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

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»

Автореферат диссертации
на соискание учёной степени кандидата технических наук

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

Москва — 2025

Общая характеристика работы

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 **??**).

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

Содержание работы

Первая глава – Анализ существующих решений и постановка задачи

Анализ существующих подходов идентификации светофоров

Постановка задачи детектирования светофоров обычно формулируется и решается на изображении, при этом может использоваться априорная информация о карте. Цифровая карта используется для ограничения области поиска светофоров на изображении. Для детектирования положения светофора и определения его сигнала используются детектирующие и классификационные сети, которые в настоящее время реализуются на базе глубоких нейронных сетей.

Задача детектирования светофоров может решаться как для ночных, так и дневных условий, учитывать особенности, связанные с небольшим размером светофора на изображении. Использовать информацию с карты. Кроме того, для решения задачи сопоставления с картой, проекции и репроекции данных могут использоваться лидары, стереокамеры или несколько камер. Постановки задачи также могут учитывать разные локации: городская среда или автострада; типы светофоров: автомобильные или трамвайные.

Общее решение задачи, как правило, складывается из решения следующих подзадач:

- детектирование и классификация сигналов светофоров на имеющемся изображении;
- трекинг светофоров между кадрами и между датчиками;
- фильтрация измерений и сопоставление полученных измерений с картой.

Краткий обзор работ по предложенной теме представлен в таблице 1.

Таблица 1 — Анализ существующих подходов

Датчики и особенности	Детектирование	Трекинг и фильтрация	Цифровая карта
Стереокамера, одометрия и инерциальный модуль. Светофоры на изображении имеют крайне небольшие размеры	YOLO, автомобильные светофоры со стрелками, отдельная нейронная сеть для классификации	На основе отдельной нейронной сети	Не используется

Лидар и три камеры. Беспилотный трамвай. Пригород	SSD детектор. Трамвайные светофоры	Скрытая марковская модель. Межкамерная обработка	Высокоточная лидарная карта, использование лидара для формирования областей поиска
Лидар, одна камера	YOLO, светофоры двух классов. Автомобильные светофоры	Накопление детекций в буфере, использование кластеризации	Высокоточная лидарная карта, использование лидара для формирования областей поиска
Несколько камер с разными фокусными расстояниями, GPS, инерциальный модуль, городская среда	YOLOv5 для каждой камеры. Автомобильные светофоры	Венгерский алгоритм и скрытая марковская модель	Цифровая карта для формирования областей поиска
Стереопара, GPS	Нейронная сеть	Фильтр Калмана и венгерский алгоритм для формирования карты светофоров	Не используется

Анализ работ показал, что решение задачи распознавания светофоров строится на основе многоэтапных алгоритмов. При этом реализованные в настоящее время алгоритмы используют информацию двух и более одинаковых (для формирования стерео) камер или лидаров, что существенно облегчает и повышает точность решения задачи. На этапе детектирования чаще всего используется априорная информация от цифровой карты, которая также получается на основе высокоточных измерений. Однако в известных работах не представлены решения, использующие камеры с существенно разными фокусными расстояниями, новые архитектуры нейронных сетей, способные формировать 3D-информацию о дорожной сцене. Кроме того, решения редко проверяются в условиях городской среды (как наиболее трудном случае), а также не связаны с задачей проезда перекрестка исходя из детектирования только «нужных» светофоров, а не всех возможных.

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) совпадает с центром переднего бампера и находится на земле, рисунок 1. 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 — 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 2 показан результат работы сети.

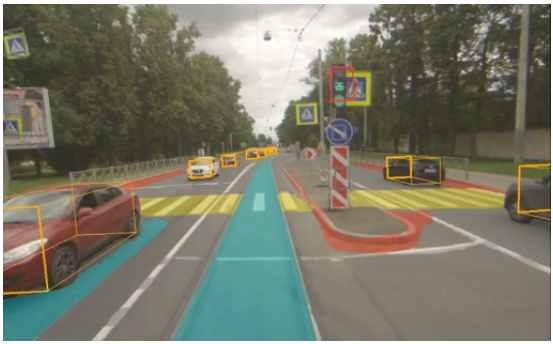

Рисунок 2 — Визуализация распознавания дорожной сцены (3D детекции – машины (желтые), светофоры (красные), область сегментации – бордюры (красная), трамвайное полотно (бирюзовое)

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, м;
- 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 - эмбеддинг (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. 3):

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.

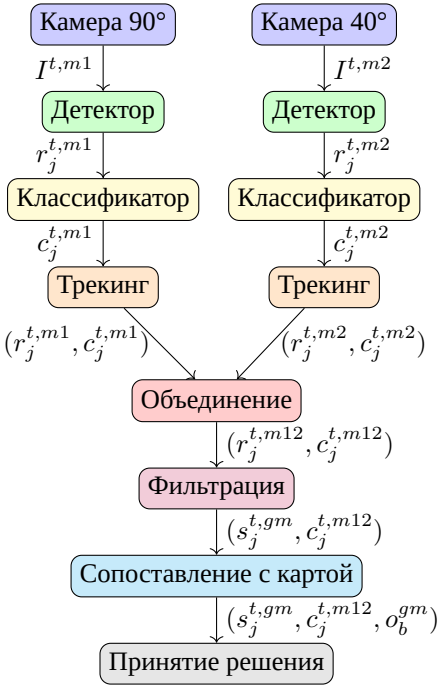

Рисунок 3 — Схема алгоритма идентификации светофоров

Вторая глава – Методика сопоставления детекций в двухкамерной разнофокусной системе

Трекинг

В современных задачах компьютерного зрения трекинг объектов, таких как светофоры, требует комбинирования методов, устойчивых к специфическим условиям городской среды. Классические алгоритмы, такие как SORT, основанные на моделях движения и метриках геометрического перекрытия (Intersection over Union), демонстрируют ограниченную эффективность в сценариях с малым размером объектов (например, светофоры на удалении) и их плотным расположением (”гроздьями”на столбах). В таких случаях ошибки трекинга возникают из-за:

- Низкой точности предсказания перемещений объектов на основе фильтра Калмана (модель движения светофоров статична);
- Невозможности различить объекты с близкими координатами и идентичными визуальными характеристиками.

Для решения этих проблем в работе предложен комплексный гибридный метод для трекинга объектов, сочетающий:

- Модели движения с использованием адаптированного фильтра Калмана;
- Ре-идентификация (re-ID) на основе визуальных признаков с использованием эмбеддингов (DeepSORT, FairMOT).

Архитектура трекинга включает три ключевых компонента:

1. **Детекция и генерация эмбеддингов**
   Для каждого j-го детектируемого объекта формируется 128-мерный вектор признаков E_j с помощью re-ID ветви нейронной сети (аналогично FairMOT).
2. **Модель движения на основе фильтра Калмана**
   Для предсказания положения объектов используется линейная модель с матрицей перехода F. Вектор состояния объекта $x = [x, y, v_x, v_y]^T$ включает координаты центра (x, y) и скорости по осям (v_x, v_y):

$$\hat{x}_t = F \cdot \hat{x}_{t-1} + w_t,\ F = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{1}$$

   где Δt — временной интервал между кадрами, w_t — шум процесса.
3. **Ассоциация треклетов и детекций**
   Под треклетом понимается последовательность детекций одного объекта на нескольких последовательных кадрах, представленная вектором состояния и усредненным эмбеддингом. Процесс ассоциации треклетов и детекций организован в три каскадных этапа:

а) **Первичное сопоставление по эмбеддингам**
Для каждого активного треклета T_i вычисляется косинусное расстояние до всех новых детекций D_j:

$$d_{cos}(T_i, D_j) = 1 - \frac{\bar{E}_i E_j}{|\bar{E}_i||E_j|} \tag{2}$$

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

$$\bar{E}_i^{(t)} = \beta \cdot \bar{E}_i^{(t-1)} + (1 - \beta) \cdot E_j^{(t)},\ \beta = 0.9. \tag{3}$$

Пары с $d_{cos} < \tau_1$ сопоставляются через венгерский алгоритм.

б) **Вторичная проверка по геометрическому перекрытию**
Для оставшихся несопоставленных треклетов и детекций вычисляется IoU:

$$IoU(B_i, B_j) = \frac{B_i \cap B_j}{B_i \cup B_j}, \tag{4}$$

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

в) **Инициализация новых треклетов**
Детекции, оставшиеся несопоставленными после этапов 1-2, инициируют новые треклеты с начальным эмбеддингом $\bar{E}_i(0) = E_j$.

Объединение результатов детекций между камерами

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 полученными с двух камер необходимо объединять, т.е. находить между ними соответствие. Поскольку под трекингом чаще всего понимается процедура межкадрового слежения, а не межкамерного, чтобы не путать читателя используется термин «объединение».

Задача объединения данных не является тривиальной в силу достаточно большого количества переменных и может быть решена с использованием следующих функций расстояния:

– Косинусное расстояние между эмбеддингами (см. уравнение (2)). Данная техника широко распространена на практике и не зависит от параметров настроек камеры. Однако эмбеддинги могут совпасть у похожих объектов, которые при этом находятся в разных частях изображения и эмбеддинг может меняться от кадра к кадру (напр. из-за мигания светофора или смены его сигнала).
– Мера пересечения двух боксов (Intersection over Union (IoU) 2D) 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:

$$c_j^{vr} = R_1^{vr\top}(d_1(K_1^{-1}c_j^1) - T_1^{vr}) \tag{5}$$
$$c_j^2 = K_2(R_2^{vr}c_j^{vr} + T_2^{vr}) \tag{6}$$

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 (4) между перенесенным и детектированным bbox. Данный подход требует точной калибровки камер и надежной оценки глубины.
– 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:

$$IoU_{3D} = \frac{V_{intersection}}{V_{union}} = \frac{\text{vol}(B_1 \cap B_2)}{\text{vol}(B_1 \cup B_2)} \tag{7}$$

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 расхождениям в оценке глубины объектов, что либо исключает пересечение кубоидов в пространстве, либо создает ложные пересечения.
– Евклидово расстояние (L2) между центрами объектов (в пространстве) рассчитывается как:

$$d_{L2} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2} \tag{8}$$

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.

– Расстояние Махаланобиса, близкое по своей природе 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:

$$d_M = \sqrt{(x_1 - x_2)^T S^{-1} (x_1 - x_2)} \tag{9}$$

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

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

$$d_{pixel} = \sqrt{(u_1 - u_2)^2 + (v_1 - v_2)^2} \tag{10}$$

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 (5). 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.

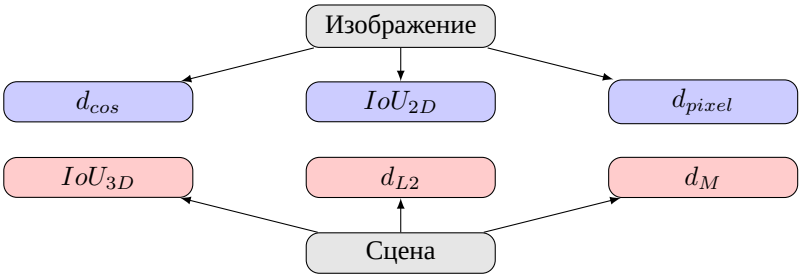

Рисунок 4 — Классификация метрик по доменам: изображение (2D) и сцена (3D)

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, при использовании нескольких метрик необходимо определить коэффициенты их влияния.

Следует отметить, что метрики, основанные на трехмерной сцене (см. рисунок 4), можно считать производными от данных изображений, поскольку 3D-координаты получаются путем преобразования 2D-данных. Поэтому для построения итоговой метрики было решено использовать только оригинальные метрики, работающие непосредственно с изображениями. На основании проведенных экспериментов была создана итоговая метрика, которая объединяет данные о взаимном расположении центров 2D-объектов, эмбеддинги и IoU 2D-объектов. С последними двумя компонентами вводятся веса, позволяющие адаптировать их вклад в итоговую метрику. Итоговая метрика соответствия M вычисляется следующим образом:

$$M = \begin{cases} w_{emb} \cdot d_{cos} + w_{iou} \cdot (1 - \text{IoU}), & \text{если } d_{pixel} < d_{thresh} \\ \infty, & \text{иначе} \end{cases} \tag{11}$$

где d_{cos} – косинусное расстояние между эмбеддингами, 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 определения оптимальных значений весовых коэффициентов w_{emb} и w_{iou} был проведен ряд экспериментов с использованием размеченного набора данных, содержащего эталонные соответствия детекций светофоров между двумя камерами.

Задача оптимизации весов формулировалась как минимизация функции ошибки E:

$$E(w_{emb}, w_{iou}) = \sum_{i=1}^{N} \sum_{j=1}^{K} L(y_{ij}, \hat{y}_{ij}(w_{emb}, w_{iou})) \tag{12}$$

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 (использовалась бинарная кросс-энтропия):

$$L(y, \hat{y}) = -y \log(\hat{y}) - (1 - y) \log(1 - \hat{y}) \tag{13}$$

При этом накладывались ограничения:

$$\begin{cases} w_{emb} + w_{iou} = 1 \\ w_{emb}, w_{iou} \geq 0 \end{cases} \tag{14}$$

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 5 представлены результаты оптимизации весовых коэффициентов.

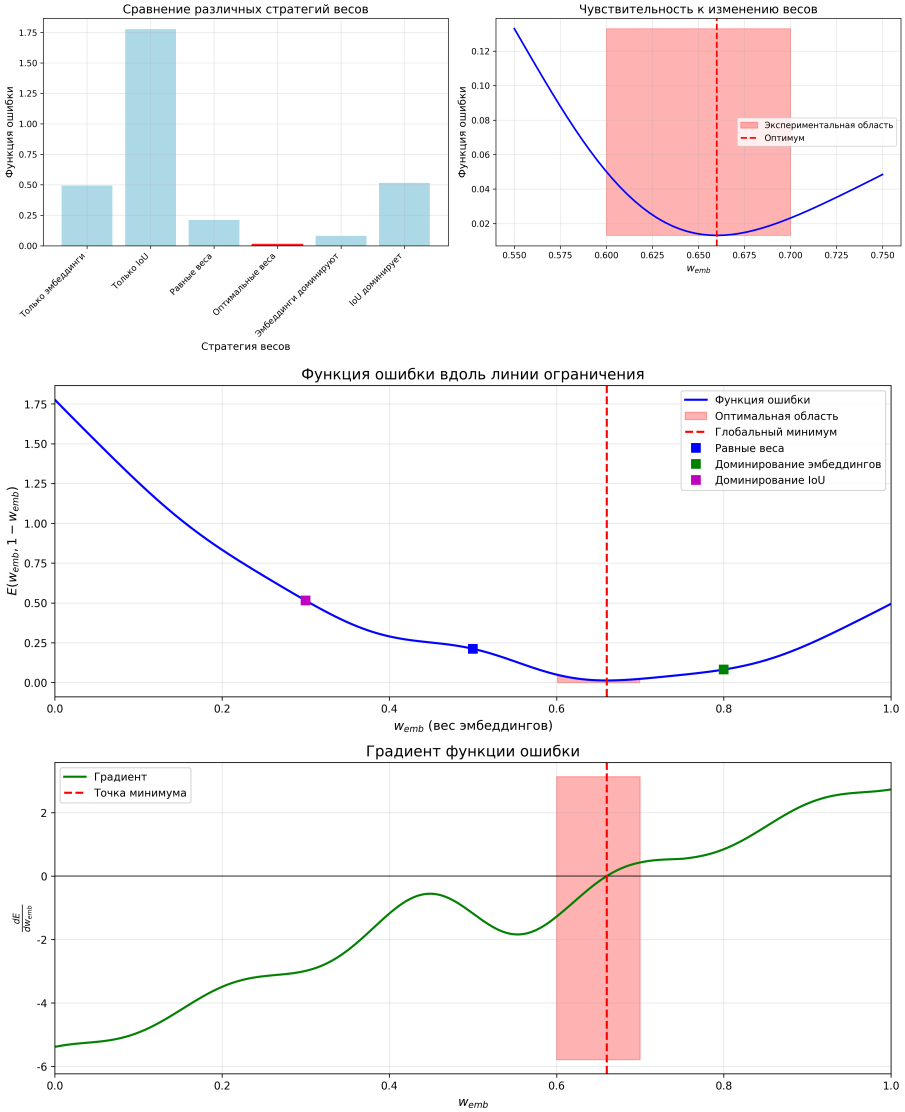

Рисунок 5 — Анализ оптимизации весовых коэффициентов

Итоговое сопоставление светофоров между камерами осуществляется с помощью венгерского алгоритма, который решает задачу о назначениях, минимизируя суммарную стоимость паросочетаний. В качестве матрицы стоимости используется матрица метрик M_{ij} между всеми парами светофоров с двух камер. Математически это можно записать как:

$$\min_X \sum_{i=1}^{n} \sum_{j=1}^{m} M_{ij} X_{ij} \tag{15}$$

при ограничениях:

$$\sum_{j=1}^{m} X_{ij} = 1, \quad \sum_{i=1}^{n} X_{ij} = 1, \quad X_{ij} \in \{0, 1\} \tag{16}$$

где X_{ij} – бинарная матрица назначений, n и m – количество светофоров на каждом изображении.

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 эмбеддингов детектируемых областей.

Третья глава – Методика фильтрации пространственных координат светофоров

Оптимальная фильтрация координат местоположения светофоров

Результаты, полученные от нейронной сети, часто содержат значительный уровень случайных неточностей или «шума», что требует их дополнительной фильтрации. С этой целью применяется фильтр Калмана. В ходе разработки системы были реализованы и протестированы два варианта фильтрации.

Первоначально использовался известный в литературе и используемый для решения подобного рода задач расширенный фильтр Калмана (EKF) с константной моделью движения

$$s_{t+1}^{gm} = s_t^{gm} + w_t \tag{17}$$

где $s_t^{gm} = [x_{gm}, y_{gm}]$ – вектор состояния в GM, w_t – шум модели.

Уравнение измерений имеет вид

$$z_t^{vr} = [r, \phi] = h(s_t^{gm}) + v_t \tag{18}$$

где r, ϕ – полярные координаты светофора в VR, v_t – шум измерений. Такое представление вектора измерений в полярных координатах было выбрано для более точного описания шума измерений, поскольку угол пеленга ϕ определяется значительно точнее, чем дальность r до объекта при измерениях с помощью

камеры. Функция преобразования h из вектора состояния в вектор измерений имеет вид

$$h(s_t^{gm}) = \begin{bmatrix} \sqrt{(x_{gm})^2 + (y_{gm})^2} \\ \arctan(\frac{y_{gm}}{x_{gm}}) \end{bmatrix} \tag{19}$$

Якобиан H, необходимый для EKF, вычисляется как

$$H = \begin{bmatrix} \frac{x_{gm}}{\sqrt{(x_{gm})^2+(y_{gm})^2}} & \frac{y_{gm}}{\sqrt{(x_{gm})^2+(y_{gm})^2}} \\ -\frac{y_{gm}}{(x_{gm})^2+(y_{gm})^2} & \frac{x_{gm}}{(x_{gm})^2+(y_{gm})^2} \end{bmatrix} \tag{20}$$

Алгоритм EKF состоит из двух этапов. Этап предсказания рассчитывается по формулам

$$s_t^{gm} = s_{t-1}^{gm} \tag{21}$$

$$P_t = P_{t-1} + Q_t \tag{22}$$

Этап коррекции рассчитывается по формулам

$$K_t = P_t H_t^T (H_t P_t H_t^T + R_t)^{-1} \tag{23}$$

$$\hat{s}_t^{gm} = s_t^{gm} + K_t(z_t^{vr} - h(s_t^{gm})) \tag{24}$$

$$\hat{P}_t = (I - K_t H_t) P_t \tag{25}$$

где Q_t и R_t – ковариационные матрицы шумов модели и измерений соответственно, K_t – усиление Калмана, P_t – ковариационная матрица ошибки оценки состояния.

Итоговое состояние $\hat{s}_t^{gm}$ представляет собой оптимальную оценку положения светофора с учетом как предсказанного состояния, так и полученных измерений, взвешенных через коэффициент усиления фильтра.

Тем не менее, применение данного фильтра не дало желаемых результатов вследствие ограниченности модели. В результате был разработан новый алгоритм, основанный на UKF фильтре с расширенными векторами состояния и измерений.

Уравнение состояния нового алгоритма описывается уравнением (17). Вектор состояния значительно расширен для более точного определения положения светофора

$$s_t^{gm} = [x_{gm}, y_{gm}, z_{gm}, w, l, h, \theta_{gm}] \tag{26}$$

где $[x_{gm}, y_{gm}, z_{gm}]$ – координаты светофора в GM, $[w, l, h]$ – размеры светофора в метрах, θ_{gm} – угол поворота светофора (рысканье). Дополнительные

параметры в новом векторе состояния позволяют существенно повысить точность оценки положения светофора. Координата z_{gm} учитывает высоту установки светофора, что необходимо для корректного сопоставления с координатами в пикселях на изображении. Угол рысканья θ_{gm} позволяет определить ориентацию светофора и вносит дополнительные геометрические ограничения в модель. Размеры w, l, h обеспечивают более стабильную оценку габаритов светофора, т.к. в измерениях эти значения могут сильно варьироваться.

Особенностью новой версии является реализация тесносвязанной схемы комплексирования, где измерения представляют собой непосредственно данные от нейронной сети

$$z_t^{vr} = [u_{cam}, v_{cam}, d_{cam}, \theta_{vr}, w, l, h] \tag{27}$$

Функция преобразования h из вектора состояния в вектор измерений выполняется последовательно

$$p_{vr} = R_m^{gm/vr} \cdot p_{gm} + T_m^{gm/vr} \tag{28}$$

$$p_{cam} = R_m^{vr/cam} \cdot p_{vr} + T_m^{vr/cam} \tag{29}$$

$$\begin{bmatrix} u_{cam} \\ v_{cam} \end{bmatrix} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \end{bmatrix} \cdot \begin{bmatrix} x_{cam}/z_{cam} \\ y_{cam}/z_{cam} \\ 1 \end{bmatrix} \tag{30}$$

$$d_{cam} = z_{cam} \tag{31}$$

$$\theta_{vr} = \theta_{gm} + \Delta\theta_{gm/vr} \tag{32}$$

где f_x, f_y, c_x, c_y – внутренние параметры камеры, $\Delta\theta_{gm/vr}$ – разница углов рысканья между GM и VR системами координат. Размеры объекта w, l, h передаются без изменений.

Алгоритм UKF состоит из следующих этапов. Сначала генерируются сигма-точки

$$\chi_{t-1} = [\hat{s}_{t-1} \;\; \hat{s}_{t-1} + \gamma\sqrt{P_{t-1}} \;\; \hat{s}_{t-1} - \gamma\sqrt{P_{t-1}}] \tag{33}$$

где $\gamma = \sqrt{n + \lambda}$, n – размерность вектора состояния.

Выполняется этап предсказания

$$\chi_{t|t-1} = f(\chi_{t-1}) \tag{34}$$

$$\hat{s}_t^- = \sum_{i=0}^{2n} W_i^m \chi_{i,t|t-1} \tag{35}$$

$$P_t^- = \sum_{i=0}^{2n} W_i^c [\chi_{i,t|t-1} - \hat{s}_t^-][\chi_{i,t|t-1} - \hat{s}_t^-]^T + Q \tag{36}$$

Выполняется этап коррекции

$$\mathcal{Z}_{t|t-1} = h(\chi_{t|t-1}) \tag{37}$$

$$\hat{z}_t^- = \sum_{i=0}^{2n} W_i^m \mathcal{Z}_{i,t|t-1} \tag{38}$$

$$P_{zz} = \sum_{i=0}^{2n} W_i^c [\mathcal{Z}_{i,t|t-1} - \hat{z}_t^-][\mathcal{Z}_{i,t|t-1} - \hat{z}_t^-]^T + R \tag{39}$$

$$P_{xz} = \sum_{i=0}^{2n} W_i^c [\chi_{i,t|t-1} - \hat{s}_t^-][\mathcal{Z}_{i,t|t-1} - \hat{z}_t^-]^T \tag{40}$$

$$K_t = P_{xz} P_{zz}^{-1} \tag{41}$$

$$\hat{s}_t = \hat{s}_t^- + K_t(z_t - \hat{z}_t^-) \tag{42}$$

$$P_t = P_t^- - K_t P_{zz} K_t^T \tag{43}$$

где W_i^m, W_i^c – весовые коэффициенты для среднего значения всех точек и ковариации соответственно.

Замена EKF на UKF была обусловлена сложностью функции преобразования координат h, для которой вычисление якобиана – задача сложная как алгебраически, так и численно. Кроме того, UKF в целом обеспечивает более высокую точность оценки состояния для существенно нелинейных систем.

Сравнительный анализ двух версий фильтров показал значительное превосходство второго варианта. Использование расширенного вектора состояния позволяет отслеживать все параметры светофора, включая его ориентацию и габариты. Тесносвязанная архитектура обеспечивает более точную фильтрацию за счет прямого использования измерений от нейронной сети без промежуточных преобразований. Применение UKF вместо EKF улучшило устойчивость фильтра и точность оценки параметров.

Сопоставление результатов с картой

По результатам работы фильтра формируется вектор координат светофоров, который далее сравнивается с информацией о местоположении светофоров, имеющейся в цифровой карте.

Цифровая карта представляет собой базу данных, в которой хранятся координаты местоположения светофоров, участки маршрутов, которые они регулируют, классы сигналов для разрешения проезда участков маршрута и другая служебная информация.

Сопоставление векторов координат местоположений светофоров с рассчитанными значениями осуществляется с использованием расстояния Махаланобиса, которое определяется по следующей формуле:

$$d_{ib} = \sqrt{(s_i^{gm} - o_b^{gm})^T P^{-1} (s_i^{gm} - o_b^{gm})} \tag{44}$$

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. Для найденных дистанций составляется матрица «стоимости» со значениями метрики обратно пропорциональной расстоянию Махаланобиса. Далее с использованием венгерского алгоритма по уравнению (15) определяются лучшие пары «измерение – карта».

Классификация сигналов светофоров

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 6.

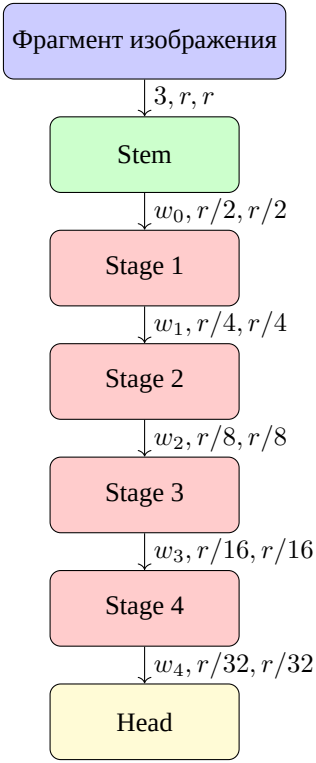

Рисунок 6 — Архитектура RegNet

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

$$CE(p_t) = -\sum_{c=1}^{C} y_c \log(p_c), \tag{45}$$

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

$$FL(p_t) = -\sum_{c=1}^{C} \alpha_c y_c (1 - p_c)^{\gamma} \log(p_c), \tag{46}$$

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 7.

Рисунок 7 — Примеры светофоров разных классов

Для стабилизации распознавания сигналов и устранения «дрожания» предсказаний применяется временная фильтрация на основе параллельных гипотез. Для каждого светофора j система поддерживает набор гипотез $\mathcal{P}_j$, где каждая гипотеза $p \in \mathcal{P}_j$ описывается кортежем (c_p, A_p, L_p): класс сигнала, накопленное время подтверждения и метка времени последнего наблюдения.

При поступлении нового результата C_j^t в момент τ_t для существующей гипотезы p с классом $c_p = C_j^t$ время A_p увеличивается на $(\tau_t - \tau_{t-1})$, а метка L_p обновляется. Если гипотеза для C_j^t отсутствует и класс не совпадает с текущим стабильным состоянием S_j, создается новая.

Стабильное состояние S_j переключается, когда накопленное время A_p для какой-либо гипотезы p превышает порог T_{thresh}:

$$S_j^t = \begin{cases} c_p, & \text{если } \exists p \in \mathcal{P}_j : A_p \geq T_{\text{thresh}} \\ S_j^{t-1}, & \text{в противном случае} \end{cases} \tag{47}$$

После переключения все гипотезы для светофора j сбрасываются. Для отсева шумных распознаваний гипотезы, не наблюдавшиеся дольше порога T_{timeout} (т.е. $\tau_t - L_p > T_{\text{timeout}}$), удаляются из $\mathcal{P}_j$.

Четвертая глава – Методика оценки точности алгоритма идентификации светофоров

Анализ существующих подходов оценки точности работы подсистемы распознавания светофоров

Оценка точности работы подсистемы распознавания светофоров может производиться на нескольких уровнях. Первый уровень оценки точности — это изображение. Как правило, для оценки точности используются классические метрики точности (precision) и полнота (recall). 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 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 (multiple object tracking accuracy), которая акцентирует внимание на общих ошибках отслеживания (трекинга) и детектирования. Примерами датасетов для оценки качества системы на первом и втором уровне являются LISA и DriveU.

Однако описанные выше подходы имеют и недостатки:

- подготовка и разметка данных требует существенных временных и человеческих затрат;
- собранные датасеты, как правило, являются несбалансированными, поскольку содержат преимущественно данные о часто встречающихся сигналах и редко включают информацию о дополнительных секциях светофора, таких как стрелки;
- при обучении сетей на практике может достигаться обратный эффект, когда повышение качества работы сети по отдельным классам не влияет или даже ухудшает общее качество работы подсистемы в силу малой встречаемости светофоров данного класса.

В силу данных обстоятельств оценка точности подсистемы идентификации светофоров может проводиться и в комплексе с системой управления. Такие эксперименты прежде всего востребованы в автопилотируемых решениях, поскольку от качества решения данных задач зависит «скорость» движения робота по улицам. При этом факт правильного распознавания сигналов светофоров оценивается путем остановки на перекрестке на запрещающий сигнал или проезда (начала движения) на разрешающий сигнал.

Применительно к рассматриваемой системе помощи водителю трамвая, автоматически оценить точность работы системы помощи водителя с активной

подсистемой контроля сигналов светофора невозможно в силу того, что водители зачастую сами останавливают трамвай перед стоп-линией и начинают движение, когда загорается разрешающий сигнал светофора, поэтому оценка точности системы может производиться только в теневом режиме, путем сравнения действий водителя и действий умного помощника.

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 2.

Таблица 2 — 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 2 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.

Для исполнения своей непосредственной функции система контролирует ситуацию при подъезде трамвая к стоп-линии (первый сценарий) и, находясь непосредственно на ней (второй сценарий). Второй сценарий активируется, если в первом было принято решение – ОСТАНОВКА.

Первый сценарий наступает при приближении к перекрестку, когда одна из камер детектирует светофор. На данном этапе принимается решение об остановке трамвая, если сигнал светофора запрещающий, или продолжении движения, если разрешающий. Второй сценарий реализуется, если трамвай

находился на стоп-линии, в данном сценарии контролируется смена сигнала светофора на разрешающий для выдачи действия – ДВИЖЕНИЕ.

Сценарии работы системы:

– подъезд к перекрестку (S_1);
– трогание с перекрестка (S_2).

Входная информация

В качестве входной информации – с одной стороны выступают действия водителя, с другой, действия подсистемы. Действия водителя определяются D_i – положением ручки контроллера хода, v_i – скоростью движения трамвая. Действие подсистемы определяется положением устанавливаемой ручки хода контроллера H_i и формируемой целевой скоростью для торможения s_i.

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».

Критерии для формирования событий

Для определения полной группы событий в каждом сценарии задаются критерии, представленные в таблице 3 и таблице 4. Эти критерии основаны на анализе действий системы и водителя и помогают классифицировать события как TP, TN, FP и FN.

Таблица 3 — Критерии формирования группы событий для сценария S_1

Событие	Действие	Критерий ручки хода	Критерий скорости
TP	A_P	$H_i \leq 0,\ D_i \leq 0$	$s_i \leq v_i,\ v_i \to 0$
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	$H_i \geq 0,\ D_i \leq 0$	$s_i > v_i,\ v_i \to 0$

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 $s_i \leq v_i,\ v_i \to 0$ 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 $v_i \to 0$ указывают на то, что водитель тормозит на перекрестке.

Таблица 4 — 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. **TP:** 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.

Результаты оценки точности системы

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 5 — 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 5 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 8 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.

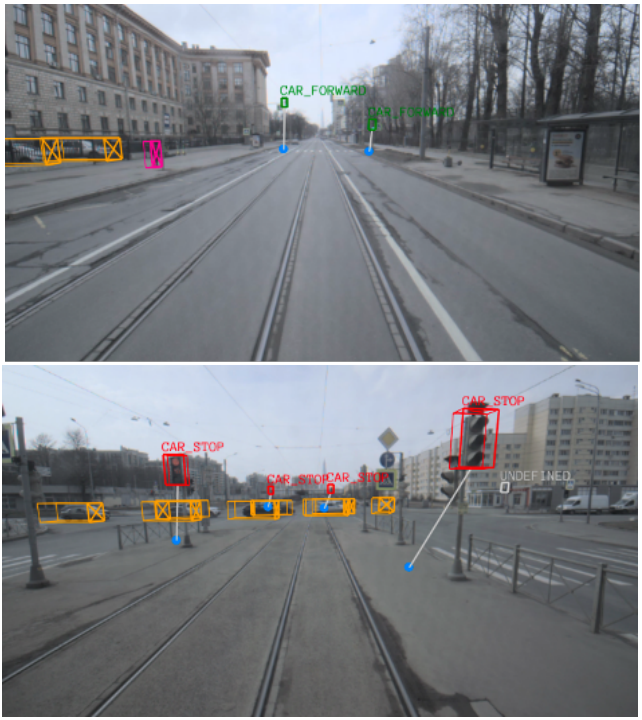

Рисунок 8 — Визуализация результатов распознавания на перекрестках города Санкт-Петербург

Заключение

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 эмбеддингов 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.

Основные положения диссертационной работы изложены в следующих опубликованных работах

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 B61 L 27/00. № 2023102622; Заявл. 06.02.23; Опубл. 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; Опубл. 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; Опубл. 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.