

На правах рукописи



Зубов Илья Геннадьевич

**ОБРАБОТКА И АНАЛИЗ ВИДЕОДАННЫХ В СИСТЕМЕ УПРАВЛЕНИЯ
БЕСПИЛОТНОГО АВТОМОБИЛЯ**

Специальность

2.2.13 – Радиотехника, в том числе системы и устройства телевидения

Автореферат

диссертации на соискание ученой степени

кандидата технических наук

Санкт-Петербург — 2022

Работа выполнена в ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)» на кафедре «Телевидения и видеотехники».

Научный руководитель: **Обухова Наталия Александровна**

доктор технических наук, профессор, заведующий кафедрой «Телевидения и видеотехники», ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)»

Официальные оппоненты: **Луцив Вадим Ростиславович**

доктор технических наук, профессор кафедры аэрокосмических компьютерных и программных систем, ФГАОУ ВО «Санкт-Петербургский государственный университет аэрокосмического приборостроения» (ГУАП), г. Санкт-Петербург

Николаев Дмитрий Петрович

кандидат физико-математических наук, заместитель директора по науке, федеральное государственное бюджетное учреждение науки Институт проблем передачи информации им. А.А. Харкевича Российской академии наук (ИППИ РАН), г. Москва

Ведущая организация: Акционерное общество «Научно-исследовательский институт телевидения» (АО «НИИ телевидения»), г. Санкт-Петербург

Защита состоится «20» сентября 2022 г. в «14-00» часов на заседании диссертационного совета 24.2.387.01 при ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)» по адресу: 197022, Санкт-Петербург, ул. Проф. Попова, дом 5, литера Ф.

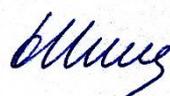
С диссертацией можно ознакомиться в библиотеке ФГАОУ ВО «Санкт-Петербургский государственный электротехнический университет «ЛЭТИ» им. В. И. Ульянова (Ленина)» и на сайте университета <https://etu.ru> в разделе «Наука» - «Подготовка кадров высшей квалификации» - «Объявления о защитах».

Отзывы на автореферат в двух экземплярах, заверенные печатью, просим направлять по адресу: 197022, Санкт-Петербург, улица Профессора Попова, д. 5, литера Ф.

Автореферат разослан «27» июня 2022 г.

Ученый секретарь

диссертационного совета 24.2.387.01



Шевченко М. Е.

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

Актуальность работы.

Дорожно-транспортные происшествия являются главной причиной смертности среди детей и молодых людей. Как правило, причинами аварий являются нарушения правил дорожного движения и недостаточная квалификация водителей. Одним из наиболее эффективных способов решения проблем, вызванных влиянием «человеческого фактора», является широкое внедрение технологий автономного вождения.

Система технического зрения является одним из основных модулей беспилотных транспортных средств (ТС). Главной задачей такой системы является анализ окружающего пространства беспилотного ТС как 3D-сцены, что позволяет оценивать взаимодействие объектов и предсказывать их возможное поведение в физическом пространстве сцены.

Благодаря развитию вычислительной и телевизионной техники задача анализа окружающего пространства ТС как 3D-сцены на основе видеоряда становится все более актуальной. Для анализа реального взаимодействия трехмерных объектов в видеопотоке необходимо определять и анализировать информацию об их траектории и пространственной ориентации. Сложность траекторных измерений состоит в том, что в отличие от активных датчиков видеопоток монокулярной камеры не предоставляет информацию о пространственной ориентации объектов.

подавляющим большинством объектов, окружающих беспилотный транспорт, являются другие ТС. В данной работе под объектами интереса подразумеваются ТС (автомобили и автобусы). С учетом того, что ТС имеет жесткую структуру и движется по плоскости дорожного полотна, его положение определено как $p = (x, y, \alpha, m)$, где (x, y) – координаты точки проекции объекта, m – масштаб, α – ракурс (угол рыскания).

Таким образом, для оценки пространственной ориентации 3D-объекта в видеопотоке необходимо локализовать и классифицировать этот объект, определить его габаритные размеры и ракурс.

Методы определения ракурса объекта на основе изображения изучены слабо в сравнении с такими задачами, как локализация, классификация и сегментация объекта. Основными проблемами в задачах оценки ракурса ТС являются: сложность аннотирования ракурса и отсутствие широкомасштабных наборов данных, учитывающих различные условия наблюдения. Необходимый набор данных может быть получен из проекций трехмерных моделей объектов (синтетических изображений). Однако с учетом того, что особенности синтетических и реальных изображений различны, необходимо использовать дескрипторы, которые будут одинаковыми на обеих выборках. К таким дескрипторам можно отнести внешнюю форму и расположение ключевых точек объекта.

Современные методы анализа изображений обладают рядом существенных недостатков:

- классические методы не обеспечивают достаточной точности сегментации и локализации ключевых точек для задачи оценки пространственной ориентации;
- нейросетевые методы требуют больших наборов аннотированных данных и являются неэффективными из-за медленной скорости анализа.

Таким образом, актуальной является задача разработки и исследования каскадного подхода анализа видеоданных, обеспечивающего автоматическую оценку пространственной ориентации объекта, основанного на накопленных знаниях и опыте, а также обладающего возможностью к развитию в соответствии с объективными изменениями условий функционирования. Для этого также необходимо разработать новые методы сегментации изображения; локализации ключевых точек объекта интереса; определения ракурса, основанного на сопоставлении дескрипторов целевого объекта и трехмерной модели.

Цель диссертационной работы: разработка методов обработки и анализа видеоданных для систем автономного транспорта, позволяющих повысить точность оценки пространственной ориентации транспортных средств.

Основные задачи диссертационной работы.

1. Оценить возможность и эффективность использования существующих методов определения пространственной ориентации объектов на основе анализа видеоряда в прикладных ТВ системах беспилотного транспорта.

2. Разработать каскадный подход автоматического анализа видеоданных, необходимый для оценки пространственной ориентации объектов, учитывающий специфику видеоданных, полученных в сложных условиях наблюдения.

3. Разработать метод автоматической сегментации изображений, учитывающий априорную информацию о местоположении объектов, позволяющий анализировать изображения в сложных условиях видеонаблюдения.

4. Разработать метод локализации ключевых точек объекта интереса, основанный на анализе внутренних слоев СНС, обеспечивающий устойчивый анализ для различных архитектур.

5. Разработать метод автоматической оценки ракурса на основе совокупности дескриптора формы и локализации ключевых точек объекта интереса.

Методы исследования.

В работе использованы методы цифровой обработки изображений, математической статистики, машинного обучения, регрессионного анализа, методы моделирования на ЭВМ, экспериментальные исследования.

Основные положения, выносимые на защиту.

1. Каскадный подход анализа изображений, включающий в себя локализацию и сегментацию объекта интереса, локализацию ключевых точек, а также оценку ракурса, обеспечивает высокую точность сегментации и долю корректного определения ракурса на 10% выше имеющихся аналогов, таких как AAVER и HSSR.

2. Метод сегментации изображения на основе преобразования в полярную систему координат и поиска кратчайшего пути во взвешенном графе, позволяющий в автоматическом режиме выделить объект на сложном фоне с учетом априорной оценки положения объекта. Метод обеспечивает: меру сходства Жаккара на 6% выше имеющихся аналогов, таких как GrabCut и DEEPLABv3; устойчивость к шумам различного вида, таким как импульсный, мультипликативный и аддитивный белый гауссовский шум.

3. Метод локализации ключевых точек ТС, основанный на анализе внутренних слоев СНС, позволяющий повысить информативность метода локализации объектов на

изображении и обеспечивающий: долю корректно локализованных ключевых точек на 10% выше имеющихся аналогов, таких как WPOD-NET и Mask-RCNN; устойчивый анализ для различных архитектур.

4. Метод оценки ракурса ТС на основе дескриптора объекта интереса в виде сигнатуры формы и локализации ключевых точек, обеспечивающий: долю корректного определения ракурса на 5% выше, чем у аналога HSSR, устойчивость к окклюзии объекта и не требующий трудозатратной ручной аннотации.

Научная новизна.

1. Разработан новый каскадный подход анализа видеоданных, включающий в себя локализацию объектов и их ключевых точек, сегментацию изображения и определение ракурса.

2. Разработан новый метод сегментации изображений на основе преобразования в полярную систему координат и поиска кратчайшего пути во взвешенном графе, опирающийся на априорную информацию в виде локализации объекта интереса на изображении.

3. Разработан новый метод локализации ключевых точек объекта интереса на изображении на основе анализа внутренних слоев СНС.

4. Разработан новый метод определения ракурса ТС, на основе дескриптора формы и локализации ключевых точек объекта интереса.

Практическая значимость полученных результатов.

1. Разработанный каскадный подход обеспечивает повышение эффективности телевизионной системы компьютерного зрения беспилотного автомобиля, позволяя оценивать пространственную ориентацию объектов без активных датчиков, таких как лидары и радары.

2. Применение разработанного метода сегментации изображений позволило увеличить меру сходства Жаккара для задачи сегментации изображения на 6%.

3. Применение разработанного метода локализации ключевых точек объекта интереса позволило повысить информативность метода локализации объектов, а также повысить долю корректной локализации ключевых точек объекта интереса на 10%.

4. Применение разработанного метода определения ракурса позволило повысить долю корректного определения ракурса ТС на 5%.

Реализация результатов работы.

Основные результаты использованы в научно-исследовательских работах ООО «НЕКСТ», ООО «Инновационный центр «КАМАЗ», а также в учебном процессе кафедры телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета.

Апробация работы.

Основные результаты диссертационной работы докладывались и обсуждались на научных конференциях:

- Конференция ЦИФРОВАЯ ОБРАБОТКА СИГНАЛОВ И ЕЁ ПРИМЕНЕНИЕ – DSPA, Москва 2019, 2020 г.
- 15-я Международная конференция «Телевидение: передача и обработка изображений», Государственный электротехнический университет (ЛЭТИ), Санкт-Петербург, 2018 г.
- Конференция IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), Санкт-Петербург, 2019, 2020 г.
- Научно-техническая конференция молодых специалистов «Вопросы радиоэлектроники: техника телевидения», Санкт-Петербург, 2021 г.

Публикации.

Содержание и основные результаты диссертации изложены в 9 научных работах, которые включают 4 статьи, опубликованные в журналах, входящих в список рецензируемых научных журналов ВАК при Министерстве науки и высшего образования Российской Федерации. Опубликовано 5 работ в трудах научно-технических конференций, в том числе международных.

Структура и объем диссертации.

Диссертация состоит из введения, пяти глав, заключения и списка литературы. Работа содержит 134 страницы основного текста и включает 26 таблиц, 45 рисунков и содержит список литературы из 161 наименования.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во введении обоснована актуальность темы диссертационной работы, сформулированы цель и задачи исследования. Изложены научные положения, выносимые на защиту, определены их научная новизна и практическая значимость, приведены сведения об апробации и внедрении результатов работы.

В первой главе выявлены основные тенденции развития современных методов анализа и обработки видеоданных в ТВ системах беспилотного транспорта, определены функциональные особенности и сформулирован перечень требований к методам анализа видеоданных в системах беспилотного транспорта. Приведен аналитический обзор методов обработки и анализа видеоданных в современных ТВ системах беспилотного транспорта. Показана необходимость и определены основные направления разработки методов представления и анализа видеоданных для ТВ систем беспилотного транспорта. Сделана постановка задач, подлежащих решению в данной работе.

На основании проведенного обзора и анализа современных подходов обработки и анализа видеоданных в системах автономного управления автомобилей сделаны следующие выводы:

- Актуальной задачей анализа видеоданных в системе автономного управления автомобилей является оценка пространственной ориентации объектов, базой для которой является определение ракурса объектов интереса.

- Современные методы анализа изображений сфокусированы на сквозном обучении СНС для решения одновременно задач локализации, классификации, сегментации, определение ракурса и т. д. Как правило, решение нескольких задач одной моделью является целесообразным, так как в случае использования ансамбля СНС для решения

различных задач существенно уменьшается скорость анализа изображений. Сквозные подходы на основе СНС позволяют достичь значительных результатов оценки пространственной ориентации объектов в видеопотоке, однако точность таких подходов в основном зависит от наборов обучающей выборки, их объема и т. д.

- Открытые наборы данных, как правило, не включают в себя условия различной погоды, времени суток, ракурсов наблюдения, из-за чего являются слабо масштабируемыми в реалии российских дорог.

- Анализ жестких объектов, как совокупности взаимно расположенных частей, позволяет более подробно описать поведение объекта в пространстве, понять степень окклюзии объекта, оценить ракурс объекта и т. д. Однако в результате отсутствия широкомасштабных и разнородных аннотированных баз данных, такие методы не распространены.

Таким образом, для создания телевизионной системы анализа окружающего пространства автономного транспорта требуется разработать подход оценки ракурса ТС, основанный на каскадном анализе информации. Разрабатываемый подход должен: обеспечивать извлечение информации необходимой для оценки пространственной ориентации объекта; опираться на накопленные знания и опыт; работать в различных условиях наблюдения, включая сложные; анализировать изображение без существенных требований к базам данных. Исходя из перечисленных требований подхода оценки ракурса ТС необходимо разработать: метод сегментации изображений, опирающийся на априорную информацию в виде локализации объекта интереса на изображении; метод локализации ключевых точек объекта интереса на основе анализа внутренних слоев СНС; метод определения ракурса ТС, на основе дескриптора формы и локализации ключевых точек объекта интереса.

Во второй главе предложен новый каскадный подход анализа видеoinформации, обеспечивающий автоматическую оценку ракурса ТС, обладающий возможностью к развитию в соответствии с объективными изменениями условий функционирования. Произведен анализ современных методов локализации и классификации объектов на изображении, как классических, так и нейросетевых. Выполнено кросс-валидационное тестирование современных методов локализации и классификации объектов на изображении, с последующим выбором метода, удовлетворяющего условиям работы системы анализа и обработки видеоданных беспилотного ТС. В общем виде функциональная схема предложенного каскадного подхода оценки ракурса ТС изображена на рисунке 1.



Рисунок 1 – функциональная схема каскадного подхода оценки ракурса ТС.

На первом шаге производится локализация объектов интереса на изображении на основе СНС (рис. 2-б, 2-в). На основе предложенного алгоритма анализа внутренних сверточных слоев нейронной сети (НС) производится модификация архитектуры НС, позволяющая локализовать ключевые точки, принадлежащие объекту интереса (рис. 2-г). На основе априорной информации о местоположении ТС на изображении, производится сегментация изображения (рис. 2-д). Затем извлекается сигнатура (форма объекта) как зависимости расстояния от ключевой точки до границ объекта в виде функции угла. Фурье образ таких сигнатур является ключевым признаком, на основе которого производится определение ракурса ТС.

На основе трехмерных моделей производится извлечение шаблонов для каждого ракурса ТС. Определение целевого ракурса ТС происходит за счет сопоставления признаков исходного объекта и шаблонов, полученных на основе 3D-моделей. Ракурс целевого ТС определяется, как ракурс шаблона, имеющего наибольший коэффициент корреляции (рис. 2-е).

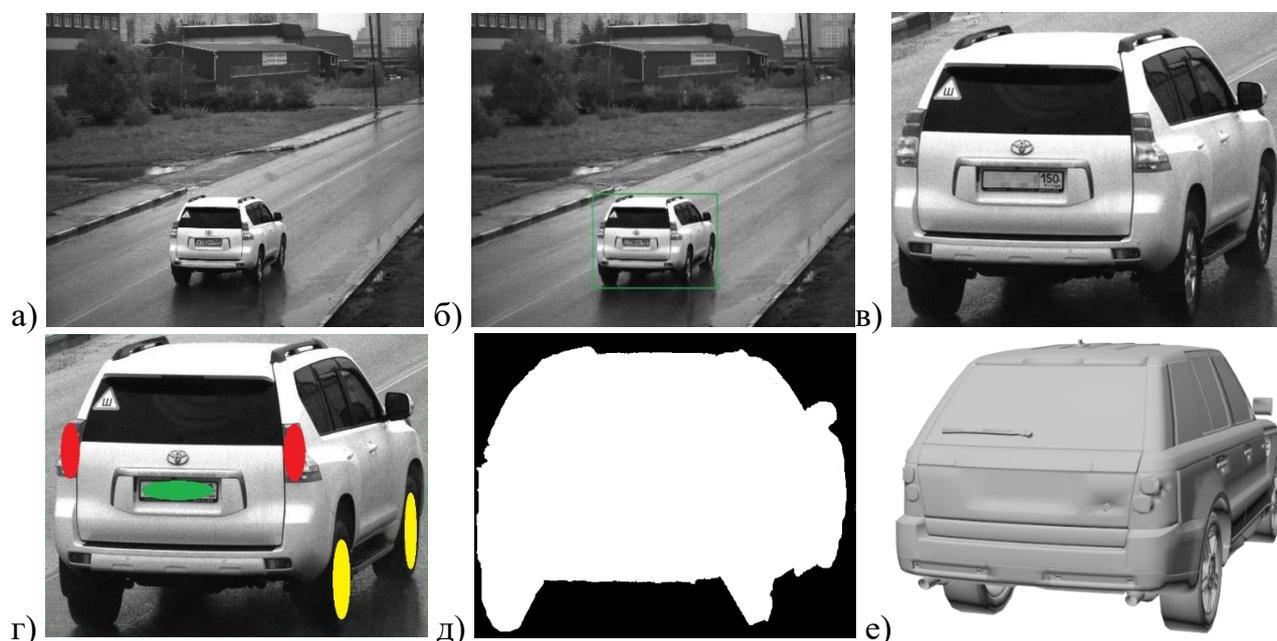


Рисунок 2 – пример работы каскадного подхода оценки ракурса ТС: а) исходное изображение; б) локализация ТС на изображении; в) часть изображения, содержащая объект интереса; г) локализация ключевых точек; д) сегментация изображения; е) совмещение маски объекта и 3D модели ТС.

Во второй главе определена реализация первого шага предлагаемого подхода, а именно локализация объекта интереса на изображении. ТС являются одним из основных объектов, присутствующих во многих классических наборах данных обнаружения и распознавания объектов на изображении, поэтому целесообразно не обучать нейросетевую модель с нуля, а использовать известные модели для обнаружения ТС, выбрав их с учетом нескольких критериев. С одной стороны, необходима высокая скорость обработки изображения, с другой стороны, необходима высокая точность, поскольку любое обнаруженное ТС будет подвержено дальнейшему анализу. Соответствующая точность локализации определяет точность последующих шагов.

Развертывание моделей, обученных на общедоступных наборах изображений в крупномасштабной сети реальных камер, может привести к снижению производительности из-за недостаточного многообразия целевого набора данных. Для

оценки обобщающей способности методов локализации ТС к внешним факторам было произведено кросс-валидационное тестирование современных методов локализации объектов, разработанных специально для практических задач, анализирующих видеопоток в реальном времени таких как: YOLOv4, MobileNetV2_SSD, RetinaNet_ResNet50, FCOS_ResNet-50, MobileNetV1_SSD. В качестве тестового множества использовался новый набор данных, содержащий более 49000 изображений, на которых присутствует более 61000 автомобилей. В результате для первого шага каскадного подхода оценки ракурса объекта интереса была выбрана одноступенчатая СНС YOLOv4 обеспечивающая метрику качества локализации ТС на изображении (average precision) 97%, превосходя аналоги более чем на 12%.

В третьей главе проведен анализ современных методов сегментации изображений, как классических, так и нейросетевых. Разработан новый метод сегментации изображений на основе априорной информации о местоположении объекта интереса. Выполнено экспериментальное исследование качества метода сегментации изображений и устойчивости разработанного метода к шумам различного вида.

Результатом локализации является информация о местоположении объектов на изображении (x, y, w, h, c, p) , где (x, y) – координаты центра ограничивающего объект прямоугольника, (w, h) – ширина и высота объекта на изображении соответственно, c – класс объекта (выбранный метод позволяет отличать легковые и грузовые автомобили, а также автобусы), p – вероятность предсказания класса объекта.

Основываясь на данных, полученных при локализации объектов, разработан метод автоматической сегментации изображений.

Основные шаги предложенного метода:

1) Преобразование изображения из декартовой системы координат в полярную. Пусть масштабированное изображение локализованного объекта имеет разрешение (W, H) . Каждый пиксел изображения преобразуется в полярную систему координат, используя центр изображения, как начало координат $O = \left(\frac{W}{2}, \frac{H}{2}\right)$:

$$r = \sqrt{\left(x' - \frac{W}{2}\right)^2 + \left(y' - \frac{H}{2}\right)^2},$$

$$\phi = \begin{cases} \arctg\left(\frac{2y'-H}{2x'-W}\right), & x' > \frac{W}{2}, y' \geq \frac{H}{2} \\ \arctg\left(\frac{2y'-H}{2x'-W}\right) + 2\pi, & x' > \frac{W}{2}, y' < \frac{H}{2} \\ \arctg\left(\frac{2y'-H}{2x'-W}\right) + \pi, & x' < \frac{W}{2}, \\ \frac{\pi}{2}, & x' = \frac{W}{2}, y' > \frac{H}{2} \\ \frac{3\pi}{2}, & x' = \frac{W}{2}, y' < \frac{H}{2} \end{cases}$$

где r – расстояние от пиксела до начала координат; ϕ – полярный угол; (x', y') – координаты пиксела изображения.

2) Результат полярного преобразования представляется в прямоугольной системе координат (рис. 3б) I_ϕ с разрешением $(w_\phi; h_\phi)$ и формируется контурный препарат изображения, производя операцию свертки изображения оператором Собеля с ядром размера 13x13:

$$L_{\text{конт.}}(\phi, r) = |I_\phi(\phi, r) * G_r|,$$

$$G_r = \begin{bmatrix} -1 & -12 & -66 & -220 & -495 & -792 & -924 & -792 & -495 & -220 & -66 & -12 & -1 \\ -10 & -120 & -660 & -2200 & -4950 & -7920 & -9240 & -7920 & -4950 & -2200 & -660 & -120 & -10 \\ -44 & -528 & -2904 & -9680 & -21780 & -34848 & -40656 & -34848 & -21780 & -9680 & -2904 & -528 & -44 \\ -111 & -1320 & -7260 & -24200 & -54450 & -87120 & -101640 & -87120 & -54450 & -24200 & -7260 & -1320 & -110 \\ -165 & -1980 & -10890 & -36300 & -81675 & -130680 & -152460 & -130680 & -81675 & -36300 & -10890 & -1980 & -165 \\ -132 & -1584 & -8712 & -29040 & -65340 & -104544 & -121968 & -104544 & -65340 & -29040 & -8712 & -1584 & -132 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 132 & 1584 & 8712 & 29040 & 65340 & 104544 & 121968 & 104544 & 65340 & 29040 & 8712 & 1584 & 132 \\ 165 & 1984 & 10890 & 36300 & 81675 & 130680 & 152460 & 130680 & 81675 & 36300 & 10890 & 1980 & 165 \\ 110 & 1320 & 7260 & 24200 & 54450 & 87120 & 101640 & 87120 & 54450 & 24200 & 7260 & 1320 & 110 \\ 44 & 528 & 2904 & 9680 & 21780 & 34848 & 40656 & 34848 & 21780 & 9680 & 2904 & 528 & 44 \\ 10 & 120 & 660 & 2200 & 4950 & 7920 & 9240 & 7920 & 4950 & 2200 & 660 & 120 & 10 \\ 1 & 12 & 66 & 220 & 495 & 792 & 924 & 792 & 495 & 220 & 66 & 12 & 1 \end{bmatrix}$$

3) Представление контурного препарата $L_{\text{конт.}}(\phi, r)$ в виде ориентированного взвешенного графа G . В качестве множества вершин графа использованы множество пикселей изображения. Ребрами соединены только соседние пиксели. При том два пиксела q и p , являются соседними, если удовлетворяют условию:

$$(\phi'_p - \phi'_q = 1) \cap (|r'_p - r'_q| \leq 1),$$

где $\phi'_q, r'_q; \phi'_p, r'_p$ – координаты пикселей q и p в представлении графа в прямоугольной системе координат, соответственно.

Веса ребер, разделяющих два пиксела, в точке с координатами q и p , представим как:

$$d(p, q) = \frac{2h_\phi - (r'_p + r'_q)}{L_{\text{конт.}}(\phi'_q, r'_q) + L_{\text{конт.}}(\phi'_p, r'_p) + 1}.$$

Для поиска замкнутого контура $r_n(\phi)$ использован алгоритм поиска кратчайшего пути. В качестве начальной и конечной вершин графа выбраны пиксели с координатами $(0, \frac{h_\phi}{2})$ и $(w_\phi, \frac{h_\phi}{2})$, соответственно.

На рисунке 3в представлен результирующий путь с наименьшим весом. Используя обратное преобразование выражений (1), (2), полученный путь преобразуется в декартову систему координат. В результате формируется замкнутый контур, описывающий объект интереса. Пиксели исходного изображения, расположенные внутри замкнутого контура, выделяются в качестве изображения объекта (рис. 3г).

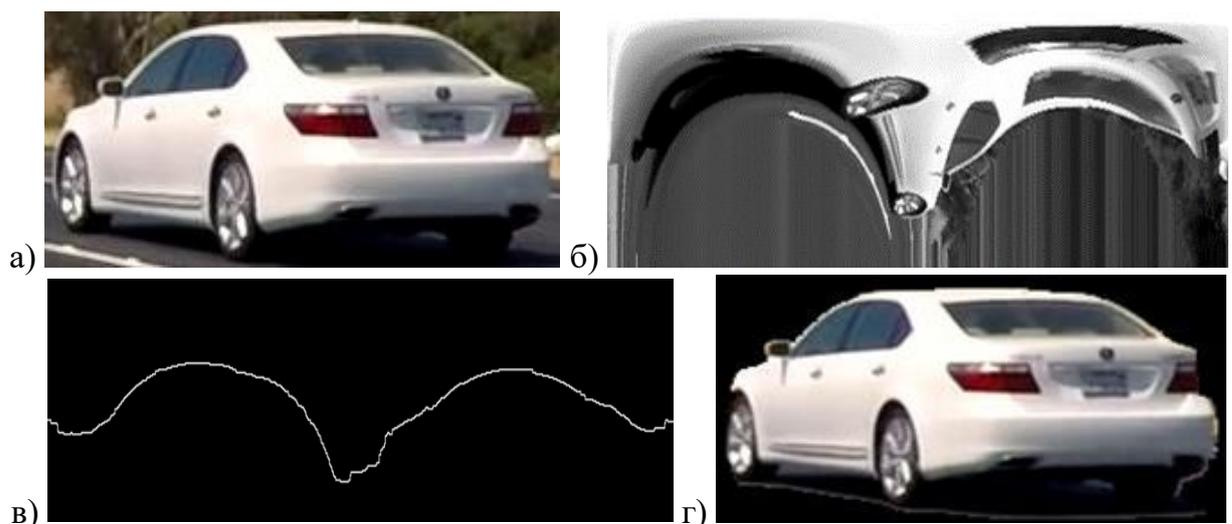


Рисунок 3 Пример работы метода сегментации изображения: а) исходное изображение; б) результат преобразования; в) результат выделения кратчайшего пути графа; г) результат работы метода.

Экспериментальное исследование показывает: разработанный метод сегментации изображений на основе априорной информации о местоположении объекта интереса, обеспечивает меру сходства Жаккара 89%, превосходя аналоги (k-means, GrabCut, DeepLabv3) более чем на 10%; в сложных условиях наблюдения, таких, как влияние шумов различного вида, разработанный метод сегментации изображения при ПОСШ, составляющего 20 Дб, обеспечивает меру сходства Жаккара 79%, 80% и 79% при влиянии импульсного, мультипликативного и аддитивного белого гауссовского шума соответственно; в отличие от современных нейросетевых методов сегментации изображений, разработанный метод не требует трудоемких вычислений и аннотирования данных.



Рисунок 4 Примеры сегментации изображений с добавлением шума.

На рисунке 4 представлены результаты сегментации изображений разработанным методом для различных значений ПОСШ

В четвертой главе произведен анализ современных методов локализации ключевых точек объекта на изображении. Разработан метод локализации ключевых точек объекта интереса на основе анализа внутренних слоев предварительно обученных СНС. Выполнено экспериментальное исследование точности разработанного метода.

Чтобы произвести анализ каждого слоя СНС, необходимо изменить её архитектуру. Каждый слой СНС может быть потенциально значимым и нести информацию об интересующей семантической части объекта. Для анализа внутренних слоев был добавлен выход к каждому сверточному слою, с последующим масштабированием выходной матрицы до разрешения входного изображения.

Модифицированная структура имеет дополнительный выход в виде матрицы всех промежуточных сверточных слоев базовой структуры с размерами $W \times H \times N$ где W и H – высота и ширина входного изображения соответственно; N – количество промежуточных сверток НС.

Для анализа СНС необходим набор аннотированных данных X – множество изображений (I) и масок (A_k), содержащих информацию о местоположении k -й интересующей части объекта.

Для определения ключевой точки выделяются сверточные слои и каналы, доля сосредоточения активации в области интереса, в которых удовлетворяет двум условиям:

– активирующиеся в интересующей области:

$$\hat{l}_{ijk} = \{Q_{ijk} \geq th_1\};$$

– не активирующиеся в интересующей области:

$$\tilde{V}_{ijk} = \{Q_{ijk} \leq th_2\},$$

$$Q_{ijk} = \sum_{\{I, A_k \in X\}} \frac{\sum_{\substack{y \in 1, H \\ x \in 1, W}} f_{ij}[I(x, y)] A_k(x, y)}{\sum_{\substack{y \in 1, H \\ x \in 1, W}} f_{ij}[I(x, y)]},$$

где $f_{ij}(I)$ – результат функции активации i -го слоя, j -го канала, масштабированный до размера входного изображения; пороги th_1 и th_2 определяются видом функции активации, используемой в анализируемой архитектуре.

Элементы, не входящие в множества \hat{l}_{ijk} и \tilde{V}_{ijk} , удаляются из дополнительного выходного слоя модифицированной архитектуры СНС.

Агрегируя данные с выхода базовой структуры СНС, получено новое изображение I' , элементы которого, локализованные в интересующей части объекта, обладают наибольшими значениями интенсивности (рис. 5):

$$F(x, y) = \sum_{\substack{i, j, k \in \hat{l}_{ijk} \\ y \in 1, H \\ x \in 1, W}} f_{ij}[I'(x, y)] - \sum_{\substack{i, j, k \in \tilde{V}_{ijk} \\ y \in 1, H \\ x \in 1, W}} f_{ij}[I'(x, y)].$$

НС YOLOv4 принимает на вход изображение и выдает координаты ограничивающего объект прямоугольника: левого верхнего угла x_{min} , y_{min} и правого нижнего угла x_{max} , y_{max} . Для получения координат ключевой точки объекта на изображении, производится поиск максимума в области ограничивающего объект прямоугольника:

$$x_{кл}, y_{кл} = \underset{\substack{x' \in x_{min}, x_{max} \\ y' \in y_{min}, y_{max}}}{argmax} F(x', y').$$

Экспериментальное исследование показывает: предложенный метод обеспечивает долю корректной локализации ключевой точки объекта интереса 97% для базовой структуры Darknet53 СНС YOLOv4, в среднем превосходя аналоги такие как MaskRCNN и WPODNet на 10%; разработанный метод не требует больших наборов аннотированных данных и обучения нейросетевой модели; предложенный метод обеспечивает высокую точность определения ключевых точек при различных условиях наблюдения; предложенный метод эффективен для различных базовых структур СНС, доля корректной локализации ключевой точки объекта интереса для базовых структур СНС VGG19 и SqueezNet составляет 97% и 95% соответственно.

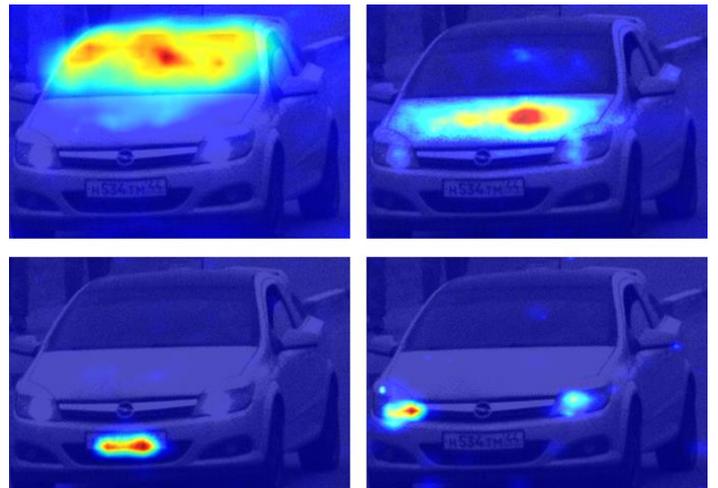


Рисунок 5 Пример визуализации агрегированных выходов базовой структуры НС YOLOv4.

В пятой главе произведен анализ существующих методов сопоставления форм объекта на изображении с шаблоном. Разработан новый метод определения ракурса ТС, основанный на сопоставлении дескриптора формы объекта интереса с трехмерной моделью. Проведено экспериментальное исследование предложенного метода.

Удаление нерелевантной информации и локализация ключевых точек объекта интереса позволяет перейти к анализу ракурса. Ракурс ТС определяется, как ракурс шаблона, с которым дескриптор формы имеет наибольшее значение коэффициента корреляции.

Внешний контур объекта описан с помощью сигнатуры (одномерной функции). Сигнатура определена как зависимость расстояния от ключевой точки объекта до его границы $r(\phi)$, нормированного на максимальное значение:

$$r_n(\phi) = \frac{r(\phi)}{\max r(\phi)}.$$

Полученные таким образом сигнатуры, инвариантны по отношению к параллельному переносу и изменению масштаба, однако, зависят от поворота. Чтобы получить инвариантность относительно поворота полученную сигнатуру представим в виде Фурье-спектра. Вычислим ДПФ:

$$a = \sum_{\phi=0}^{N-1} |r_n(\phi) e^{-i2\pi u\phi/N}|.$$

Ракурс целевого ТС определяется, как ракурс шаблона, с которым ракурс целевого ТС имеет наибольшее значение коэффициента корреляции:

$$\alpha = \frac{\sum_{u=1}^N (a_u - \bar{a})(t_u - \bar{t})}{\sqrt{\sum_{u=1}^N (a_u - \bar{a})^2 \sum_{u=1}^N (t_u - \bar{t})^2}},$$

где, \bar{a} и \bar{t} – математические ожидания Фурье дескрипторов формы целевого ТС и шаблона соответственно.

В качестве набора эталонных данных были использованы изображения ТС разных ракурсов, полученных на основе трехмерных моделей. Трехмерные модели выбраны таким образом, чтобы представить внутриклассовые вариации (седан, купе, внедорожник, хэтчбек и т. д.). Высота точки наблюдения и количество ракурсов ТС выбираются в зависимости от условий эксплуатации.

Зависимость доли корректного определения ракурса ТС разработанным методом от шага дискретизации угла наблюдения изображена на рисунке 6.

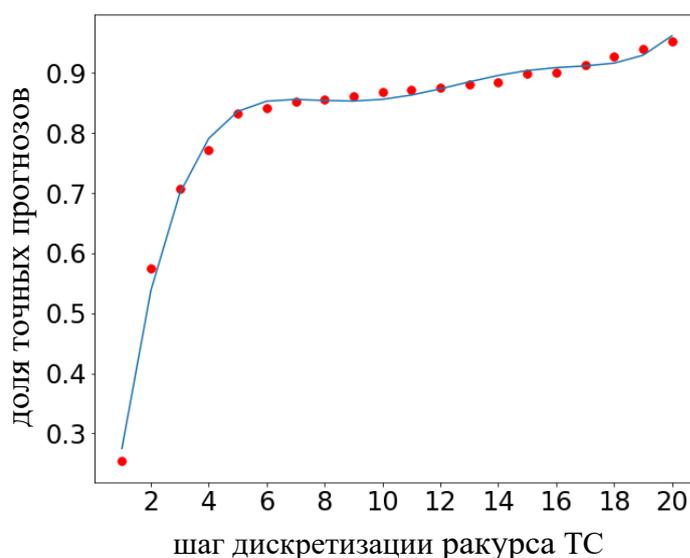


Рисунок 6 – зависимость доли точных прогнозов от шага дискретизации.

На рисунке 7 представлены примеры изображений ТС различного ракурса.

Зависимость доли корректного определения ракурса автомобиля от окклюзии изображена на рисунке 8, здесь 0% означает, что объект виден полностью (рис. 9а), 50% объект виден на половину (рис. 9б).

Экспериментальные исследования показывают: величина средней ошибки определения ракурса предложенным методом для 16 ракурсов, составляет 2.3%, притом 1,8% ложных предсказаний составляет разница в 180 градусов, а оставшиеся 0.5% - соседние ракурсы; использование ключевых точек, принадлежащих номерному знаку и колесам, позволяет уменьшить ошибку определения ракурса предложенным методом на 2%: разработанный метод обеспечивает долю корректного определения ракурса ТС 80% с дискретом в 6 градусов; доля корректного определения ракурса ТС составляет 80% для 16 ракурсов при окклюзии объекта, составляющей 60%; определение ракурса ТС на основе дескриптора Фурье сигнатуры объекта позволяет использовать в качестве эталонных данных трехмерные модели, что исключает необходимость ручной аннотации данных; каскадный подход определения ракурса ТС на основе СНС YOLOv4 и разработанных методов локализации ключевых точек, сегментации и определения ракурса на основе формы объекта обеспечивает долю корректного определения ракурса 89% превосходя аналоги, такие как HSSR и AAVR.



Рисунок 7 – изменения проекции ТС с шагом дискретизации в 22 градуса.

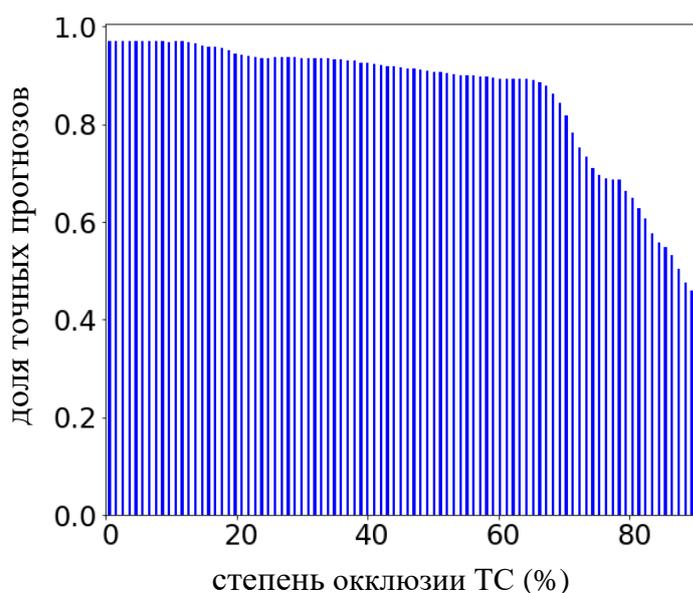


Рисунок 8 – зависимость доли точных прогнозов от окклюзии.



Рисунок 9 – а) ТС без окклюзии; б) окклюзия ТС 50%.

В заключении сформулированы основные результаты, полученные в диссертационной работе:

1. Разработан новый метод сегментации изображения на основе преобразования в полярную систему координат и поиска кратчайшего пути во взвешенном графе, позволяющий в автоматическом режиме выделить объект на сложном фоне при условии предварительной локализации объекта. Величина средней ошибки сегментации разработанного метода, для ПОСШ, составляющего 20 Дб, не превышает 22%. В отличие от современных нейросетевых методов сегментации изображений, разработанный метод не требует трудоемких вычислений и аннотирования данных.

2. Разработан новый метод локализации ключевых точек ТС, основанный на анализе внутренних слоев СНС, позволяющий повысить информативность метода локализации объектов на изображении и обеспечивающий устойчивый анализ для различных архитектур и условий наблюдения. Разработанный метод не требует больших наборов аннотированных данных и обучения нейросетевой модели, а также не зависит от архитектуры СНС.

3. Метод определения ракурса ТС на основе дескриптора формы и локализации ключевых точек объекта интереса, обеспечивающий долю точных прогнозов определения ракурса 97% для дискрета в 16 градусов. Доля точных прогнозов 80% для дискрета 16 ракурсов достигается при окклюзии объекта, составляющей 60%. Определение ракурса ТС на основе формы объекта позволяет использовать в качестве эталонных данных трехмерные модели, что исключает необходимость ручной аннотации данных.

4. Разработан новый каскадный подход анализа изображения, включающий в себя локализацию и сегментацию объекта интереса, локализацию ключевых точек, а также оценку ракурса обеспечивает: долю корректного определения ракурса на 10% выше имеющихся аналогов, включая методы, построенные на сквозном обучении; возможность к развитию в соответствии с объективными изменениями условий функционирования.

Общий вывод. Полученные результаты диссертационной работы повышают точность анализа окружающего пространства телевизионной системы автономного вождения, обеспечивая необходимый анализ видеопотока для оценки пространственной ориентации ТС. Результаты работы использованы в научно-исследовательских работах ООО «НЕКСТ», ООО «Инновационный центр «КАМАЗ», а также в учебном процессе кафедры телевидения и видеотехники Санкт-Петербургского государственного электротехнического университета, о чём имеются соответствующие акты.

Таким образом, совокупность полученных выводов и научных положений показывает, что поставленные задачи разработки методов обработки и анализа видеоданных в системе управления беспилотного автомобиля, выполнены.

СПИСОК ОПУБЛИКОВАННЫХ РАБОТ ПО ТЕМЕ ДИССЕРТАЦИИ

В журналах из списка, рекомендованного ВАК России:

1. Зубов, И.Г. Метод автоматической сегментации транспортных средств на изображении / И.Г. Зубов // Известия высших учебных заведений России. Радиоэлектроника. - 2019. - № 22(5). – С. 6–16.

2. Зубов, И.Г. Метод автоматического определения ключевых точек объекта на изображении / И.Г. Зубов // Известия высших учебных заведений России. Радиоэлектроника. - 2020. - № 23(6). – С. 6–16.

3. Зубов, И.Г. Метод автоматического определения трехмерной траектории транспортных средств на изображении / И.Г. Зубов, Н.А. Обухова // Известия высших учебных заведений России. Радиоэлектроника. - 2021. - № 24(3). – С. 49–59.

4. Зубов, И.Г. Обработка и анализ видеоданных в системе управления беспилотного автомобиля / И.Г. Зубов // Вопросы радиоэлектроники. Сер. Техника телевидения. – 2021. – Вып. 3. – С. 96–99.

Другие статьи и материалы конференций:

5. Зубов, И.Г. Алгоритм сегментации транспортных средств на ТВ изображениях / И.Г. Зубов // 15-я межд. конф. «Телевидение: передача и обработка изображений»: Материалы конф., г. С.-Петербург, 26–27 июня 2018 г. – СПб: изд-во ООО «Технолит». – 2018. – С. 135–137.

6. Зубов, И.Г. Оценка ракурса транспортного средства, полученного монокулярной камерой / И.Г. Зубов // Цифровая обработка сигналов и её применение – DSPA-2019: Материалы конф., г. Москва, 27–29 марта 2019 г. – С. 381–386.

7. Zubov, I.G. Estimation of Vehicle Pose with Monocular Camera / I.G. Zubov // 2019 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 28-31 Jan 2019. – P. 395–397.

8. Zubov, I.G. Vehicle Pose Estimation Based on Object Contour / I.G. Zubov // 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 27-30 Jan 2020. – P. 1452–1454.

9. Зубов, И.Г. Автоматический метод определения ракурса автомобиля на изображении / И.Г. Зубов // Цифровая обработка сигналов и её применение – DSPA–2020: Материалы конф., г. Москва, 14–15 апреля 2020 г. – С. 430–435.