Road Sign Detection and Recognition Using Deep Learning [Текст] / J. De Guia, P. Deveraj // International Journal of Advanced Computer Science and Applications. — 2024. — Т. 15, № 1. — С. 95—104. — DOI: [10.14569/IJACSA.2024.0151095](https://doi.org/10.14569/IJACSA.2024.0151095).
17. Deep Traffic Light Detection for Self-driving Cars from a Large-scale Dataset [Текст] / J. Kim [и др.] // 2018 21st International Conference on Intelligent Transportation Systems (ITSC). — 2018. — DOI: [10.1109/ITSC.2018.8569575](https://doi.org/10.1109/ITSC.2018.8569575).
18. Road vehicles – Functional safety [Текст]. — International Organization for Standardization, 2018. — URL: <https://www.iso.org/standard/68383.html> ; Parts 1-12.
19. Автотранспортные средства. Функциональная безопасность. Часть 1. Термины и определения [Текст]. — Росстандарт, 2020. — Адаптированный международный стандарт ISO 26262.
20. Дорожные транспортные средства. Функциональная безопасность. Часть 2. Управление функциональной безопасностью [Текст]. — Росстандарт, 2014. — Адаптированный международный стандарт ISO 26262.
21. Дорожные транспортные средства. Функциональная безопасность. Часть 5. Разработка аппаратных средств [Текст]. — Росстандарт, 2021. — Адаптированный международный стандарт ISO 26262.
22. *Schildbach, G.* On the Application of ISO 26262 in Control Design for Automated Vehicles [Текст] / G. Schildbach // arXiv preprint arXiv:1804.04349. — 2018. — URL: <https://arxiv.org/pdf/1804.04349.pdf>.
23. Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles [Текст]. — SAE International, 04.2021. — Standard J3016:2021.

24. *Ronchi, L.* Development, Testing, and Validation of ADAS L2/L3 Systems: A KPI-Based Methodology [Текст] / L. Ronchi, C. Annicchiarico, R. Capitani // Engineering Proceedings. — 2025. — Т. 85, № 1. — С. 38. — DOI: [10.3390/engproc2025085038](https://doi.org/10.3390/engproc2025085038). — URL: <https://www.mdpi.com/2673-4591/85/1/38>.
25. *Synopsys.* The 6 Levels of Vehicle Autonomy Explained [Текст] / Synopsys. — 2025. — URL: <https://www.synopsys.com/blogs/chip-design/autonomous-driving-levels.html> ; Accessed on July 29, 2024. Synopsys Automotive.
26. *Neemeh, S.* What is ISO 26262 Functional Safety in Transport Vehicles? [Текст] / S. Neemeh. — 05.2024. — URL: <https://www.lhpes.com/blog/what-is-iso-26262-functional-safety-in-transport-vehicles>. LHP Engineering Solutions Blog.
27. *Yang, D.* Traffic Light Detection using Attention-Guided Continuous Conditional Random Fields [Текст] / D. Yang, S.-W. Seo // 2022 International Conference on Electronics, Information, and Communication (ICEIC). — 2022. — DOI: [10.1109/iceic54506.2022.9748468](https://doi.org/10.1109/iceic54506.2022.9748468).
28. Traffic Object Detection and Recognition Based on the Attentional Visual Field of Drivers [Текст] / M. Shirpour [и др.] // IEEE Transactions on Intelligent Vehicles. — 2023. — Т. 8. — С. 594—604. — DOI: [10.1109/TIV.2021.3133849](https://doi.org/10.1109/TIV.2021.3133849).
29. Real-Time Traffic Sign and Signal Recognition for AI-Based Vehicular Control [Текст] / A. S. Utane [и др.] // Advances in Nonlinear Variational Inequalities. — 2025. — DOI: [10.52783/anvi.v28.3656](https://doi.org/10.52783/anvi.v28.3656).
30. *Siemens Mobility.* The Autonomous Siemens Tram. On the way to smart and autonomous trams: A Siemens Mobility research project [Текст] / Siemens Mobility. — 2019. — URL: <https://assets.new.siemens.com/siemens/assets/api/uuid:bc2811c4-3d26-460d-9472-9372d5ce32d7/autonomous-tram.pdf> (дата обр. 21.07.2024).
31. *Khanmohamadi, M.* Smart Intersections and Connected Autonomous Vehicles for Sustainable Smart Cities: A Brief Review [Текст] / M. Khanmohamadi,

- M. Guerrieri // Sustainability. — 2025. — Т. 17, № 7. — С. 3254. — DOI: [10.3390/su17073254](https://doi.org/10.3390/su17073254). — URL: <https://www.mdpi.com/2071-1050/17/7/3254>.
32. Estimating Phase Duration for SPaT Messages [Текст] / S. Ibrahim [и др.] // arXiv preprint arXiv:1710.05394. — 2017. — URL: <https://arxiv.org/abs/1710.05394v2> ; arXiv:1710.05394v2 [cs.SY].
 33. C-V2X Communications for the Support of a Green Light Optimized Speed Advisory (GLOSA) Use Case [Текст] / I. P. Chochliouros [и др.] // Journal of Telecommunications and Information Technology. — 2021. — № 2. — С. 93—106. — DOI: [10.26636/jtit.2021.152321](https://doi.org/10.26636/jtit.2021.152321). — URL: <https://jtit.pl/jtit/article/view/467>.
 34. A Review of Research on Intersection Control Based on Connected Vehicles and Data-Driven Intelligent Approaches [Текст] / K. Gao [и др.] // Electronics. — 2020. — Т. 9, № 6. — С. 885. — DOI: [10.3390/electronics9060885](https://doi.org/10.3390/electronics9060885). — URL: <https://www.mdpi.com/2079-9292/9/6/885>.
 35. *СМАРТС*. Оценка стоимости развертывания инфраструктуры V2X в России [Текст] / *СМАРТС*. — 2024. — URL: <https://ru.wikipedia.org/wiki/%D0%A1%D0%9C%D0%90%D0%A0%D0%A2%D0%A1> ; Дата доступа: 2024-07-30. Данные приведены в статье Wikipedia о компании.
 36. Mande, S. A Comprehensive Survey on Challenges and Issues in V2X and V2V Communication in 6G Future Generation Communication Models [Текст] / S. Mande, N. Ramachandran // Information Sciences Letters. — 2024. — Т. 13, № 3. — С. 107—116. — DOI: [10.18280/isi.290315](https://doi.org/10.18280/isi.290315). — URL: <https://doi.org/10.18280/isi.290315>.
 37. *CPDEVICE*. What is V2X (Vehicle to Everything)? [Текст] / *CPDEVICE*. — 2024. — URL: <https://ru.cpdevice.com/about-v2x/> ; Дата доступа: 2024-07-30. Technical article.
 38. A Tutorial on 5G NR V2X Communications [Текст] / М. Н. С. Garcia [и др.] // IEEE Communications Surveys & Tutorials. — 2021. — Т. 23, № 3. —

- C. 1972—2026. — DOI: [10.1109/comst.2021.3057017](https://doi.org/10.1109/comst.2021.3057017). — URL: <http://dx.doi.org/10.1109/COMST.2021.3057017>.
39. *Asrat, K. T. A Comprehensive Survey on High-Definition Map Generation and Maintenance [Текст] / K. T. Asrat, H.-J. Cho // ISPRS International Journal of Geo-Information. — 2024. — Т. 13, № 7. — С. 232. — DOI: [10.3390/ijgi13070232](https://doi.org/10.3390/ijgi13070232). — URL: <https://www.mdpi.com/2220-9964/13/7/232>.*
 40. *Lanelet2: A high-definition map framework for the future of automated driving [Текст] / F. Poggenschans [и др.] // 2018 21st International Conference on Intelligent Transportation Systems (ITSC). — 2018. — С. 1672—1679. — DOI: [10.1109/ITSC.2018.8569929](https://doi.org/10.1109/ITSC.2018.8569929).*
 41. *Requirements for Automotive LiDAR Systems [Текст] / Z. Dai [и др.] // Sensors. — 2022. — Т. 22, № 19. — С. 7532. — DOI: [10.3390/s22197532](https://doi.org/10.3390/s22197532). — URL: <https://www.mdpi.com/1424-8220/22/19/7532>.*
 42. *Li, Y. LiDAR for autonomous driving: the principles, challenges, and trends for automotive LiDAR and perception systems [Текст] / Y. Li, J. Ibanez-Guzman // IEEE Signal Processing Magazine. — 2020. — Т. 37, № 4. — С. 50—61. — DOI: [10.1109/MSP.2020.2973615](https://doi.org/10.1109/MSP.2020.2973615).*
 43. *Sensor and Sensor Fusion Technology in Autonomous Vehicles: A Review [Текст] / D. J. Yeong [и др.] // Sensors. — 2021. — Т. 21, № 6. — DOI: [10.3390/s21062140](https://doi.org/10.3390/s21062140). — URL: <https://www.mdpi.com/1424-8220/21/6/2140>.*
 44. *Comparison Review on LiDAR vs Camera in Autonomous Vehicle [Текст] / N. Mugunthan [и др.] // International Research Journal of Engineering and Technology (IRJET). — 2020. — Т. 7, № 8. — URL: <https://www.irjet.net/archives/V7/i8/IRJET-V7I8731.pdf> ; e-ISSN: 2395-0056, p-ISSN: 2395-0072.*
 45. *Monninger, T. Semantic Map Learning of Traffic Light to Lane Assignment based on Motion Data [Текст] / T. Monninger, A. Weber, S. Staab. — 2023. — DOI: [10.48550/arXiv.2309.14793](https://doi.org/10.48550/arXiv.2309.14793). — arXiv: [2309.14793](https://arxiv.org/abs/2309.14793) [cs.CV].*

