

Чигорин Александр Александрович

**Алгоритмы и программная система для выделения и
распознавания объектов в видеопоследовательности**

Специальность 05.13.11 – математическое и программное обеспечение
вычислительных машин, комплексов и компьютерных сетей

АВТОРЕФЕРАТ

диссертации на соискание учёной степени
кандидата физико-математических наук

Москва – 2014

Работа выполнена на кафедре автоматизации систем вычислительных комплексов факультета вычислительной математики и кибернетики Московского государственного университета имени М. В. Ломоносова.

Научный руководитель: кандидат физико-математических наук, доцент
Конушин Антон Сергеевич, заведующий
Лабораторией компьютерной графики и
мультимедиа МГУ имени М.В. Ломоносова

Официальные оппоненты:

Ведущая организация:

Защита состоится “___” месяца 2015 г. в ____ часов на заседании диссертационного совета Д002.024.01 созданного на базе ФГБУН Институт прикладной математики имени М.В. Келдыша РАН, расположенного по адресу: 125047 Москва, Миусская пл., д.4.

С диссертацией можно ознакомиться в библиотеке и на сайте ФГБУН Институт прикладной математики им. М.В. Келдыша РАН: www.keldysh.ru

Автореферат разослан “___” месяц 2015 г.

Учёный секретарь
диссертационного совета
доктор физико-математических наук

Т.А. Полилова

Общая характеристика работы

Объект исследования и актуальность работы

Объем информации, получаемой с помощью видеокамер растет экспоненциально от года к году. Например, количество часов видео, загружаемых каждую минуту на сайт youtube.com, увеличилось с 5 часов в 2007 году до 100 часов в 2014. На сегодняшний день подавляющее большинство видеоконтента не подвергается автоматическому анализу. В то же время извлечение информации из видео может быть полезно во многих практических приложениях, таких как мобильное картографирование, видеонаблюдение, поиск по видеоконтенту, робототехника или системы помощи водителю.

Ключевой задачей во всех перечисленных приложениях является задача выделения и распознавания объектов в видеопоследовательности. Для ее решения необходимо:

- выделить объекты на отдельных кадрах видеопоследовательности
- классифицировать объекты
- сопоставить объекты между кадрами.

Для апробации предлагаемых алгоритмов в данной работе рассматривается задача мобильного картографирования - автоматического нанесения объектов на карту. Входные данные представляют собой геопривязанный видеоряд с камеры, установленной на движущейся платформе, и параметры камеры. В результате работы алгоритма должны быть определены тип и положение искомым объектов на карте. В качестве таких объектов могут выступать любые объекты придорожной инфраструктуры: знаки дорожного движения, дорожная разметка, столбы, остановки, здания и так далее. Примеры основных этапов решения задачи приведены на рисунке 1.

В качестве области практического приложения предлагаемых алгоритмов рассматривается задача автоматического картографирования дорожных знаков. Знак дорожного движения является одним из ключевых объектов придорожной инфраструктуры. Знание о расположении и классе знаков может быть полезно при построении и обновлении навигационных карт, в системах помощи водителю, при решении задачи учета и управления дорожной инфраструктурой. В первом случае

знание о положении знаков дорожного движения, их ориентации и классе позволит автоматизировать построение дорожного графа. Уменьшение доли человеческого участия в данном случае может существенно уменьшить интервал между получением данных и полным завершением их обработки. Во втором случае информация о знаках позволит уведомлять водителя о текущей ситуации на дороге (рекомендуемую скорость, запрещенные повороты, запрет обгона и так далее).



Рисунок 1. Визуализация работы алгоритма автоматического нанесения объектов на карту. На первом этапе происходит выделение и распознавание объектов на нескольких кадрах видеопоследовательности. На втором выделенные объекты сопоставляются между кадрами и производится их локализация на карте.

Дорожные знаки сделаны, чтобы быть заметными, и имеют отличительные цвет и форму. Но разнообразие типов знаков и трансформаций над ними, встречаемых при решении задачи в промышленных масштабах, оставляют задачу высокоточного обнаружения знаков нерешенной до сих пор.

В данной работе рассматривается общая схема современных алгоритмов выделения и распознавания объектов и формулируется проблема их зависимости от наличия большой и репрезентативной обучающей выборки. Эта проблема особенно актуальна в тех случаях, когда создание обучающей выборки требует много сил и материальных затрат. Например, в случае знаков дорожного движения создание обучающей выборки может потребовать разметки десятков километров проезда. Более того, разметку приходится повторять для каждой новой страны, так как вид знаков сильно различается в разных странах.

В результате проведенного анализа в первой главе диссертации предлагается алгоритм автоматического создания обучающей выборки и ряд методов, позволяющих

оценить качество получаемых искусственных данных. Далее, для каждого из этапов решения исходной задачи, предлагаются алгоритмы, созданные для обучения на искусственных данными. Во второй главе рассматривается задача выделения объектов на отдельных кадрах видеопоследовательности и предлагается новый алгоритм, позволяющий использовать в качестве обучающей выборки синтетические данные. Третья глава посвящена задаче классификации. В ней также рассмотрена проблема неточной локализации объектов за счет множественных срабатываний алгоритма обнаружения вокруг области интереса. В результате анализа предложен новый алгоритм, позволяющий уточнить положение объекта, отделить его от фона и распознать его класс. Оба предложенных алгоритма способны обучаться на искусственно созданных данных. В четвертой главе более подробно описана задача мобильного картографирования, используемая для апробации предложенных алгоритмов на практике. Предложена схема автоматизированной системы для решения задачи. Дано описание каждого из модулей системы, состоящее из входных и выходных данных, решаемой модулем задачи и описания способа ее решения.

Цель диссертационной работы

Целью данной диссертационной работы является разработка алгоритмов и программной системы для выделения и распознавания объектов в видеопоследовательности, способных к обучению на искусственно созданных данных. Программная система не должна требовать пользовательского ввода и должна позволять обрабатывать большие объемы данных дешевле и быстрее ручной разметки операторами.

Научная новизна работы

В данной работе разработан алгоритм создания искусственной обучающей выборки, позволяющий сократить затраты на разметку обучающих данных. Предложены методы оценки качества получаемых синтетических данных в сравнении с их реальными аналогами. На примере задачи классификации знаков дорожного движения показано, что за счет обучения на большом объеме искусственно созданных данных можно улучшить точность классификации по сравнению с обучением на реальных данных. Предложены модификации классического алгоритма обнаружения объектов Виолы-Джонса,

позволяющие увеличить скорость и точность его работы.

Также, в рамках работы над задачей классификации объектов, предложен новый алгоритм сегментация объектов от фона. Алгоритм создан для обучения на искусственных данных и позволяет оценивать параметры преобразований, которым был подвержен объект интереса. Предложена схема обучения на искусственных данных с использованием глубокой сверточной нейронной сети, позволяющая обучаться на выборках большого объема.

Практическая значимость и реализация

В рамках работы создана программная реализация разработанных алгоритмов, удовлетворяющая всем требованиям и ограничениям, сформулированным в цели диссертации. Для экспериментальной оценки предложенных алгоритмов была создана база знаков дорожного движения Российской Федерации (Russian Traffic Signs Dataset, RTSD)¹. Данные для разметки были предоставлены ЗАО “Геоцентр-Консалтинг.” Чтобы позволить дальнейшее сравнение с другими методами данные сделаны общедоступными.

Предложенные алгоритмы разрабатывались в рамках проекта по автоматическому картографированию знаков дорожного движения в компании ООО “GeoCV”.

Тестирование предложенных алгоритмов выделения объектов проводилось на вышеупомянутой базе данных RTSD. Алгоритмы распознавания также тестировались на трех общедоступных базах - German Traffic Signs Recognition Benchmark², KUL Belgium Traffic Sign Classification Benchmark³, Sweden Traffic Signs Dataset⁴.

На всех вышеперечисленных наборах были получены согласованные результаты, что подтверждает широкую применимость предложенных методов.

Апробация работы

Основные результаты работы докладывались и обсуждались на:

- 15-й международной конференции по компьютерному зрению ACIVS (Advanced Concepts for Intelligent Vision Systems), Польша, 2013

¹ Данные доступны по адресу <ftp://anonymous@kiviuq.graphicon.ru/AnonymousFTP/RTSD/>

² Данные доступны по адресу <http://benchmark.ini.rub.de/?section=gtsrb&subsection=about>

³ Данные доступны по адресу <http://www.vision.ee.ethz.ch/~timofter/>

⁴ Данные доступны по адресу <http://www.cvl.isy.liu.se/research/traffic-signs-dataset>

- международной конференции ISPRS (International Society for Photogrammetry and Remote Sensing), CMRT (City Models Roads And Traffics) workshop, Турция, 2013
- 3-й научно-технической конференции “Техническое зрение в системах управления”, Россия, 2012
- 15-й международной конференции DSPA (Digital Signal Processing and Applications), Россия, 2013
- 14-й международной конференции DSPA (Digital Signal Processing and Applications), Россия, 2012
- 22-й международной конференции по компьютерной графике и машинному зрению Graphicon, Россия, 2012
- научной конференции «Ломоносовские чтения», Россия, 2013.
- 6-й летней школе Microsoft для аспирантов (Microsoft Research PhD Summer School), Англия, Кембридж, 2011
- семинаре аспирантов кафедры АСВК факультета ВМК МГУ под руководством Л. Н. Королева, 2013
- семинаре группы компьютерного зрения лаборатории компьютерной графики и мультимедиа под руководством А. С. Конушина, ВМК МГУ, 2011

Публикации

По теме диссертации автором опубликовано 9 научных работ, из них 3 статьи в рецензируемых журналах, включенных в перечень ВАК [4,5,6]. Статья [4] была опубликована в журнале Lecture Notes in Computer Science издательства Springer. Доклад по статье [7] победил в номинации за лучшую презентацию во время проведения семинара CMRT 2013 на конференции ISPRS2013-SSG.

Содержание работы

Во введении описывается задача выделения и распознавания объектов в видеопоследовательности и обосновывается ее актуальность. Отдельно рассмотрена задача мобильного картографирования, как одна из наиболее важных практических задач анализа видео. Обосновывается актуальность и выделяются четыре основных этапа ее решения:

1. выделение объектов интереса
2. классификация объектов
3. сопоставление изображений объекта на соседних кадрах видеопоследовательности
4. локализация выделенных объектов на карте.

Рассмотрены ограничения на входные данные. Предполагается, что на вход алгоритму поступает видеопоследовательность в цветовом формате RGB с разрешением кадров не менее 800x600 пикселей. Частота кадров видеопоследовательности должна быть не меньше 5 кадров в секунду при скорости мобильной платформы в 50 км/час. Если максимальная скорость мобильной платформы может быть выше 50 км/час, то частота кадров должна быть увеличена пропорционально изменению скорости. При этом предполагается, что частота кадров в обрабатываемом фрагменте видеопоследовательности постоянна. Главная ось камеры должна быть сонаправлена с направлением движения мобильной платформы. Минимальный размер объектов интереса на отдельных кадрах видеопоследовательности должен быть равен 24 пикселям по каждой стороне ограничивающего прямоугольника. Для успешного обнаружения объект интереса должен быть виден минимум на двух последовательных кадрах видеопоследовательности.

Формулируется проблема сбора репрезентативной обучающей выборки, необходимой для обучения алгоритмов выделения и классификации объектов. Во многих случаях создание подобной выборки требует значительных финансовых затрат на разметку. В качестве примера приводится задача разметки данных для распознавания знаков дорожного движения, требующая просмотра десятков километров проезда и выделения на них всех встречающихся знаков с отметкой класса знака. При этом разметку необходимо повторять для каждой новой страны, так как внешний вид знаков разных стран сильно различается.

Первая глава посвящена задаче создания искусственной обучающей выборки (рисунок 2), которая сформулирована следующим образом: пусть дан набор пар (p_i, r_i) и набор возможных трансформаций над пиктограммой $T(\theta) = \{t_1(\theta_1), t_2(\theta_2), \dots\}$, где p_i – пиктограмма, соответствующая объекту под номером i ,

r_i – реальное изображение i -го объекта,

θ_i - параметры трансформации $\theta = \{\theta_1, \theta_2, \dots\}$.

Целью алгоритма создания искусственных данных на этапе обучения является поиск плотности распределения параметров трансформаций $p(\theta)$, позволяющей создавать искусственную выборку, максимально похожу на реальную по некоторому критерию C оценки качества синтетических данных.



Рисунок 2. Слева пиктограмма объекта, к которой применяется набор трансформаций, переводящих ее в набор изображений (справа), максимально приближенных по своим характеристикам к реальным данным.

Далее предложено несколько критериев C для оценки качества получаемых искусственных данных:

1. через оценку точности получаемых в результате обучения на искусственных данных классификаторов (оценка точности классификаторов производилась на реальных данных)
2. с помощью оценки точности приближения реальных данных
3. с помощью визуальной оценки качества синтетических данных.

Предложен набор возможных трансформаций, которым может быть подвергнута пиктограмма во время преобразования в реальное изображение, состоящий из следующих трансформаций:

1. изменение яркости V и насыщенности S
2. вращение пиктограммы и маски знака относительно трех осей координат на углы R_x, R_y, R_z

3. гауссово размытие со среднеквадратичным отклонением σ_B и добавлением нормального шума со среднеквадратичным отклонением σ_N
4. добавление отступов (dx_l, dx_r, dx_u, dx_d) с четырех сторон объекта
5. масштабирование до нескольких размеров s_1, s_2, \dots, s_N и обратно до целевого размера классификатора
6. добавление фона из реальных изображений.

После этого произведена оценка важности вышеописанных трансформаций с помощью критерия, основанного на качестве получаемых классификаторов. Оценка производилась за счет использования классификатора на основе поиска ближайшего соседа и дескриптора на основе гистограмм ориентированных градиентов⁵. В процессе экспериментов изменению подвергались как параметры дескриптора, так и параметры трансформаций. В результате проведенных измерений сделаны следующие выводы о важности отдельных параметров трансформаций по критерию увеличения точности классификации:

1. отсутствие смещений и поворотов приводит к наиболее низкой точности классификации в 85.28%
2. добавление дополнительных смещений увеличивает точность классификации до 90.53%
3. увеличение плотности семплирования значения параметра смещения увеличивает точности до 91.08%
4. добавление поворотов дополнительно увеличивает точность до 93.15%
5. вариации параметров σ_B, σ_N, V и S не оказывают существенного влияния на точность классификации.

Таким образом, по критерию качества получаемых классификаторов наиболее важными параметрами трансформаций являются повороты и смещения объекта.

Вторым предложенным критерием оценки качества получаемых искусственных данных является метод на основе оценки точности приближения реальных данных. Для каждого реального изображения точность его приближения можно оценить за счет

⁵ Dalal N., Triggs W. Histogram of oriented gradients for human detection // Proc. IEEE Conf. Comput. Vis. and Pattern Recog. 2005. P. 886-893

поиска ближайшего соседа в наборе искусственных изображений. В данной работе были проведены эксперименты, использующие в качестве метрики похожести SAD (sum of absolute differences, сумма абсолютных разностей), которую можно записать следующим образом:

$$D = \sum_{p \text{ in pixels}} |I_1^{(p)} - I_2^{(p)}|, \text{ где}$$

$I_i^{(p)}$ - p -й пиксель первого i -го изображения.

Для многоканальных изображений вычисления производятся аналогичным образом за счет обработки каждого канала по отдельности и последующего суммирования. В качестве оценки качества приближения синтетической выборкой набора реальных изображений предложено использовать усредненное расстояние по всем реальным изображениям до ближайших искусственных примеров:

$$Q = \frac{\sum_{i=1}^N \min_{j=1, M} (SAD(I_r^{(i)}, I_s^{(j)}))}{N}, \text{ где}$$

N, M - количество реальных и искусственных изображений, соответственно,

$I_r^{(i)}$ - i -е реальное изображение,

$I_s^{(j)}$ - j -е искусственное изображение.

Сначала (по-анalogии с методом на основе точности получаемых классификаторов) была произведена оценка важности трансформаций на основе точности приближения по метрике SAD. Показано, что оба предложенных метода дают похожие результаты.

Затем была показана положительная корреляция (коэффициент корреляции Пирсона равен 0.98) между точностью аппроксимации реальных данных и точностью классификации, что является свидетельством в пользу выбранной метрики аппроксимации. После этого был проведен эксперимент по оценке плотности распределения параметров трансформаций $p(\theta)$. Для этого для каждого изображения из набора реальных изображений производился поиск ближайшего соседа в выборке искусственных изображений, созданных с помощью семплирования параметров трансформаций из равномерного распределения. Таким образом для каждого реального изображения была произведена оценка параметров трансформаций θ_i , которым оно

было подтверждено. После этого были построены гистограммы распределения значений параметров трансформаций (таблица 1), являющиеся аппроксимацией плотности распределения $p(\theta)$.

Чтобы показать, что параметры трансформаций были оценены корректно был проведен дополнительный эксперимент, в ходе которого классы были разделены на два непересекающихся набора A и B. По набору A были оценены плотности распределения $p(\theta)$. После чего было создано два варианта искусственной обучающей выборки для набора B. В первом варианте значения параметров трансформаций семплировались из равномерного распределения, а во втором из оцененного распределения $p(\theta)$. После этого была измерена точность классификации при использовании в качестве обучающих данных двух вариантов искусственной выборки. Результаты приведены в таблице 2. Видно, что оценка плотности $p(\theta)$ по реальным данным позволяет значительно увеличить точность классификации.

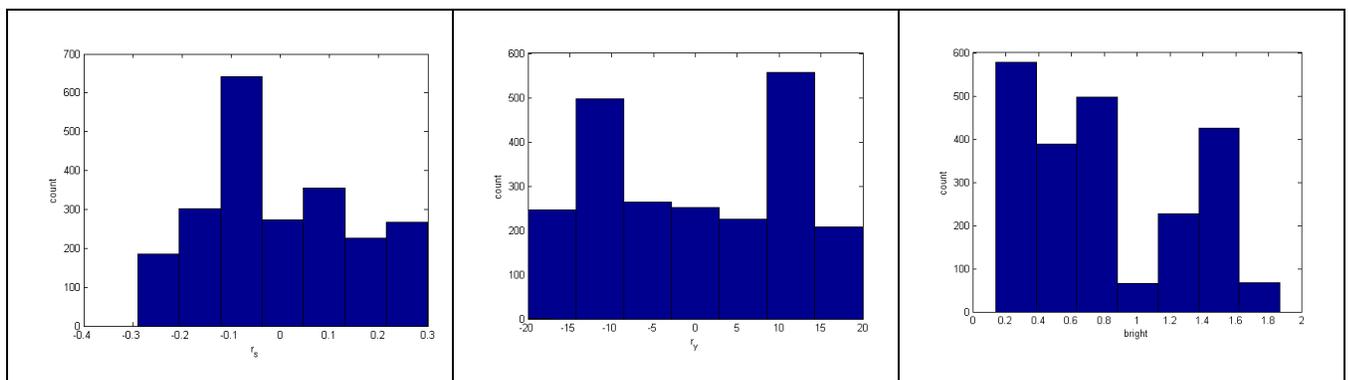


Таблица 1. Примеры гистограмм значений некоторых параметров трансформаций (слева направо: сдвиг справа, вращение вдоль оси y, коэффициент яркости).

Распределение на параметры трансформаций	Точность аппроксимации (SAD)	Точность классификации (1-NN) [%]
равномерное	32.99	70.2
оцененное по набору A	30.88	79.2

Таблица 2. Зависимость точности классификации от плотности распределения параметров трансформаций, использованной при создании искусственных изображений.

В начале **второй главы** формулируется задача выделения объектов и критерий оценки качества выделения. Пусть на входе дано изображение I со списком присутствующих на нем объектов (o_1, \dots, o_n) , заданных с помощью ограничивающих прямоугольников $o_i = (x, y, h, w)$. Считают, что прямоугольник r_i верно выделяет объект o_j , если отношение их площади пересечения к площади объединения больше некоторого заданного порога τ . В задаче выделения объектов, используя только входное изображение, необходимо найти минимально возможный набор прямоугольников (r_1, \dots, r_n) , выделяющих как можно большее количество искоемых объектов.

Для оценки предлагаемых алгоритмов была создана и выложена в общий доступ база российских знаков дорожного движения (RTSD), состоящую из 9508 изображений со знаками и 71050 изображений с фоном.

Для ускорения скорости работы модуля обнаружения при большом количестве классов объектов предложено разбивать множество объектов на типы и обучать отдельный детектор для каждого типа объектов. Такой подход позволяет существенно ускорить время работы детектора, особенно в условиях большого количества похожих классов объектов. В экспериментах со знаками дорожного движения объекты были разбиты на четыре типа (по цвету и форме) - *красные треугольники, красные круги, синие круги, синие квадраты*.

В качестве базового алгоритма обнаружения объектов был выбран метод, предложенный в работе Виолы и Джонса⁶. Предложен ряд модификаций исходного алгоритма, призванных улучшить скорость и точность его работы. Первая модификация состоит в использовании на ранних этапах каскада признаков на основе разделенных диполей (dissociated dipoles)⁷, извлекаемых из специально подготовленных цветовых каналов изображения, подчеркивающих различные цвета. Для примера приведена формула извлечения областей доминирования синего цвета⁸:

$$f_{blue}(x) = \max(0, \min(\frac{x_b - x_r}{x_r + x_g + x_b}, \frac{x_b - x_g}{x_r + x_g + x_b})) ,$$

⁶ Viola P., Jones M. Robust Real-Time Face Detection // International Journal of Computer Vision. VOL 57(2), P. 137-154

⁷ Balas B., Sinha P. Dissociated Dipoles: Image Representation via Non-local Comparisons // CBCL Paper #229, 2003

⁸ Ruta, A., Porikli, F., Watanabe, S., Li, Y., 2011. In-vehicle camera traffic sign detection and recognition. Mach. Vis. Appl., VOL 22(2), P. 359–375

где x_r, x_g, x_b - значения пикселя x в трех каналах RGB-изображения.

Экспериментально подтверждается превосходство цветных признаков перед признакам, извлекаемыми из изображений в градациях серого (рисунок 3, рисунок 4) в задаче обнаружения знаков дорожного движения.

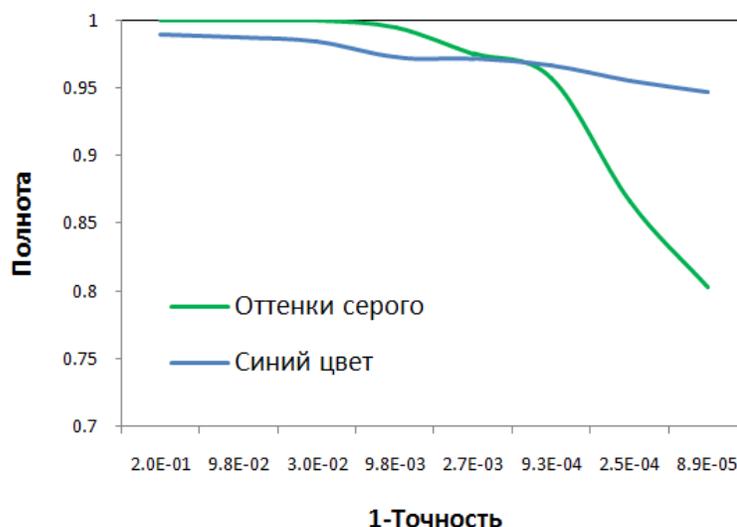


Рисунок 3. Соотношение полноты и точности для случаев использования цветных признаков и признаков на основе оттенков серого.

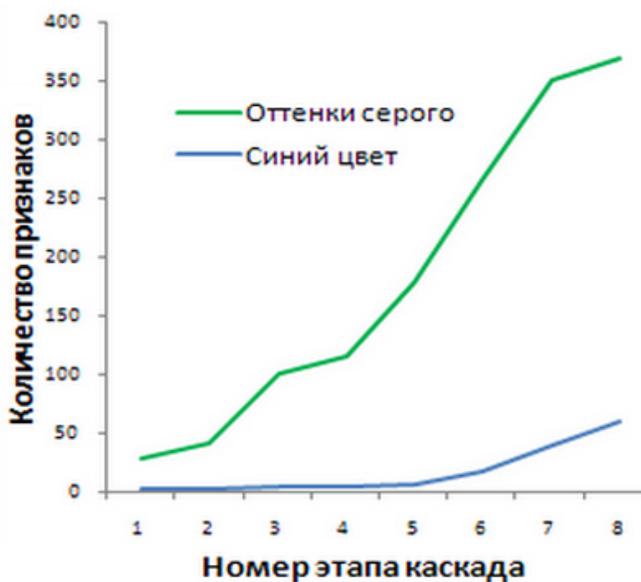


Рисунок 4. Зависимость количества необходимых признаков от этапа каскада для случаев использования цветных признаков и признаков на основе оттенков серого.

Вторая модификация заключается в использовании различных признаков и классификаторов на различных этапах каскада. Такой подход позволяет не терять в полноте обнаружения на поздних этапах каскада. Проводится сравнение AdaBoost-классификатора на основе диполей с SVM-классификатор на основе гистограмм ориентированных градиентов (HOG) и сверточной нейронной сетью (CNN), способной к обучению на больших объемах искусственных данных за счет использования современных графических сопроцессоров. Экспериментально доказывается, что использование сверточной нейронной сети вместо AdaBoost+Dipoles в качестве классификатора на последнем этапе каскада позволяет увеличить полноту обнаружения в среднем на 7% при том же уровне ложных срабатываний (рисунок 5).

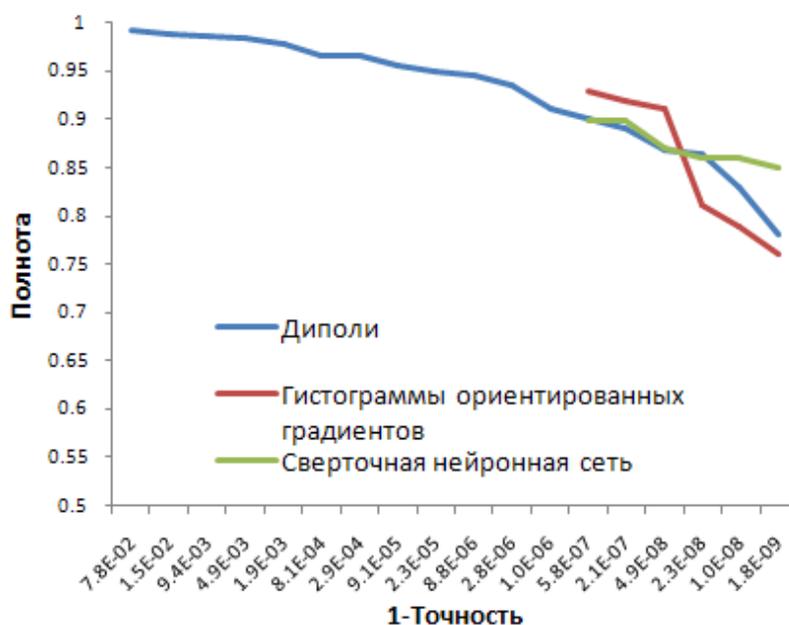


Рисунок 5. Соотношение полноты и точности при обучении каскада классификаторов типа знаков *синие квадраты* на основе различных признаков.

Производится экспериментальное подтверждение возможности обучения модуля обнаружения с учетом вышеописанных модификаций на искусственно созданной выборке. Для этого на базе RTSD измеряется точность работы предложенного алгоритма на четырех основных типах знаков дорожного движения (таблица 3).

Тип знака	Доля ложных срабатываний	Полнота (по изображениям)	Полнота (по физическим знакам)
Синие квадраты	$7 \cdot 10^{-10}$	77%	92.18%
Красные треугольники	$7 \cdot 10^{-10}$	73.1%	82.35%
Синие круги	$6 \cdot 10^{-10}$	79.2%	83%
Красные круги	$2 \cdot 10^{-9}$	73.8%	84.7%

Таблица 3. Точность работы каскада детектора на четырех типах знаков.

Третья глава посвящена задаче классификации выделенных объектов. В начале главы рассмотрена проблема множественных обнаружений детектора, возникающая в практических системах (рисунок 6).



Рисунок 6. Ограничивающие прямоугольники объекта, получаемые на выходе модуля обнаружения.

Для ее решения предложен алгоритм, обучаемый на искусственно созданных данных, и позволяющий не только находить точный ограничивающий прямоугольник объекта интереса, но и отделять объект интереса от фона. Предложенный метод основан на поиске ближайшего соседа в большой выборке искусственно созданных изображений объектов, для которых известны параметры преобразования и маска фона. Имея параметры преобразования нескольких ближайших соседей с помощью усреднения можно вывести параметры преобразования, которому был подвержен объект

на входном реальном изображении. Усреднение найденных масок фона также позволяет отделить объект интереса от фона (рисунок 7).



Рисунок 7. Примеры найденных ближайших соседей по метрике евклидова расстояния на HOG-дескрипторах. Левый столбец - входное реальное изображение. Остальные столбцы - найденные ближайшие соседи.

Для подтверждения гипотезы о полезности предложенного алгоритма сегментации объекта от фона для решения задачи классификации был проведен эксперимент, показывающий изменение точности классификации при использовании предложенного алгоритма. Для этого было выбрано три часто используемых на практике классификатора (LDA, Linear SVM, k-NN). Каждый из них был обучен на двух выборках. Первая состояла из синтетически созданных изображений объектов, вторая - из того же числа синтетических изображений, полученных путем применения предложенного алгоритма сегментации от фона к изображениям из первой выборки. Сравнение полученных результатов классификации (таблица 4, обучающий набор 1 и 3) показывает, что использование предложенного алгоритма позволяет существенно повысить точность классификации.

Обучающий набор #	Тип обучающей выборки	LDA [%]	Linear SVM [%]	k-NN [%]	CNN [%]
1	синтетическая	43.6	79.01	93.15	97.87
2	реальная	93.28	95.7	72.81	96.3
3	синтетическая, после сегментации от фона	83.22	91	96.91	---

Таблица 4. Сравнение точности работы различных классификаторов, обученных на реальных и синтетических данных базы GTSRB.

Проведенные эксперименты по сравнению вышеупомянутых классификаторов со сверточной нейронной сетью (CNN) показали, что использование CNN, обучаемой на большой выборке искусственных данных, позволяет достигать наилучших результатов (таблица 4, обучающий набор 1). Данные результаты можно объяснить в первую очередь тем, что современные реализации нейронных сетей способны к обучению на выборках большого размера, что позволяет использовать для их обучения большое количество искусственных данных, доступных в неограниченном количестве.

Также на базе российских знаков дорожного движения (RTSD) были проведены эксперименты по сравнению точности работы классификаторов, обученных на реальных и искусственно созданных данных. Результаты эксперименты показаны в таблице 5. Видно, что использование искусственных данных позволяет увеличить точность классификации.

Тип данных	Количество классов	Количество обучающих примеров на класс	Точность классификации [%]
Реальные	42	> 15	93.7
Синтетические	42	> 9000	94.1

Таблица 5. Сравнение точность классификации RTSD сверточной нейронной сетью при обучении на реальных и на синтетических данных.

В четвертой главе дано описание программной системы, разработанной на основе предложенных алгоритмов. Система имеет модульную структуру и состоит из следующих модулей:

1. модуль обнаружения объекта
2. модуль сегментации объекта от фона
3. модуль классификации отдельного изображения объекта
4. модуль слежения
5. модуль уточнения класса физического объекта
6. модуль локализации
7. модуль объединения результатов локализации
8. модуль визуализации результатов.

Модули обнаружения, сегментации и классификации работают на отдельных кадрах входной видеопоследовательности и готовят результат для анализа модулями более высокого уровня. Модули слежения, уточнения класса физического объекта и локализации работают на уровне соседних кадров видеопоследовательности. Модуль объединения результатов локализации комбинирует результаты работы системы после анализа видеопоследовательностей, полученных с одного и того же участка проезда.

В заключении перечислены основные результаты работы.

Основные результаты работы

В ходе работы над данной диссертацией разработаны и реализованы новые алгоритмы и программные модули, которые стали частью программного комплекса по автоматическому обнаружению, распознаванию и нанесению объектов на карту. К основным результатам работы можно отнести:

1. разработан алгоритм создания искусственной обучающей выборки, позволяющий сократить затраты на разметку и улучшить точность классификации объектов по сравнению с обучением на реальных данных
2. предложены способы оценки качества получаемых искусственных данных в сравнении с реальными изображениями

3. предложены модификации классического алгоритма выделения объектов Виолы-Джонса, позволяющие обучить модель на искусственно созданных данных, а также увеличить скорость и точность работы модуля обнаружения
4. разработан новый алгоритм сегментации объектов от фона. Алгоритм создан для обучения на искусственных данных и позволяет оценить параметры трансформаций объекта
5. предложена схема обучения на синтетических данных с использованием глубокой нейронной сети, позволяющей обучаться на выборках большого объема. За счет обучения на синтетических данных удалось улучшить точность классификации объектов по сравнению с обучением на реальных данных.

Публикации по теме диссертации

1. Chigorin A., Krivovyaz G., Velizhev A., Konushin A. A method for traffic sign detection in an image with learning from synthetic data // 14th International Conference Digital Signal Processing and its Applications. 2012. V. 2. P. 316-319.
2. Моисеев Б., Чигорин А. Классификация автодорожных знаков с помощью свёрточной нейросети, обученной на синтетических данных // Graphicon. 2012. P. 284-287.
3. Konev A., Chigorin A., Krivovyaz G., Velizhev A., Konushin A. Traffic signs recognition on images with training on synthetic data // Technical vision in computer systems. 2012. P. 65-66.
4. Moiseyev B., Konev A., Chigorin A., Konushin A. Evaluation of Traffic Sign Recognition Methods Trained on Synthetically Generated Data // Advanced Concepts for Intelligent Vision Systems. 2013. Springer LNCS. Vol. 8192. P. 576-83.
5. Чигорин А., Конушин А. Система автоматического картографирования знаков дорожного движения // Программные продукты и системы. 2013. С. 288-291.
6. Чигорин А., Конушин А. Эксперименты с обучением методов распознавания дорожных знаков на синтетических данных // Наука и образование: электронное научно-техническое издание. 2013. Т. 8. С. 315-24.
7. Chigorin A., Konushin A. A system for large-scale automatic traffic sign recognition and mapping // City Models, Roads and Traffic 2013. 2013. V. II-3/W3. P. 13-7.

8. