

УДК 656.13
EDN IGOZNL

Н. С. ГУЖВА, Р. Н. САДЕКОВ

АЛГОРИТМЫ ИДЕНТИФИКАЦИИ СВЕТОФОРОВ НА ОСНОВЕ НЕЙРОННЫХ СЕТЕЙ В МУЛЬТИКАМЕРНЫХ СИСТЕМАХ ПОМОЩИ ВОДИТЕЛЮ

В статье рассматривается задача идентификации (детектирования, фильтрации и сопоставления с картой) светофоров по последовательности изображений в активных системах помощи водителю трамвая, оснащенных несколькими камерами с разными фокусными расстояниями. Подробно описывается ход решения задачи – от формирования измерений на выходе нейронной сети (детекций) для каждой из камер до сопоставления результатов с картой. В отличие от других исследований на эту тему авторы настоящей работы используют 3D-измерения в качестве выходных данных для нейронной сети и сигма-точечный фильтр Калмана (UKF) для определения положения светофора, а также новый способ объединения данных двух камер. Эффективность предложенного алгоритма и его модификации оценивалась в натурных условиях. Результаты экспериментов показали, что совместно с системой управления трамваем, оснащенной двумя камерами, алгоритм позволяет достичь точности 76,19% и полноты 97,46%.

Ключевые слова: компьютерное зрение, детектирование светофоров, распознавание светофоров, монокулярная камера, беспилотный трамвай.

Введение

Автономное движение транспортного средства возможно при одновременном решении целого ряда задач, связанных с распознаванием дорожной ситуации и прогнозом ее развития. Одной из таких задач является идентификация сигналов светофоров, которая осуществляется с помощью систем технического зрения или обмена информацией между автомобилем и инфраструктурой (Vehicle to Infrastructure). Такая система позволяет транспортному средству получать данные о сигнале светофора напрямую, без их распознавания, однако при этом оно должно быть оборудовано специальными датчиками или иметь выход в интернет с доступом к цифровой городской сети [1], управляющей в том числе и переключением светофоров. Целесообразнее остановиться на последнем варианте, поскольку он не требует оснащения светофоров средствами приема–передачи сигналов. Тем не менее как отечественный, так и зарубежный опыт показывает, что внедрение таких систем в городскую среду обходится очень дорого, поэтому они есть далеко не во всех городах. В связи с этим широко распространен альтернативный подход – идентификация сигналов светофора на основе систем технического зрения.

Гужва Никита Сергеевич. Аспирант, НИТУ «МИСИС»; Cognitive Technologies (Москва).

Садеков Ринат Наилевич. Доктор технических наук, профессор кафедры НИТУ «МИСИС»; Cognitive Technologies. Действительный член международной общественной организации «Академия навигации и управления движением». ORCID 0000-0001-6286-358X.

Впервые распознавание сигналов светофоров (Traffic Light Recognition) появилось в системах помощи водителю, чтобы сократить время простоя на перекрестках. При переключении сигнала светофора с запрещающего на разрешающий подавался звуковой сигнал о возможности возобновить движение [2], при этом в само управление автомобилем система не вмешивалась. Создание системы, способной осуществлять торможение транспортного средства при попытке пересечь перекресток на запрещающий сигнал светофора, по ряду причин является значительно более сложной задачей. Так, система помощи водителю трамвая [3, 4] должна автоматически распознавать сигналы светофоров и полностью останавливать транспортное средство (не дальше стоп-линии) в том случае, если водитель продолжает движение вопреки запрещающему сигналу. Успешное решение задачи в такой постановке приводит к повышению безопасности дорожного движения и минимизирует риск дорожно-транспортных происшествий.

Распознавание сигналов светофора по-прежнему сопряжено с многочисленными трудностями, хотя метод широко востребован в различных системах автопилотирования, а используемые при этом алгоритмы глубокого обучения [5] активно развиваются. Причин этому несколько:

- погодные условия и плохая видимость усложняют идентификацию сигналов светофора;
- большое разнообразие светофоров, установленных на улицах города;
- кучное размещение светофоров на перекрестках;
- наличие других многочисленных источников света (уличное освещение, фары автомобилей и т.п.);
- различные способы крепления светофоров, включая специальные мачты, что приводит к возникновению «слепых зон» при подъезде к перекрестку;
- снижение яркости светофоров в течение срока их эксплуатации, а также различный уровень яркости их сигналов;
- установка временных светофоров при проведении ремонта улиц.

Перечисленные обстоятельства требуют создания сложных алгоритмов, позволяющих распознать светофоры на изображении, отфильтровать их, а также вычислить их координаты и сопоставить с картой. Карта необходима для правильного выбора светофоров, чтобы принять решение о пересечении перекрестка. Цифровые карты содержат информацию о положении светофоров, их типах, сигналах о возможности проезда в определенных направлениях и параметрах установки. В настоящей работе мы не будем рассматривать вопросы построения и обновления цифровых карт, а сфокусируемся на задаче распознавания светофоров, под которой понимаются их обнаружение на изображении, фильтрация, определение положения и сопоставление с картой.

В качестве измерений выступают 3D-детекции нейронной сети, полученные по изображению с каждой камеры. Детекция представляет собой результат работы нейронной сети в виде кубоида (прямоугольного параллелепипеда), обрамляющего объект сцены в пространстве. Внутренние параметры используемых камер значительно отличаются, в результате получаются изображения с разным разрешением, что усложняет задачу их сопоставления друг с другом, а также их фильтрацию и на последнем этапе сравнение с картой.

В статье предложены модифицированные алгоритмы объединения данных с двух камер, а кроме того, подходы к построению фильтров для оптимального оценивания

положения светофоров. Рассматривается эффективность разных комбинаций методов с использованием натуральных данных.

Статья состоит из следующих разделов: анализ существующих подходов, постановка задачи и этапы ее решения – детектирование, трекинг данных во времени, сопоставление изображений, полученных с каждой камеры, их обработка с помощью UKF, определение класса сигнала, сравнение реального положения светофора с указанным в карте, порядок принятия решения о проезде. В конце статьи приводится оценка точности предложенного подхода.

Анализ существующих подходов

Задача идентификации светофоров может ограничиваться собственно их распознаванием на изображении и классификацией, например в зависимости от времени суток (ночь [5], день [6]) или размеров [6]. Как правило, в исследованиях на эту тему подробно рассматриваются архитектуры используемых при этом нейронных сетей, а точность решения оценивается исключительно на основе классических метрик распознавания объектов. Ряд других работ посвящен решению задачи целиком – от получения изображения и сравнения его с картой [7, 8] до установления правил проезда перекрестка в режиме автопилота; в них акцент делается не столько на архитектуре нейронных сетей, сколько на алгоритмах постпроцессинга данных. Настоящая статья относится к последнему типу, в ней описывается полный цикл решения задачи, при этом особое внимание уделяется вопросам трекинга, фильтрации и сопоставления результатов, полученных после обработки изображений нейронной сетью.