46. The ATLAS of Traffic Lights: A Reliable Perception Framework for Autonomous Driving [Текст] / R. Polley [и др.]. — 04.2025. — DOI: [10 . 48550/arXiv.2504.19722](https://doi.org/10.48550/arXiv.2504.19722).
47. *Mentasti, S.* Traffic lights detection and tracking for HD map creation [Текст] / S. Mentasti, Y. C. Simsek, M. Matteucci // *Frontiers in Robotics and AI*. — 2023. — T. 10. — C. 1065394. — DOI: [10.3389/frobt.2023.1065394](https://doi.org/10.3389/frobt.2023.1065394).
48. Surround-view Fisheye Camera Perception for Automated Driving: Overview, Survey and Challenges [Текст] / V. R. Kumar [и др.]. — 2023. — arXiv: [2205. 13281 \[cs.CV\]](https://arxiv.org/abs/2205.13281). — URL: <https://arxiv.org/abs/2205.13281>.
49. *Zaremba, A.* Distance Estimation with a Stereo Camera and Accuracy Determination [Текст] / A. Zaremba, S. Nitkiewicz // *Applied Sciences*. — 2024. — T. 14, № 23. — DOI: [10 . 3390 / app142311444](https://doi.org/10.3390/app142311444). — URL: [https : //www.mdpi.com/2076-3417/14/23/11444](https://www.mdpi.com/2076-3417/14/23/11444).
50. *Luxonis.* Depth Accuracy [Текст] / Luxonis. — 10.2023. — URL: <https://docs.luxonis.com/hardware/platform/depth/depth-accuracy/> ; Accessed: 2024-07-15.
51. aUToLights: A Robust Multi-Camera Traffic Light Detection and Tracking System [Текст] / S. Wu [и др.]. — 2023. — arXiv: [2305.08673 \[cs.R0\]](https://arxiv.org/abs/2305.08673). — URL: <https://arxiv.org/abs/2305.08673>.
52. *Herman, S.* Single Camera Object Detection for Self-Driving Vehicle: A Review [Текст] / S. Herman, K. Ismail // *Journal of the Society of Automotive Engineers Malaysia*. — 2017. — T. 1, № 1. — C. 59—70. — URL: [https : / / pdfs . semanticscholar.org/9378/4c7887bd6d1e64c73c9ff88f4d06e0123de1.pdf](https://pdfs.semanticscholar.org/9378/4c7887bd6d1e64c73c9ff88f4d06e0123de1.pdf).
53. Traffic Light Recognition using Convolutional Neural Networks: A Survey [Текст] / S. Pavlitska [и др.] // arXiv preprint arXiv:2309.02158. — 2023. — arXiv: [2309.02158](https://arxiv.org/abs/2309.02158). — URL: <https://arxiv.org/abs/2309.02158>.
54. *Fukuchi, T.* Design and Optimization of a Deep Neural Network Architecture for Traffic Light Detection [Текст] / T. Fukuchi, M. O. Ikechukwu, A. Ben Abdallah // *SHS Web of Conferences*. T. 77. — EDP Sciences. 2020. —

- C. 01002. — DOI: [10.1051/shsconf/20207701002](https://doi.org/10.1051/shsconf/20207701002). — URL: https://www.shs-conferences.org/articles/shsconf/pdf/2020/05/shsconf_etltc2020_01002.pdf.
55. *Гужва Н.С.* Алгоритмы идентификации светофоров в системах помощи водителю трамвая [Текст] / Гужва Н.С., Садеков Р.Н. // Журнал “Гироскопия и навигация”. — 2024. — Т. 32, № 3. — С. 47—65.
 56. Multiple Object Tracking: A Literature Review [Текст] / W. Luo [и др.] // Artificial Intelligence. — 2021. — Т. 293. — С. 103448. — DOI: [10.1016/j.artint.2020.103448](https://doi.org/10.1016/j.artint.2020.103448). — arXiv: [1409.7618v5](https://arxiv.org/abs/1409.7618v5) [cs.CV].
 57. *Kalman, R. E.* A New Approach to Linear Filtering and Prediction Problems [Текст] / R. E. Kalman // Transactions of the ASME, Journal of Basic Engineering. — 1960. — Т. 82, № 1. — С. 35—45. — DOI: [10.1115/1.3662552](https://doi.org/10.1115/1.3662552).
 58. *Bergmann, P.* Tracking without bells and whistles [Текст] / P. Bergmann, T. Meinhardt, L. Leal-Taixé // Proceedings of the IEEE/CVF International Conference on Computer Vision. — 2019. — С. 941—951. — DOI: [10.1109/ICCV.2019.00103](https://doi.org/10.1109/ICCV.2019.00103).
 59. *Zhou, X.* Tracking objects as points [Текст] / X. Zhou, V. Koltun, P. Krähenbühl // European Conference on Computer Vision. — Springer. 2020. — С. 474—490. — DOI: [10.48550/arXiv.2004.01177](https://doi.org/10.48550/arXiv.2004.01177).
 60. Deep learning in video multi-object tracking: A survey [Текст] / G. Ciaparrone [и др.] // Neurocomputing. — 2020. — Т. 381. — С. 61—88. — DOI: [10.1016/j.neucom.2019.11.023](https://doi.org/10.1016/j.neucom.2019.11.023).
 61. Simple Online and Realtime Tracking [Текст] / A. Bewley [и др.] // 2016 IEEE International Conference on Image Processing (ICIP). — 2016. — С. 3464—3468. — DOI: [10.1109/ICIP.2016.7533003](https://doi.org/10.1109/ICIP.2016.7533003). — arXiv: [1602.00763](https://arxiv.org/abs/1602.00763).
 62. *Kuhn, H. W.* The Hungarian method for the assignment problem [Текст] / H. W. Kuhn // Naval Research Logistics Quarterly. — 1955. — Т. 2, № 1/2. — С. 83—97. — DOI: [10.1002/nav.3800020109](https://doi.org/10.1002/nav.3800020109).