Многообразие методов распознавания светофоров можно классифицировать по набору следующих признаков:

- применяемые датчики – камеры (одна [5] или несколько [7]), лидары [9];
- подходы к распознаванию – глубокие нейронные сети или классические методы обработки изображений [1];
- окружающая обстановка – городская среда [10] или автострада [6];
- типы светофоров – автомобильные или трамвайные [8];
- постановка задачи – идентификация светофоров на изображении [6, 11, 12] или сравнение с картой [8].

Финальное решение задачи, как правило, складывается из следующих этапов:

- обнаружение и классификация сигналов светофоров на имеющемся изображении;
- трекинг светофоров между кадрами;
- сопоставление изображений, полученных разными камерами;
- фильтрация изображений;
- сопоставление полученных координат светофоров с данными карты.

В работе [8] анализируется задача распознавания сигналов светофоров для трамвайного автопилота. Трамвай оснащен высокоточным лидаром и несколькими фронтально направленными камерами. Обнаружение и классификация светофоров производятся на основе аппарата нейронных сетей, при этом на изображении предварительно формируются области расположения светофоров исходя из данных цифровой карты. Каждая камера независимо от других распознает светофоры с помощью нейронной сети архитектуры Single Shot Detector. Затем результаты, полученные с нескольких камер (в общем случае – трех), объединяются и фильтруются на осно-

ве скрытой марковской модели, а итоговые данные используются планировщиком, чтобы принять решение об остановке или проезде. Особенности работы является применение высокоточного приемника спутниковой навигации, цифровой карты на основе данных лидара, а также распознавание светофоров только трамвайного типа.

В [7] рассматривается идентификация светофоров, занимающих на изображении небольшую площадь (каждый менее 10 пикселей). Предложен подход к распознаванию сигналов классических трехцветных светофоров с нанесенными на них знаками направления движение (стрелками). Авторы используют стереозрение и модель движения объекта для решения задачи в условиях предельно малых значений площадей сигналов. Задача решается в несколько этапов. На первом задействуется детектор на основе нейросети архитектуры YOLO, на вход которого поступает фрагмент верхней части изображения (поскольку светофоры располагаются на мачтах), а на выходе формируется 2D-область светофоров. Далее с помощью специальной нейронной сети классифицируются 2D-детекции, полученные на изображении. Для компенсации ошибок детектора выполняется процедура межкадрового слежения (трекинга). Каждая новая 2D-детекция учитывает историю (если она есть) ранее полученных координат расположения светофоров, что повышает вероятность правильного прогноза. Положение 2D-детекций светофоров предсказывается посредством стереозрения и данных инерциальной и одометрической систем. Такие факторы, как мерцание светофоров (из-за разных частот съема информации и собственной частоты светофоров), их небольшой размер на изображении, наличие деревьев и теней, требуют привлечения нейронной сети. На ее вход подаются два изображения – с детекцией и полученное на основе модели движения объекта. На выходе формируется окончательная 2D-область светофора в кадре, а также погрешность определения его положения. Светофоры, причем исключительно автомобильного типа, идентифицируются преимущественно в пригородах, при этом их количество на перекрестках невелико.

В [9] обнаружение светофоров осуществляется с привлечением данных лидара, изображений с камеры и нейронной сети семейства YOLO, способной распознавать сигналы светофора двух классов (красный и желтый и зеленый). На выходе такой сети формируются ограничивающие области с вероятностями соответствующих классов. На основе данных лидара и цифровой карты на изображении с некоторым допуском образуются области расположения светофоров, которые далее сравниваются с полученными нейросетью 2D-детекциями. Детекции, попадающие в области, накапливаются в некотором буфере, затем при помощи методов кластеризации определяется окончательное положение светофора как центра кластера. При решении задачи были задействованы лидар и одна камера. Метод годится только для обнаружения автомобильных светофоров, при этом, как признаются сами авторы, в силу ошибок локализации фильтр кластеризации иногда необходимо сбрасывать вручную.

В [11] рассматривается решение задачи распознавания светофоров с использованием нескольких камер с разными фокусными расстояниями, а также данных приемника спутниковой навигации и инерциального модуля. На первом этапе по изображениям каждой камеры в отдельности формируются 2D-детекции светофоров посредством нейронной сети архитектуры YOLOv5. На втором этапе полученные детекции сравниваются с данными высокоточной цифровой карты. С этой целью данные с карты проецируются на изображение в виде 2D-боксов (тоже 2D-детекции), при этом известны точные параметры калибровки камеры и координаты ав-

томобилей. Формируется полный набор возможных комбинаций текущих детекций с синтезированными по карте. Далее, исходя из соответствия типов светофоров и расстояний между положениями областей, на базе специальной меры определяется «стоимость» для каждой пары детекций. Чтобы найти пары, «стоимость» которых оказывается наименьшей, применяется венгерский алгоритм оптимизации [13]. Сопоставление детекций, полученных по изображениям разных камер, осуществляется на основе скрытой марковской модели. В исследовании задействованы две камеры с разными фокусными расстояниями, высокоточный приемник спутниковых сигналов и цифровая карта. Алгоритм применим для распознавания автомобильных светофоров в условиях плотной городской застройки.

В [12] предлагается метод создания карты светофоров по информации двух камер и датчика спутниковой навигации. На первом этапе с использованием нейронной сети формируются как области расположения светофоров на изображении, так и их ориентация. Далее на основе данных GPS и карты глубины изображения определяются положения светофоров на 3D-карте. Посредством фильтра Калмана и венгерского алгоритма производится конечная фильтрация полученных измерений, благодаря которой ложные детекции удаляются с карты. В работе применяется стереозрение и решается задача классификации светофоров с учетом их ориентации.

Анализ приведенных исследований показал, что для распознавания светофоров требуются многоэтапные алгоритмы. Известные в настоящее время алгоритмы используют с этой целью 2D-информацию на выходе нейронных сетей. Аппаратная часть систем содержит не менее двух одинаковых (для формирования стереоизображения) камер или лидаров, что существенно облегчает задачу распознавания за счет возможности точной репроекции светофора в кадр. На этапе обнаружения чаще всего привлекается априорная информация от цифровой карты.

2D-детекция не позволяет вычислить координаты светофора только по одному изображению в отличие от 3D-сетей, способных определять глубину положения светофора на изображении и тем самым устранить этот недостаток. Кроме того, в рассмотренных выше работах уделяется мало внимания вопросу применения разнофокусных камер, которое, впрочем, существенно усложняет межкамерный трекинг светофоров в силу несоответствия эмбедингов детекций. Под эмбедингом понимается векторное представление признаков, позволяющих идентифицировать объекты сцены на основе последовательности изображений. Для 3D-сетей это особенно актуально, поскольку точность определения глубины изображения зависит от дальности. В большинстве случаев точность решения оценивается на базе классических метрик сравнения детекций на изображениях, при этом на практике, в силу большого количества классов светофоров и их неравномерного распределения на маршруте, такая метрика может не отражать действительное повышение качества работы системы при увеличении точности обнаружения светофоров определенного класса. Все это отличает представленную работу от других известных.

Постановка задачи и этапы ее решения

Система помощи водителю трамвая состоит из двух камер с разными внутренними параметрами, которые установлены в передней части трамвая и функционируют в ближней и дальней зоне. С каждой камерой связана своя система координат (СК)

хуз с центром в точке C . СК камеры обозначим индексом cam (рис. 1), а номер камеры – m . СК трамвая vr совпадает с центром переднего бампера, и ее плоскость $X^{vr}Y^{vr}$ располагается параллельно земле осью Y^{vr} вперед.

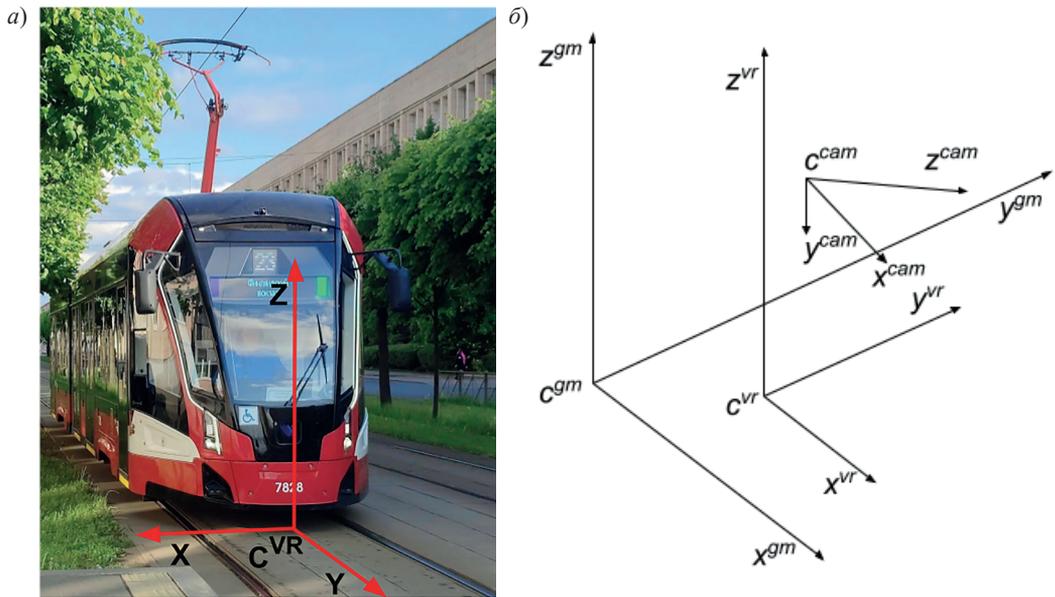


Рис. 1. Изображение трамвая и СК vr (а), gm , vr , cam (б)

Камеры считаются откалиброванными и имеющими известные внутренние K_m^{cam} и внешние R_m^{cam}, T_m^{cam} параметры установки относительно системы координат vr . K_m^{cam} является классической матрицей внутренних параметров с фокусными расстояниями f_x, f_y и координатами принципиальной точки c_x, c_y и имеет размерность 3×3 :

$$K_m^{cam} = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix}. \quad (1)$$

R_m^{cam} – матрица вращения с размерностью 3×3 , полученная для трех углов Эйлера, связывающих СК vr и cam . Вектор T_m^{cam} содержит три параметра смещения центра камеры относительно vr , в которой производится первичный анализ дорожной сцены. При этом для работы со светофорами необходимо ввести глобальную метрическую СК gm , т.е. трехмерную пространственную СК с осями x (направлена на восток), y (направлена на север), z (направлена вертикально) и центром в фиксированной точке. Навигационная система трамвая определяет его положение и ориентацию (vr центра) в СК gm . Вектор координат трамвая может быть представлен как $T_t = [x_t \ y_t \ z_t]$, а его ориентация – в виде матрицы направляющих косинусов $R_t(\alpha_t, \beta_t, \gamma_t)$. Обозначив текущее изображение как I_t , запишем цепочку преобразований из СК изображения в gm так, как показано на рис. 2.

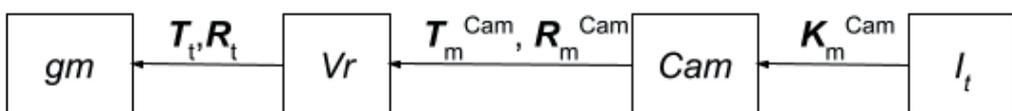


Рис. 2. Последовательность перехода из СК изображения в gm

Координаты светофоров в gm заранее известны. Представим их в виде вектора $o_b^{gm} = [x, y, z, \gamma]$, где b – порядковый номер светофора; x, y, z – координаты; γ – угол ориентации светофора. На каждом перекрестке светофоры (основные и дублирующие) объединены в группы, регулирующие движение в определенном направлении и имеющие общую стоп-линию для остановки транспортного средства. Для компактности выражений дополнительный индекс для группы светофоров вводить не будем. Каждая группа имеет вектор классов сигналов $W_b = [k_1, \dots, k_n]$, разрешающих проезд, где k_n – номера классов сигналов светофора. Например, сигнал светофора, разрешающий движение прямо, может одновременно разрешать также движение прямо и налево, прямо и направо и т.д., что ведет к необходимости ввести вектора W_b .

Распознавание светофоров производится «многоголовой» (многозадачной) нейронной сетью по изображениям I_t^m , формируемым каждой камерой. Каждая из «голов» сети имеет свои функции, например решение задачи сегментации, 3D-детектирования, создание эмбедингов и другие. Архитектура такой сети подробно описывается в работе [3]. Необходимость построения многоголовой сети обусловлена проблемой быстродействия, поскольку решение каждой задачи последовательно с точки зрения объема вычислений всегда более затратно, чем параллельные вычисления. На практике такой подход достаточно распространен, и современные нейронные сети могут одновременно выполнять десятки задач. Быстродействие решения определяется вычислением признаков низкого уровня в едином – общем для всех – блоке. Тем не менее, хотя производительность повышается, усложняется процесс обучения сети, в ходе которого может потребоваться полная разметка датасета, даже если нужно улучшить метрики только по одной из задач. На деле это вызывает затруднения, в том числе из-за необходимости получения пространственной разметки сцены на основе данных лидаров. Кроме того, разные «головы» нейронной сети в результате обучения получают разный прирост качества решения задачи. В таком случае качество решения одной задачи может деградировать на фоне улучшения качества решения другой, что связано с перестройкой весов, общих для обеих задач. Ввиду этого для обучения и построения архитектуры таких сетей нужно знать, как решения взаимосвязаны между собой. Более подробно процедура обучения и варианты ее реализации рассматриваются в [3].

Все объекты сцены опознаются как 3D-кубоиды, что на практике позволяет работать с объектами, которые частично или полностью заслонены и положение которых в связи с этим невозможно вычислить альтернативными способами, например по нижней точке. На рис. 3 показан результат работы такой сети. Нейронная сеть находит объекты: транспортные средства (включая легковые и грузовые автомобили, общественный транспорт), пешеходов, знаки дорожного движения, светофоры – и аппроксимирует их параллелепипедами (3D-детекциями).

Детекцию светофора, как и любого другого объекта, выявленного нейронной сетью в момент времени t , для камеры m обозначим как $r_j^{t,m}$, где j – ее номер. Вектор измерений $r_j^{t,m}$ содержит (см. рис. 4):

- координаты светофора на изображении $B = [u, v, du, dv]$, где u, v – координаты центра 2D-области, du, dv – ширина и высота соответственно (пиксели);
- дальность до светофора d в СК камеры (м);
- углы ориентации светофора в пространстве – тангажа (pitch), крена (roll), рысканья (yaw);

- физические размеры светофора в пространстве: w, l, h – ширина, длина и высота соответственно (м);
- уникальный идентификатор области – эмбединг (E), представляющий собой вектор размерностью 128.

Кроме того, используется еще одна нейронная сеть, выступающая в качестве классификатора, которая по имеющемуся вектору $B_j^{t,m}$ определяет класс сигнала светофора $L_j^{t,m}$. Подробнее ее архитектура будет рассмотрена ниже.

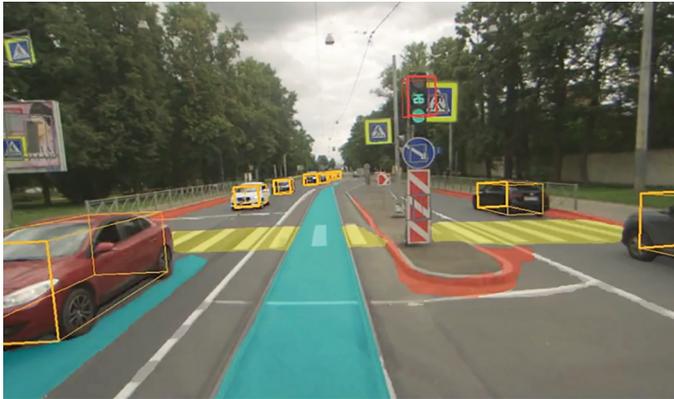


Рис. 3. Визуализация результатов распознавания дорожной сцены: 3D-детекции – машины (желтые), светофоры (красные); область сегментации – бордюры (красные), трамвайное полотно (бирюзовое)

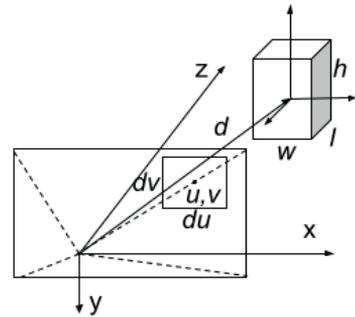


Рис. 4. 3D-детекции нейронной сети в СК камеры

Описанные выше параметры, за исключением 2D-координат и вектора эмбедингов, формируются в СК камеры. Углы ориентации объекта также определяются относительно СК камеры. Представленный вектор состояния, состоящий из 11 параметров, не считая вектора эмбедингов, является минимальным и типовым для архитектур сетей, умеющих регрессировать от 2D-измерений к 3D.

Требуется решить задачу идентификации, т.е., располагая измерениями $r_j^{t,m}$, оценить положение светофора $\hat{s}_j^{gm} = f(r_j^{t,m})$ и $\hat{b} = \operatorname{argmin}(\hat{s}_j^{gm}, o_b^{gm})$, а затем на основе полученных результатов принять решение о возможности движения по перекрестку, если $L_j^{t,m} \in W_b$.

Задача состоит из следующих этапов:

- трекинга – сопоставления детекций светофоров между кадрами для каждой камеры в отдельности, т.е. поиск соответствия между $r_j^{t,m}$ в моменты времени t и $t + 1$;
- сопоставления детекций с разных камер (межкамерный трекинг), т.е. поиска соответствия между $r_j^{t,m}$ для t и $t+1$ камер;
- оценки положения светофоров с помощью фильтра Калмана – определения координат светофоров \hat{s}_j^{gm} по измерениям $r_j^{t,m}$;
- сравнения координат светофоров, вычисленных с помощью фильтра \hat{s}_j^{gm} , с координатами светофоров o_b^{gm} , имеющимися на карте;
- классификации – идентификации класса сигнала светофора $L_j^{t,m}$ по $r_j^{t,m}$, чтобы принять решение о возможности проезда;
- принятия решения о проезде.

Трекинг

В современных задачах компьютерного зрения [19] слежение за движением (трекинг) объектов осуществляется исключительно с использованием эмбедингов. В состав каждой детекции $\mathbf{r}_j^{t,m}$ j -го детектируемого объекта (светофора) входит эмбединг \mathbf{E}_j . Размерность вектора эмбединга кратна двум и в данном случае составляет 128 параметров. Схожесть между $\mathbf{r}_{j1}^{t,m}$ и $\mathbf{r}_{j2}^{t+1,m}$ в моменты времени t и $t+1$ выявляется за счет косинусной меры сходства, задаваемой выражением вида

$$C = \frac{\mathbf{E}_{j1} \mathbf{E}_{j2}}{|\mathbf{E}_{j1}| |\mathbf{E}_{j2}|}, \quad (2)$$

где \mathbf{E}_{j1} , \mathbf{E}_{j2} – эмбединги объектов $j1$ и $j2$ в моменты времени t и $t+1$.

Чем ближе значение C к единице, тем выше сходство между эмбедингами и, как следствие, тем более явно две области светофоров на последовательности изображений соответствуют друг другу. Чтобы найти пары соответствия между $j1$ и $j2$, попарно вычисляется косинусная мера для всех измерений, полученных в моменты времени t и $t+1$, и далее выбираются пары, имеющие максимальные значения C .

Сопоставление детекций с разных камер

Задача идентификации предполагает одновременное обнаружение как основных, так и дублирующих светофоров. Кроме того, следует заблаговременно распознавать их сигналы, чтобы при необходимости обеспечить своевременное торможение транспортного средства. С этой целью в проекте задействованы две камеры с 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, нужно объединить, т.е. найти между ними соответствие. Поиск соответствия между детекциями на изображениях с двух камер будем называть межкамерным трекингом.

Объединение данных является нетривиальной задачей в силу ряда факторов и может быть осуществлено на базе следующих подходов.

1. Применение межкадрового трекинга (описанного выше). Это широко распространенный метод, который не зависит от параметров настроек камеры. Однако эмбединги светофоров могут сильно различаться, поскольку камеры имеют разные поле зрения и угол обзора, что приводит к многочисленным ложным сопоставлениям.
2. Вычисление параметров пересечения двух боксов на основе метрики IoU (Intersection over Union) путем переноса результатов распознавания с одного изображения на другое. Это понятная и хорошо интерпретируемая с точки зрения геометрии метрика, которая востребована на практике. Вместе с тем требуется высокоточная калибровка обеих камер, а также знание глубины изображения для переноса $\mathbf{B}_j^{t,m}$. Кроме того, $\mathbf{B}_j^{t,m}$ с разных камер не всегда накладываются друг на друга, что приводит к большому количеству ложных совпадений, особенно если объекты стоят близко друг к другу.
3. Вычисление IoU для 3D-детекций. Нейронная сеть выдает 3D-кубоиды, поэтому можно использовать эту информацию для проверки пересечения. Очевидно,

это позволяет уменьшить двойственность, когда объекты в кадре расположены рядом. Тем не менее в силу того что глубина изображения объектов, полученного с разных камер, может сильно отличаться (из-за большой разницы в фокусном расстоянии), кубоиды в пространстве могут не пересекаться совсем или, наоборот, фиксируется их ложное пересечение.

4. Расчет евклидова расстояния между центрами объектов в пространстве. Понятный и популярный подход, но из-за ошибочно предсказанной глубины изображения объекты также могут не совпадать.
5. Расчет расстояния на основе дистанции Махаланобиса, что по смыслу близко к евклидову расстоянию, но позволяет учесть погрешность определения глубины изображения объектов для каждой камеры и ковариацию (шум) определения глубины нейронной сетью.
6. Расчет дистанции между центрами $\mathbf{B}_j^{t,m}$ путем переноса координат центров выделенных областей на одно изображение и вычисления расстояния в СК изображения. Довольно простой и понятный подход, однако учет разницы в пикселях для удаленных объектов на разных камерах приводит к необходимости введения адаптивного порога, на который влияет размер области. Кроме того, метрика зависит от параметров обеих камер.

Как видим, способов объединения данных достаточно много. Чтобы выбрать оптимальную стратегию, следует учитывать следующие факторы: метрики должны выдавать нормализованные значения в диапазоне от 0 до 1, при использовании нескольких метрик необходимо определить коэффициенты их влияния.

По результатам 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 соответствия C_{res} , рассчитываемую на основе взвешенной суммы:

$$C_{res} = w_{emb} \cdot C(\mathbf{E}_j^{t,m}, \mathbf{E}_j^{t,m+1}) + (1 - w_{emb}) \cdot IOU(\mathbf{B}_j^{t,m}, \mathbf{B}_j^{t,m+1}), \quad (3)$$

$$IOU(\mathbf{B}_j^{t,m}, \mathbf{B}_j^{t,m+1}) = \frac{\mathbf{B}_j^{t,m} \cap \mathbf{B}_j^{t,m+1}}{\mathbf{B}_j^{t,m} \cup \mathbf{B}_j^{t,m+1}}. \quad (4)$$

В результате обучения на имеющемся размеченном датасете были получены значения оптимальных весов, которые для формулируемой задачи находятся на уровне $w_{emb} = 0,6 \dots 0,7$.

Оценка положения светофоров с помощью фильтра Калмана

Результаты, получаемые непосредственно с помощью нейронной сети, часто содержат много случайных неточностей или «шума», что требует их дополнительной фильтрации. С этой целью применяется фильтр Калмана. В статье описывается два подхода к решению задачи на основе расширенного (ЕКФ) и сигма-точечного (УКФ) фильтров Калмана. Поскольку задействованы две камеры, то обновление фильтров Калмана происходит по результатам двух измерений, полученных посредством межкамерного трекинга.

Построение ЕКФ. Вектор состояния системы состоит из двух параметров $\mathbf{s}_t^{gm} = [x_t^{gm}, y_t^{gm}]$. Вектор измерений представим в виде $\mathbf{z}_t = [\rho_t, \delta_t]$, где ρ_t – расстояние

до светофора, δ_i – угол между осью ОУ (в СК vr) и светофором. Дальность и угол на светофор определяются в несколько этапов:

- вычисляются координаты светофора в СК vr с привлечением координат центра кубоида на изображении $[u, v]$, глубины изображения до центра d , входящие в состав вектора измерений $\mathbf{r}_j^{l,m}$:

$$\begin{pmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \end{pmatrix} = d \begin{pmatrix} \frac{u - c_x}{f_x} \\ \frac{v - c_y}{f_y} \\ 1 \end{pmatrix}, \quad (5)$$

$$\begin{pmatrix} x_{VR} \\ y_{VR} \\ z_{VR} \end{pmatrix} = \mathbf{R}_m^{cam} \begin{pmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \end{pmatrix} + \mathbf{T}_m^{cam}; \quad (6)$$

- рассчитываются дальности по формуле $\rho_t = \sqrt{x_{vr}^2 + y_{vr}^2}$;
- определяется угол в СК vr по формуле $\delta_i = a \tan\left(\frac{-x_{vr}}{y_{vr}}\right)$.

Уравнение состояния ЕКФ содержит константную модель

$$\mathbf{s}_{t+1}^{gm} = \mathbf{s}_t^{gm} + \mathbf{w}_t^{gm}, \quad (7)$$

где \mathbf{w}_t^{gm} – шум модели.

Уравнение измерений будет иметь вид

$$\mathbf{z}_t = \mathbf{H}(s_t^{gm}) + \mathbf{v}_t, \quad (8)$$

где

$$\mathbf{H}(s_t^{gm}) = \begin{bmatrix} \sqrt{(x_t^{gm} - l_{t,x}^{gm})^2 + (y_t^{gm} - l_{t,y}^{gm})^2} \\ -a \tan\left(\frac{x_t^{gm} - l_{t,x}^{gm}}{y_t^{gm} - l_{t,y}^{gm}}\right) \end{bmatrix}. \quad (9)$$

Здесь \mathbf{l}_t^{gm} – координаты vr центра трамвая в СК gm (компоненты вектора \mathbf{T}_t), вырабатываемые навигационной системой; \mathbf{v}_t – шум измерений.

Надо заметить, предложенный подход имеет недостатки: так, при проецировании 3D-детекции светофора в кадр после фильтрации его положение существенно отличается от первоначального, полученного нейронной сетью на этом же кадре, при этом величина погрешности сильно зависит от дальности до объекта. Для уменьшения этой зависимости вводятся предположения, учитывающие непосредственное положение 2D-детекции объекта на изображении, и модернизируется решение, чтобы учесть сильносвязную схему комплексирования.

Таким образом, фильтрацию показаний светофоров можно проводить с использованием UKF на основе уравнения, аналогичного (8):

$$\mathbf{z}_t = \mathbf{H}(s_t^{gm}) + \mathbf{v}_t. \quad (10)$$

Вектор состояния системы в СК gm содержит следующие параметры: $s_t^{gm} = [x_{gm}, y_{gm}, z_{gm}, yaw_{gm}, w, l, h]$. При этом в качестве измерений выступают $\mathbf{z}_t = [u, v, d, yaw, w, l, h]$ составляющие расширенного вектора измерений \mathbf{r}_j^t . Уравнение состояния системы остается неизменным в силу статичности расположения светофора.

Первые три координаты вектора s_t^{gm} (3D-положения центра кубоида) связаны с vr внешними матрицами, формируемыми навигационной системой:

$$\begin{pmatrix} x_{vr} \\ y_{vr} \\ z_{vr} \end{pmatrix} = \mathbf{R}_t^{-1} \begin{pmatrix} x_{gm} \\ y_{gm} \\ z_{gm} \end{pmatrix} - \mathbf{T}_t. \quad (11)$$

Далее из СК vr с помощью известных матриц формируются измерения в СК камеры, которые позволяют определить составляющие u, v, d вектора измерений \mathbf{r}_j^t :

$$\begin{pmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \end{pmatrix} = \mathbf{R}^{cam-1} \begin{pmatrix} x_{vr} \\ y_{vr} \\ z_{vr} \end{pmatrix} - \mathbf{T}^{cam}, \quad (12)$$

$$\begin{pmatrix} \hat{u} \\ \hat{v} \\ \hat{d} \end{pmatrix} = \mathbf{K}^{cam} \begin{pmatrix} x_{cam} \\ y_{cam} \\ z_{cam} \end{pmatrix}. \quad (13)$$

Угол yaw получается путем перехода из gm в vr по формуле, представленной выше, w, l, h идентичны. Так как функция $h(x)$ содержит множество преобразований и нелинейности, расчет якобиана становится невозможным, поэтому процедура реализуется в виде UKF.

Недостаток применения ЕКФ заключается в сложности точного описания якобиана при всех возможных трансформациях состояний и измерений, с которым хорошо справляется UKF благодаря взвешиванию и преобразованию сигма-точек для отражения нелинейной динамики системы.

На практике матрица измерений R , используемая в уравнениях фильтра Калмана, зависит от глубины изображения объектов d на сцене. По этой причине задействуется функция вида $R(d)$, позволяющая задать уровень СКП шумов измерений как функция от значения глубины. Вид этой функции может быть получен на основе представленных ниже графиков точности определения положения (АТЕ) и углов ориентации (АТО) в зависимости от расстояния d для объектов типа «светофоры» (рис. 5). Параметры АТЕ и АТО рассчитываются как корень квадратный из суммы квадратов погрешности выработки координат и углов ориентации объектов каждого типа на основе эталонных данных о нахождении объектов, полученных посредством лидаров.

Анализ графика показывает, что на средних дистанциях (до 50 м) положение светофора определяется с точностью до нескольких метров, которая практически не изме-

няется. Напротив, углы ориентации вычисляются с погрешностью на уровне 0,5 рад, которая растет при приближении к границам требуемого диапазона дальностей.

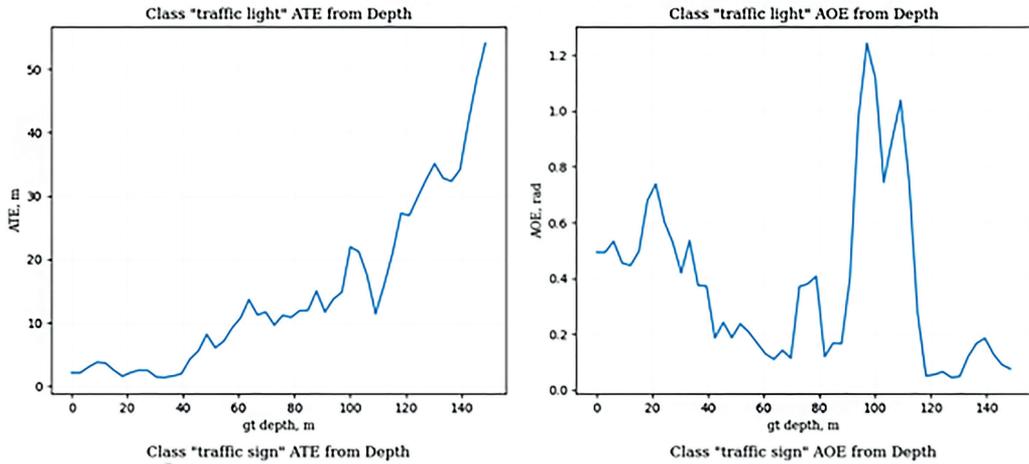


Рис. 5. Зависимость точности определения положения (ATE) и ориентации (АТО) светофоров от дальности d

Сопоставление координат светофоров

По результатам работы UKF формируется вектор координат светофоров \hat{s}_j^{gm} , который далее сравнивается с их координатами на цифровой карте.

Цифровая карта представляет собой базу данных, в которой хранятся координаты местоположения светофоров, классы сигналов для проезда по тому или иному направлению и другая служебная информация. Координаты светофоров определяются геодезическими средствами общего назначения или на основе данных высокоточной съемки (ортофотопланов), получаемых от профильных кадастровых служб города. Во втором случае с помощью специальных геодезических программ пересчитываются координаты объектов из локальных СК в глобальные WGS84.

Сопоставление векторов координат осуществляется с использованием расстояния Махаланобиса, которое вычисляется по формуле

$$d_{ib} = \sqrt{(s_i^{gm} - o_b^{gm})P^{-1}(s_i^{gm} - o_b^{gm})}, \quad (14)$$

где o_b^{gm} – вектор координат для b -го светофора карты; s_i^{gm} – вектор координат i -го светофора, полученный в результате фильтрации; P^{-1} – матрица ковариации погрешностей определения координат светофора. Для найденных дистанций составляется матрица «стоимости» со значениями метрики обратно пропорциональной расстоянию Махаланобиса и по венгерскому алгоритму отбираются лучшие пары «измерение – карта».

Классификация сигналов светофоров и принятие решения

Задача классификации решается с использованием глубокой нейронной сети. На вход сети подается фрагмент изображения светофора, формируемый по измерениям $I_t^m(\mathbf{B}_j^m)$, а на выходе сети получаем информацию о классе сигнала светофора $L_j^{l,m}$ и его ориентации e .

Классификатор содержит 25 типов сигналов светофора, часть из которых показана на рис. 6. Ориентация светофора является бинарной величиной и характеризует классы как «направлен на нас – развернут». Введение этого признака позволяет независимо фильтровать фрагменты по их ориентации и отбраковывать те, которые не являются определяющими при принятии решения.

Проще говоря, архитектура сети состоит из двух элементов – извлечения признаков и предсказания классов и ориентации. Для извлечения признаков используется нейронная сеть архитектуры RegNetY [20], позволяющая сформировать вектор признаков и обладающая высоким быстродействием. Сеть содержит два выхода, каждый из которых реализован с привлечением полносвязных слоев. Поскольку большая часть данных при обучении нейронной сети относится к типу «направлен на нас», данные по которому сильно не сбалансированы, то при обучении такой сети применяется техника Focal Loss [21]. С ее помощью можно динамически масштабировать потери функции кросс-энтропии для несбалансированных классов. Масштабный коэффициент уменьшается до нуля по мере роста уверенности нейронной сети в правильности предсказанного класса.

Чтобы разметить фрагменты изображений, содержащих светофоры, применяется разветвленная система классов, позволяющая учитывать тип светофора (автомобильный, трамвайный, пешеходный), ориентацию (направление), количество секций (пяти-, трех-, двухсекционный), а также специальные символы и их комбинации («прямо», «налево», «прямо и налево» и т.д.). Нейронная сеть обучается на основе специально собранного датасета по методу обучения с учителем. Для создания категории «класс не определен» используются изображения, из которых случайным образом вырезаются 2D-фрагменты, подаваемые далее в классификатор.

Примеры светофоров разных классов приведены на рис. 6.



Рис. 6. Некоторые типы сигналов светофоров, распознаваемые классификатором

Класс сигнала устанавливается путем проверки соответствия данных, полученных на выходе нейронной сети, информации о типе сигнала, записанной для b -го светофора в цифровую карту:

$$\gamma + \delta\gamma \gg e \geq \gamma - \delta\gamma,$$

$$L_j^{t,m} \in W_b,$$

где $\delta\gamma$ – допуск в определении ориентации.

Если класс, определенный нейросетью, соответствует имеющемуся в карте, принимается решение о проезде, если нет, то считается, что сигнал светофора запрещающий. Такой подход позволяет достаточно гибко решать задачу пересечения перекрестка, движение через который регулируется светофорами сразу в нескольких направлениях.

Оценка точности

Точность работы системы распознавания светофоров может оцениваться на нескольких уровнях. На первом – по изображению. Существующие открытые датасеты содержат большое количество фото- и видеоизображений, по которым и оценивается точность данных, выработанных нейросетевыми детекторами и классификаторами светофоров. Так, в датасетах LISA [14] и DriveU [15] имеются не только изображения, но и аннотированные видеопоследовательности, снятые при разном освещении и различных погодных условиях. Разметка данных предоставляется с учетом релевантности (регулирует светофор текущее направление движения или нет). Качество работы алгоритмов обычно зависит от того, как функционируют их детектирующая (локализирующая) и классифицирующая части. Для оценки точности локализации используется мера Жаккара [16] с пороговым уровнем для установления факта совпадения детекции и разметки. Далее формируются три типа событий: TP (True Positive) – детектор верно локализовал светофор, FP (False Positive) – детектор нашел светофор там, где его нет, FN (False Negative) – детектор не нашел светофор, хотя он есть. Событие TN (True Negative) – светофора нет и детектор его не нашел – не вводится. На основе базовых характеристик рассчитываются производные, при этом точность на видеопоследовательности может оцениваться посредством функции MOTA (Multiple Object Tracking Accuracy), которая позволяет определить общую точность отслеживания и детекции. Оценка качества классификации производится путем вычисления по каждому классу матрицы ошибок и производных метрик, таких как полнота, точность и F-мера.

В настоящей работе привлечен гораздо более широкий перечень классов светофоров, включая специальные трамвайные (светодиодные), трамвайные (двухсекционные) и пешеходные, которые отсутствуют в открытых датасетах, но встречаются на маршрутах в 70-95% случаев. Попытка улучшить классические показатели сети только на автомобильных светофорах на практике не только не приведет к росту качества работы системы, но, скорее, ее ухудшит. С этой точки зрения открытые датасеты не годятся для оценки точности работы сети в частности и алгоритма в целом. В связи с этим предлагается другой подход – учитывающий потребительские метрики качества и показывающий их изменение в зависимости от конфигурации системы и методов обработки измерений.

Оценка точности работы системы распознавания может производиться и в комплексе с системой управления [17, 18]. Такой подход прежде всего востребован в автопилотируемых решениях для определения уровня безопасности беспилотного вождения в городской среде. Обычно с этой целью используют тестовые маршруты, на которых есть перекрестки. Правильная идентификация сигнала светофора означает остановку на перекрестке при запрещающем сигнале и проезд при разрешающем. Автоматически оценить точность работы системы помощи водителю можно, если считать действия водителя эталонными. Критерий не является строгим и справедлив в ряде ситуаций – сборки стендов, тестирования нового функционала, управления трамваем водителем с большим стажем. Качество работы системы будет определяться по своевременности и правильности вырабатываемых ею сигналов управления – «движение» или «остановка».

Вводятся целевые (произведенные системой) и фактические (произведенные водителем) действия системы:

- остановка при запрещающем сигнале;
- движение при разрешающем сигнале.

В контексте пересечения перекрестка наиболее критичным действием является «остановка», поскольку при его невыполнении возникает риск дорожно-транспортного происшествия. Сформируем матрицу соответствия событий ситуации (см. табл. 1).

Т а б л и ц а 1

Матрица соответствия событий

Событие	Целевое действие	Фактическое действие системы
TN	ДВИЖЕНИЕ	ДВИЖЕНИЕ
TP	ОСТАНОВКА	ОСТАНОВКА
FN	ОСТАНОВКА	ДВИЖЕНИЕ
FP	ДВИЖЕНИЕ	ОСТАНОВКА

Т а б л и ц а 2

Сведения о маршрутах

Маршруты	Протяженность одного маршрута, км
36 маршрут	8 заездов, 17,43 км
24 маршрут	3 заезда, 14,67 км
27 маршрут	1 заезд, 18,04 км

Исходя из табл. 1 формируются события: TP – остановка при запрещающем сигнале; TN – движение при разрешающем сигнале; FP – остановка при разрешающем сигнале; FN – движение при запрещающем сигнале, которое относится к самой критичной с точки зрения безопасности ошибке, поскольку разрешает проезд на красный свет.

Оценка точности проводилась путем анализа данных, собранных на трамвайных маршрутах Санкт-Петербурга (табл. 2), для трех алгоритмов:

- задействована одна камера и ЕКФ (A_1);
- задействована одна камера и UKF (A_2).
- задействованы две камеры и UKF (A_3).

Т а б л и ц а 3

Оценка точности представленных алгоритмов

События, количество раз	Движение в городе		
	A_1	A_2	A_3
TP	160	181	192
TN	167	183	180
FP	83	57	60
FN	27	16	5
Точность, %	65,84	76,05	76,19
Полнота, %	85,56	91,88	97,46

В табл. 3 приведены результаты оценки точности системы на натурных данных, полученные посредством разных алгоритмов. Рассчитываются как базовые (TP, TN, FP, FN), так и производные значения параметров (точность, полнота), характеризующие точность работы алгоритма. Параметр «точность» показывает, сколько предсказанных ситуаций «остановка» совпало с действиями водителя. Параметр «полнота» демонстрирует, сколько ситуаций, требующих остановки, было правильно предсказано алгоритмом. Исходя из логики решаемой задачи, приоритетным является пара-

метр «полнота», поскольку проезд перекрестка на красный сигнал светофора связан с большими рисками. Как видно из анализа, при использовании одной камеры (алгоритмы A_1 , A_2) повышается эффективность второго алгоритма по сравнению с первым благодаря более «тесной» (на уровне изображения) схеме фильтрации данных, за счет чего точность предсказания возрастает на 15%, а полнота – на 7%. При этом две камеры позволяют увеличить параметр полноты на 6%, что объясняется возможностью лучшего обнаружения дальних дублирующих светофоров.

На рис. 7 представлены результаты работы алгоритма на различных перекрестках Санкт-Петербурга. Параллелепипедами обозначены все объекты, распознанные нейронной сетью (желтым цветом отмечен автомобиль, красным и зеленым – светофоры, синими точками – положение светофоров в кадре по данным карты, белые линии показывают связи между детекциями и картой).



Рис. 7. Визуализация результатов работы алгоритма

Заключение

Проблемой идентификации светофоров ученые занимаются уже достаточно продолжительное время, при этом абсолютно надежного решения до сих пор не существует в силу большого количества внешних факторов, влияющих на процесс. Между тем использование нескольких камер или лидаров, глубоких нейронных сетей и высокоточных карт позволяет существенно повысить качество решения задачи. В статье описывается алгоритм распознавания светофоров в реальном времени и с высокой точностью, предназначенный для современных систем помощи водителю трамвая. Результаты исследования продемонстрировали, что совместное применение предложенного алгоритма и системы управления трамваем с двумя камерами позволяет достичь точности 76,19% и полноты 97,46%. Ошибки системы, существующие в настоящее время, связаны прежде всего с трудностями при обнаружении специализированных трамвайных светофоров в силу их малой распространенности. Особенно это относится к сигналу поворота и снижению яркости светофора за время эксплуатации. Внедрение предлагаемого алгоритма на практике будет способствовать снижению требований к разметке карты и обеспечит высокую точность распознавания.

ЛИТЕРАТУРА

1. Хартов В.Я., Гарифуллин Д.Б. Анализ решений для задачи распознавания автомобильных светофоров // Приоритетные направления развития науки и образования: сборник статей VI Международной научно-практической конференции в 2-х частях. Часть 1. Пенза: «Наука и Просвещение», 2019. С. 64–76. EDN ZBNPVB.

2. **Tran, Tai Huu-Phuong; Pham, Cuong Cao; Nguyen, Tien Phuoc; Duong, Tin Trung; Jeon, Jae Wook**, Real-time traffic light detection using color density, *IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia)*, 2016, 1–4. doi:10.1109/ICCE-Asia.2016.7804791.
3. **Гужва Н.С., Прун В.Е., Постников В.В., Любанов М.Г., Садеков Р.Н., Шоломов Д.Л.** Использование 3D-сетей для «предсказания» моделей поведения транспортных средств в задаче беспилотного движения трамвая // 29 Санкт-Петербургская международная конференция по интегрированным навигационным системам. Санкт-Петербург, 2022. С. 304–310.
4. **Гужва Н.С., Али Б., Бакулев К.С., Садеков Р.Н., Шолохов А.В.** Оценка точности местоположения трамвая в условиях высотной застройки по данным визуально-геоинформационных систем // XXX Санкт-Петербургская международная конференция по интегрированным навигационным системам. Санкт-Петербург, 2023. С. 16–21.
5. **Али Б., Садеков Р. Н., Цодокова В.В.** Алгоритмы навигации БПЛА с использованием систем технического зрения // Гирскопия и навигация. Том 30. №4 (119). 2022. С. 87–105. DOI 10.17285/0869-7035.00105.
6. **Haltakov, V., Mayr, J., Unger, C., Ilic, S.**, Semantic Segmentation Based Traffic Light Detection at Day and at Night, *Pattern Recognition. DAGM 2015. Lecture Notes in Computer Science*, 2015, vol. 9358, https://doi.org/10.1007/978-3-319-24947-6_37.
7. **Weber, M., Wolf, P., and Zollner, J.M.**, Deeptlr: A single deep convolutional network for detection and classification of traffic lights, *IEEE IV Intelligent Vehicles Symposium*, Gotenburg, Sweden, June 19-22, 2016, pp. 342–348.
8. **Behrendt, K., Novak, L., Botros, R.**, A deep learning approach to traffic lights: Detection, tracking, and classification, *IEEE International Conference on Robotics and Automation (ICRA)*, 2017, pp. 1370–1377.
9. **Palmer, A.W., Sema, A., Martens, W., Rudolph, P., and Waizenegger, W.**, The Autonomous Siemens Tram, *IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*, Rhodes, Greece, 2020, pp. 1–6, doi: 10.1109/ITSC45102.2020.9294699.
10. **Possatti, L.C. et al.**, Traffic Light Recognition Using Deep Learning and Prior Maps for Autonomous Cars, *International Joint Conference on Neural Networks (IJCNN)*, 2019, pp. 1–8.
11. **Wang, Q.; Zhang, Q.; Liang, X.; Wang, Y.; Zhou, C.; Mikulovich, V.I.**, Traffic Lights Detection and Recognition Method Based on the Improved YOLOv4 Algorithm, *Sensors*, 2022, 22, 200, <https://doi.org/10.3390/s22010200>.
12. **Wu, S. et al.**, aUToLights: A Robust Multi-Camera Traffic Light Detection and Tracking System, *arXiv preprint arXiv*, 2023, 2305, 08673.
13. **Mentasti, S., Simsek, Y.C., and Matteucci, M.**, Traffic lights detection and tracking for HD map creation, *Front. Robot. AI*, 2023, 10:1065394, doi: 10.3389/frobt.2023.1065394.
14. https://en.wikipedia.org/wiki/Hungarian_algorithm.
15. **Jensen, M.B., Philipsen, M.P., Møgelmoose, A., Moeslund, T.B., Trivedi M.M.**, Vision for looking at traffic lights: Issues, survey, and perspectives, *IEEE Transactions on Intelligent Transportation Systems*, 2016, no. 17, pp. 1800–1815.
16. **Fregin, A., Muller, J., Krebel, U., and Dietmayer, K.**, The DriveU Traffic Light Dataset: Introduction and Comparison with Existing Datasets, *IEEE International Conference on Robotics and Automation (ICRA)*, 2018, pp. 3376–3383.
17. https://en.wikipedia.org/wiki/Jaccard_index
18. **Kim, J., Cho, H., Hwangbo, M., Choi, J., Canny, J., and Kwon, Y.P.**, Deep Traffic Light Detection for Self-driving Cars from a Large-scale Dataset, *21st International Conference on Intelligent Transportation Systems (ITSC)*, Maui, HI, USA, 2018, pp. 280–285, doi: 10.1109/ITSC.2018.8569575.
19. **Li, Z., Zeng, Q., Liu, Y., Liu, J., Li, L.**, An improved traffic lights recognition algorithm for autonomous driving in complex scenarios, *International Journal of Distributed Sensor Networks*, 2021;17(5), doi:10.1177/15501477211018374.
20. **Chelva, M.S., Halse, S.V., and Ratha, B.K.**, Object tracking in real time embedded system using image processing, *International Conference on Signal Processing, Communication, Power and Embedded System (SCOPEs)*, Paralakhemundi, India, 2016, pp. 1840–1844, doi: 10.1109/SCOPEs.2016.7955763.
21. **Radosavovic, I. et al.**, Designing network design spaces, *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 10428–10436.
22. **Lin, T.Y. et al.**, Focal loss for dense object detection, *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2980–2988.

Guzhva, N.S. and Sadekov, R.N. (National University of Science and Technology MISIS, Cognitive Technologies, Moscow)

Neural Network Based Algorithms for Traffic Lights Identification in Multi-Camera Driving Aid Systems, *Giroskopiya i Navigatsiya*, 2024, vol. 32. no. 3, pp. 47–65.

Abstract. The paper considers the problem of traffic lights identification (detection, filtering and map-matching) using successive images in active aid systems for tram drivers, equipped with multiple cameras with different focal lengths. The process of the problem solution is described in detail, from measurements (detections) formed at the neural network output for each of the cameras, and up to the results matching with a map. In contrast to other studies of this subject, the authors of this work use 3D measurements as the output data for the neural network, and unscented Kalman filter (UKF) for determining the position of the traffic lights; in addition, a new method for fusing the data from two cameras is applied. The efficiency of the proposed algorithms and its modification has been field-tested. The results of experiments have shown that the algorithm provides the accuracy of 76.19% and completeness of 97.46% when used in combination with the tram control system with two cameras.

Key words: computer vision, traffic lights detection, traffic lights recognition, monocular camera, unmanned tram.

Материал поступил 21.04.2024