63. *Wojke, N.* Simple Online and Realtime Tracking with a Deep Association Metric [Текст] / N. Wojke, A. Bewley, D. Paulus // 2017 IEEE International Conference on Image Processing (ICIP). — 2017. — С. 3645—3649. — DOI: [10.1109/ICIP.2017.8296962](https://doi.org/10.1109/ICIP.2017.8296962). — arXiv: [1703.07402](https://arxiv.org/abs/1703.07402).
64. A Review of Deep Learning-Based Visual Multi-Object Tracking Algorithms for Autonomous Driving [Текст] / S. Guo [и др.] // Applied Sciences. — 2022. — Т. 12, № 21. — DOI: [10.3390/app122110741](https://doi.org/10.3390/app122110741). — URL: <https://www.mdpi.com/2076-3417/12/21/10741>.
65. Multi-Camera Multiple 3D Object Tracking on the Move for Autonomous Vehicles [Текст] / P. Nguyen [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. — 06.2022. — С. 2221—2230. — DOI: [10.48550/arXiv.2204.09151](https://doi.org/10.48550/arXiv.2204.09151).
66. TrackFormer: Multi-Object Tracking with Transformers [Текст] / T. Meinhardt [и др.] // The IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 06.2022. — DOI: [10.48550/arXiv.2101.02702](https://doi.org/10.48550/arXiv.2101.02702).
67. DEFT: Detection Embeddings for Tracking [Текст] / M. Chaabane [и др.] //. — 2021. — DOI: [10.48550/arXiv.2102.02267](https://doi.org/10.48550/arXiv.2102.02267).
68. Monocular quasi-dense 3d object tracking [Текст] / H.-N. Hu [и др.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2022. — DOI: [10.48550/arXiv.2103.07351](https://doi.org/10.48550/arXiv.2103.07351).
69. Online Multi-camera People Tracking with Spatial-temporal Mechanism and Anchor-feature Hierarchical Clustering [Текст] / R. Cherdchusakulchai [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). — 2024. — DOI: [10.1109/CVPRW63382.2024.00715](https://doi.org/10.1109/CVPRW63382.2024.00715).
70. Standing Between Past and Future: Spatio-Temporal Modeling for Multi-Camera 3D Multi-Object Tracking [Текст] / Z. Pang [и др.] // Proceedings of the

- IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2023. — C. 9203—9213. — DOI: [10.48550/arXiv.2302.03802](https://doi.org/10.48550/arXiv.2302.03802).
71. Asynchronous Multi-Sensor Fusion Multi-Target Tracking Method [Текст] / S. Liu [и др.] // 2018 IEEE 14th International Conference on Control and Automation (ICCA). — 2018. — С. 459—463. — DOI: [10.1109/ICCA.2018.8444300](https://doi.org/10.1109/ICCA.2018.8444300).
 72. *Kaygusuz, N.* AFT-VO: Asynchronous Fusion Transformers for Multi-View Visual Odometry Estimation [Текст] / N. Kaygusuz, O. Mendez, R. Bowden. — 2022. — arXiv: [2206.12946 \[cs.CV\]](https://arxiv.org/abs/2206.12946). — URL: <https://arxiv.org/abs/2206.12946>.
 73. ByteTrack: Multi-Object Tracking by Associating Every Detection Box [Текст] / Y. Zhang [и др.] // arXiv preprint arXiv:2110.06864. — 2021. — DOI: [10.48550/arXiv.2110.06864](https://doi.org/10.48550/arXiv.2110.06864).
 74. FairMOT: On the Fairness of Detection and Re-Identification in Multiple Object Tracking [Текст] / Y. Zhang [и др.] // arXiv preprint arXiv:2004.01888. — 2020. — DOI: [10.1007/s11263-021-01513-4](https://doi.org/10.1007/s11263-021-01513-4).
 75. *Dao, M.-Q.* A two-stage data association approach for 3d multi-object tracking [Текст] / M.-Q. Dao, V. Frémont // Sensors. — 2021. — Т. 21, № 9. — С. 2894. — DOI: [10.3390/s21092894](https://doi.org/10.3390/s21092894).
 76. *Wang, G.* Recent Advances in Embedding Methods for Multi-Object Tracking: A Survey [Текст] / G. Wang, M. Song, J.-N. Hwang. — 2022. — Май. — DOI: [10.48550/arXiv.2205.10766](https://doi.org/10.48550/arXiv.2205.10766).
 77. *Welch, G.* An Introduction to the Kalman Filter [Текст] : тех. отч. / G. Welch, G. Bishop ; University of North Carolina at Chapel Hill, Department of Computer Science. — 1995. — TR 95—041.
 78. Observation-centric sort: Rethinking sort for robust multi-object tracking [Текст] / J. Cao [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2023. — С. 9686—9696. — DOI: [10.48550/arXiv.2203.14360](https://doi.org/10.48550/arXiv.2203.14360).

79. *Aharon, N.* BoT-SORT: Robust Associations Multi-Pedestrian Tracking [Текст] / N. Aharon, R. Orfaig, B.-Z. Bobrovsky. — 2022. — DOI: [10.48550/arXiv.2206.14651](https://doi.org/10.48550/arXiv.2206.14651). — arXiv: [2206.14651](https://arxiv.org/abs/2206.14651) [cs.CV].
80. Distributed multi-camera multi-target association for real-time tracking [Текст] / S. Yang [и др.] // Scientific Reports. — 2022. — Июнь. — Т. 12. — С. 11052. — DOI: [10.1038/s41598-022-15000-4](https://doi.org/10.1038/s41598-022-15000-4).
81. nuScenes: A multimodal dataset for autonomous driving [Текст] / H. Caesar [и др.]. — 2020. — arXiv: [1903.11027](https://arxiv.org/abs/1903.11027) [cs.LG]. — URL: <https://arxiv.org/abs/1903.11027>.
82. CC-3DT: Panoramic 3D Object Tracking via Cross-Camera Fusion [Текст] / T. Fischer [и др.]. — 2022. — DOI: [10.48550/arXiv.2212.01247](https://doi.org/10.48550/arXiv.2212.01247). — arXiv: [2212.01247](https://arxiv.org/abs/2212.01247) [cs.CV]. — URL: <https://arxiv.org/abs/2212.01247>.
83. *Chen, Z.* Bayesian Filtering: From Kalman Filters to Particle Filters, and Beyond [Текст] / Z. Chen // Statistics. — 2003. — Т. 182, № 1. — С. 1—69.
84. *Bar-Shalom, Y.* Estimation with Applications to Tracking and Navigation [Текст] / Y. Bar-Shalom, X. R. Li, T. Kirubarajan. — New York : Wiley-Interscience, 2001. — DOI: [10.1002/0471221279](https://doi.org/10.1002/0471221279).
85. *Wan, E. A.* The unscented Kalman filter for nonlinear estimation [Текст] / E. A. Wan, R. Van Der Merwe // Proceedings of the IEEE 2000 Adaptive Systems for Signal Processing, Communications, and Control Symposium (Cat. No.00EX373). — 2000. — С. 153—158. — DOI: [10.1109/ASSPCC.2000.882463](https://doi.org/10.1109/ASSPCC.2000.882463).
86. *Julier, S. J.* A new extension of the Kalman filter to nonlinear systems [Текст] / S. J. Julier, J. K. Uhlmann // Proc. of AeroSense'97. — 1997. — Т. 3068. — С. 182—193.
87. *Zarchan, P.* Fundamentals of Kalman Filtering: A Practical Approach [Текст] / P. Zarchan, H. Musoff. — 2nd. — American Institute of Aeronautics, Astronautics, 2005. — DOI: [10.2514/4.866777](https://doi.org/10.2514/4.866777).

88. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking [Текст] / M. S. Arulampalam [и др.] // IEEE Transactions on Signal Processing. — 2002. — Т. 50, № 2. — С. 174—188. — DOI: [10.1109/78.978374](https://doi.org/10.1109/78.978374).
89. *Doucet, A.* A tutorial on particle filtering and smoothing: Fifteen years later [Текст] / A. Doucet, A. M. Johansen // Handbook of Nonlinear Filtering / под ред. D. Crisan, B. Rozovski. — Oxford University Press, 2009. — С. 656—704.
90. *Jwo, D.-J.* Implementation and Performance Analysis of Kalman Filters with Consistency Validation [Текст] / D.-J. Jwo, A. Biswal // Mathematics. — 2023. — Т. 11, № 3. — С. 521. — DOI: [10.3390/math11030521](https://doi.org/10.3390/math11030521).
91. *Kim, J.* Unscented Kalman Filter-Aided Long Short-Term Memory Approach for Wind Nowcasting [Текст] / J. Kim, K. Lee // Aerospace. — 2021. — Т. 8, № 9. — С. 236. — DOI: [10.3390/aerospace8090236](https://doi.org/10.3390/aerospace8090236).
92. *Khodarahmi, M.* A Review on Kalman Filter Models [Текст] / M. Khodarahmi, V. Maihami // Archives of Computational Methods in Engineering. — 2023. — Т. 30. — С. 727—747. — DOI: [10.1007/s11831-022-09815-7](https://doi.org/10.1007/s11831-022-09815-7).
93. Modeling of multisensor tightly aided BDS triple-frequency precise point positioning and initial assessments [Текст] / Z. Gao [и др.] // Information Fusion. — 2020. — Т. 55. — С. 184—198. — DOI: [10.1016/j.inffus.2019.08.012](https://doi.org/10.1016/j.inffus.2019.08.012).
94. A Robust Adaptive Cubature Kalman Filter Based on SVD for Dual-Antenna GNSS/MIMU Tightly Coupled Integration [Текст] / C. Pan [и др.] // Remote Sensing. — 2021. — Т. 13, № 10. — С. 1943. — DOI: [10.3390/rs13101943](https://doi.org/10.3390/rs13101943).
95. Robust state and protection-level estimation within tightly coupled GNSS/INS navigation system [Текст] / S. Liu [и др.] // GPS Solutions. — 2023. — Т. 27, № 1. — С. 1—15. — DOI: [10.1007/s10291-023-01447-z](https://doi.org/10.1007/s10291-023-01447-z).
96. An improved Kalman filter algorithm for tightly GNSS/INS integrated navigation system [Текст] / Y. Yuan [и др.] // Mathematical Biosciences and

- Engineering. — 2024. — Т. 21, № 1. — С. 963—983. — DOI: [10.3934/mbe.2024040](https://doi.org/10.3934/mbe.2024040).
97. LVI-SAM: Tightly-coupled Lidar-Visual-Inertial Odometry via Smoothing and Mapping [Текст] / T. Shan [и др.] // 2021 IEEE International Conference on Robotics and Automation (ICRA). — IEEE. 2021. — С. 5692—5698. — DOI: [10.1109/ICRA48506.2021.9561996](https://doi.org/10.1109/ICRA48506.2021.9561996).
 98. Implementation and performance of a deeply coupled GNSS receiver with low-cost MEMS inertial sensors for vehicle urban navigation [Текст] / X. Feng [и др.] // Sensors. — 2020. — Т. 20, № 12. — С. 3397. — DOI: [10.3390/s20123397](https://doi.org/10.3390/s20123397).
 99. Sun, A. A Perception Centered Self-driving System without HD Maps [Текст] / A. Sun // arXiv preprint arXiv:2009.00782. — 2020. — URL: <https://arxiv.org/abs/2009.00782>.
 100. Cho, Y. SIO-Mapper: A Framework for Lane-Level HD Map Construction Using Satellite Images and OpenStreetMap with No On-Site Visits [Текст] / Y. Cho, J.-H. Ryu. — 2025. — arXiv: [2504.09882](https://arxiv.org/abs/2504.09882) [cs.R0]. — URL: <https://arxiv.org/abs/2504.09882>.
 101. Zhang, P. Real-time HD map change detection for crowdsourcing update based on mid-to-high-end sensors [Текст] / P. Zhang, M. Zhang, J. Liu // Sensors. — 2021. — Т. 21, № 7. — С. 2477. — DOI: [10.3390/s21072477](https://doi.org/10.3390/s21072477).
 102. High Definition Map Mapping and Update: A General Overview and Future Directions [Текст] / B. Wijaya [и др.] // arXiv preprint arXiv:2409.09726. — 2024. — URL: <https://arxiv.org/abs/2409.09726>.
 103. Designing Network Design Spaces [Текст] / I. Radosavovic [и др.] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2020. — С. 10428—10436. — DOI: [10.1109/CVPR42600.2020.01044](https://doi.org/10.1109/CVPR42600.2020.01044). — URL: https://openaccess.thecvf.com/content_

[CVPR_2020/papers/Radosavovic_Designing_Network_Design_Spaces_CVPR_2020_paper.pdf](#).

104. *Tan, M.* EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks [Текст] / M. Tan, Q. V. Le. — 2020. — arXiv: [1905.11946 \[cs.LG\]](#). — URL: <https://arxiv.org/abs/1905.11946>.
105. *Geiger, A.* Are we ready for Autonomous Driving? The KITTI Vision Benchmark Suite [Текст] / A. Geiger, P. Lenz, R. Urtasun // Conference on Computer Vision and Pattern Recognition (CVPR). — 2012.
106. The Cityscapes Dataset for Semantic Urban Scene Understanding [Текст] / M. Cordts [и др.] // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2016.
107. BDD100K: A diverse driving video database with scalable annotation tooling [Текст] / F. Yu [и др.] // arXiv preprint arXiv:1805.04687. — 2020.
108. The DriveU Traffic Light Dataset: Introduction and Comparison with Existing Datasets [Текст] / A. Fregin [и др.] // 2018 IEEE International Conference on Robotics and Automation (ICRA). — 2018. — С. 3376—3383. — DOI: [10.1109/ICRA.2018.8460737](#).
109. Scalability in Perception for Autonomous Driving: Waymo Open Dataset [Текст] / P. Sun [и др.] //. — 2020. — arXiv: [1912.04838 \[cs.CV\]](#). — URL: <https://arxiv.org/abs/1912.04838>.
110. A Comprehensive Review on Traffic Datasets and Simulators for Autonomous Vehicles [Текст] / S. Sarker [и др.]. — 2025. — arXiv: [2412.14207 \[cs.RO\]](#). — URL: <https://arxiv.org/abs/2412.14207>.
111. *Гужва Н.С.* Оценка точности системы распознавания светофоров на основе действий водителя в системах активной помощи [Текст] / Гужва Н.С., Коробков А.А. // Журнал “Известия Института инженерной физики”. — 2025. — Т. 75, № 1. — С. 18—22.

Публикации автора по теме диссертации

В перечне, рекомендованном ВАК Минобрнауки России:

1. Гужва Н.С., Садеков Р.Н., Алгоритмы идентификации светофоров в системах помощи водителю трамвая, Журнал “Гироскопия и навигация”, 2024, том 32, № 3, стр. 47–65.
2. Гужва Н.С., Коробков А.А., Оценка точности системы распознавания светофоров на основе действий водителя в системах активной помощи, Журнал “Известия Института инженерной физики”, 2025, том 75, № 1, стр. 18–22.
3. Гужва Н.С., Прун В.Е., Садеков Р.Н.; ООО “Когнитив Роботикс”. Способ агрегации множественных проездов по маршруту для расчета оптимальных профилей скорости для городского рельсового транспорта. Патент на изобретение № 2798951 РФ, МПК В61 L 27/00. № 2023102622; Заявл. 06.02.23; Оpubл. 29.06.23, Бюл. № 19.
4. Гужва Н.С., Прун В.Е., Садеков Р.Н.; ООО “Когнитив Роботикс”. Способ визуальной локализации городского рельсового транспорта на основе фильтров частиц Калмана и данных о картировании. Патент на изобретение № 2807978 РФ, МПК G01 C 21/30. № 2023102623; Заявл. 06.02.23; Оpubл. 21.11.23, Бюл. № 33.
5. Гужва Н.С., Прун В.Е.; ООО “Когнитив Роботикс”. Способ позиционирования и определения ориентации рельсового транспорта с использованием информации о карте путей в арочно-линейном виде. Патент на изобретение № 2839177 РФ, МПК G01 C 21/30. № 2024130101; Заявл. 05.10.24; Оpubл. 28.04.25, Бюл. № 13.

В других изданиях:

6. Гужва Н.С., Садеков Р.Н., Прун В.Е., Шоломов Д.Л., Лобанов М.Г., Постников В.В., “Использование 3д сетей для «предсказания» моделей поведения транспортных средств в задаче беспилотного движения трамвая”, 2022 XXIX Санкт-Петербургская международная конференция по интегрированным навигационным системам (ICINS), 2022, с. 1–6, DOI: 10.23919/ICINS51784.2022.9815388.
7. Гужва Н.С., Садеков Р.Н., Али Б., Бакулев К.С., Шолохов А.В., “Оценка точности местоположения трамвая в условиях высотной застройки по данным визуально-геоинформационных систем”, 2023 XXX Санкт-Петербургская международная конференция по интегрированным навигационным системам (ICINS), 2023, с. 1–5, DOI: 10.23919/ICINS51816.2023.10168407.
8. Гужва Н.С., Садеков Р.Н., “Алгоритмы локализации и сопоставления светофоров в системах помощи водителю”, 2024 XXXI Санкт-Петербургская международная конференция по интегрированным навигационным системам (ICINS): Сборник материалов, Санкт-Петербург, 27 мая – 29 мая 2024 года. – Санкт-Петербург: “Концерн “Центральный научно-исследовательский институт “Электроприбор”, 2024, стр. 1–6.

Список рисунков

1.1	Архитектура подсистемы идентификации светофоров в автономных транспортных средствах	19
1.2	Классификация подходов к идентификации светофоров	22
1.3	Взаимодействие транспортного средства с дорожной инфраструктурой	23
1.4	Высокоточная цифровая карта дорожной сцены	27
1.5	Проблема сопоставления светофора с полосой движения	30
1.6	Изображение трамвая и системы координат VR	36
1.7	Визуализация распознавания дорожной сцены (3D детекции – машины (желтые), светофоры (красные), область сегментации – бордюры (красная), трамвайное полотно (бирюзовое)	37
1.8	Схема алгоритма идентификации светофоров	39
2.1	Трекинг	42
2.2	Сравнение подходов к межкамерному сопоставлению объектов. А) Традиционный подход с независимым однокамерным трекингом (SCT) и последующим глобальным сопоставлением. В) Современный подход с использованием единой глобальной модели, которая одновременно обрабатывает данные со всех камер	46
2.3	Этапы трекинга	50
2.4	Косинусное расстояние между эмбедингами	60
2.5	Мера пересечения двух боксов	61
2.6	Мера пересечения двух кубоидов	62
2.7	Евклидово расстояние между центрами объектов	63
2.8	Расстояние Махаланобиса	64
2.9	Расстояние между центрами 2D детекций	65
2.10	Классификация метрик по доменам: изображение (2D) и сцена (3D) . .	65
2.11	Пример изображений из набора данных для сопоставления светофоров.	67

2.12	Анализ процесса оптимизации весовых коэффициентов	69
3.1	Сигма-точки UKF	75
3.2	Анализ сходимости фильтров в условиях сильного шума	86
3.3	Сравнение оценок положения светофоров для EKF и UKF	87
3.4	Сравнение СКО оценки положения для EKF и UKF	88
3.5	Цифровая карта светофоров и путей	90
3.6	Сопоставление результатов с картой	92
3.7	Архитектура RegNet	94
3.8	Примеры светофоров разных классов	95
4.1	Диаграмма состояний системы управления проездом перекрестка . . .	107
4.2	Примеры данных из тестовой выборки	112
4.3	Визуализация результатов идентификации светофоров на перекрестках города Санкт-Петербург	114

Список таблиц

1	Уровни автоматизации транспортных средств по классификации SAE J3016:2021	18
2	Ключевые характеристики расширенного и сигма-точечного фильтров Калмана	78
3	Параметры шумов измерений для различных сценариев моделирования	85
4	Сводные метрики оценки положения светофоров разных фильтров . .	88
5	Матрица соответствия действий	105
6	Основные внешние события и соответствующие переходы состояний системы	108
7	Критерии формирования группы событий для сценария S_1	110
8	Критерии формирования группы событий для сценария S_2	110
9	Оценка точности представленных алгоритмов	113

Приложение А

Акт о внедрении

УТВЕРЖДАЮ

Начальник Управления информационных технологий и интеллектуальных систем
СПб ГУП «Горэлектротранс»

 П.С. Поляков

« 10 » 2025 года

АКТ

о внедрении результатов диссертационной работы
Гужвы Никиты Сергеевича
«Идентификация светофоров на основе 3D измерений нейронных сетей в
мультикамерных системах помощи водителя трамвая»

10.10.2025

г. Санкт-Петербург

Комиссия в составе:

1. Начальника отдела испытаний Управления информационных технологий и интеллектуальных систем СПб ГУП «Горэлектротранс» Шепяковой А.О.;
2. Водителя трамвая-испытателя отдела испытаний Управления информационных технологий и интеллектуальных систем СПб ГУП «Горэлектротранс» Романова Я.А.;

рассмотрев автореферат диссертации Гужвы Никиты Сергеевича на тему «Идентификация светофоров на основе 3D измерений нейронных сетей в мультикамерных системах помощи водителя трамвая» установила следующее:

1. Система активной безопасности и помощи водителю (АБПВ) устанавливается на трамваях, выпускаемых с 2022 года и эксплуатируемых в СПб ГУП "Горэлектротранс". В настоящее время парк трамваев с установленной системой АБПВ насчитывает 334 комплекта. Система АБПВ имеет следующие функции:

- подает предупреждения о запрещающем сигнале светофора при движении по маршруту и плавно снижает скорость трамвайного вагона вплоть до полной остановки при отсутствии надлежащих действий со стороны водителя;

- подает предупреждения при обнаружении на пути движения трамвайного вагона, транспортного средства или человека, с которым имеется опасность столкновения, и снижает скорость трамвайного вагона вплоть до полной остановки при отсутствии надлежащих действий со стороны водителя;

- обеспечивает соблюдение безопасного скоростного режима – ограничивает скорость путем отключения или ограничения ходовых позиций контроллера водителя при достижении порога ограничения и включением тормозных позиций при значительном превышении номинала ограничения, внесенного в систему профиля скорости, при движении по маршруту.

2. Полученные Гужвой Н.С. результаты в виде программного кода используются в реализации функции «проезда перекрестка по сигналам светофора» системы АБПВ. Данная функция является крайне востребованной на практике, в силу отсутствия, с одной стороны оборудованных беспроводными системами связи перекрестков в городе, и имеющимися случаями аварий с другой.

3. Совокупность предложенных в работе методик позволяет на практике:

- использовать камеры дальнего наблюдения для заблаговременного обнаружения сигналов светофоров при подъезде к перекрестку или регулируемому пешеходному переходу;

- определять сигнал светофора в условиях, когда основной светофор заслонен или находится вне зоны видимости;
- более точно определять позицию светофоров на дорожной сцене и проводить операцию идентификации «своего» светофора.

4. Предложенная «Методика оценки точности» применяется в отделе испытаний Управления информационных технологий и интеллектуальных систем СПб ГУП «Горэлектротранс» для оценки точности работы соответствующей подсистемы АБПВ. Использование действий водителя в качестве эталона позволяет оценить точность подсистемы без использования внешней информации, что является крайне существенным на практике в силу использования большого парка трамваев, необходимости проведения оперативного мониторинга за качеством работы системы АБПВ, на любых требуемых временных отрезках.

Подписи членов комиссии:

Начальник отдела испытаний
Управления информационных технологий
и интеллектуальных систем
СПб ГУП «Горэлектротранс»



А.О. Шептякова

Водитель трамвая-испытатель
отдела испытаний Управления информационных
технологий и интеллектуальных систем
СПб ГУП «Горэлектротранс»



Я.А. Романов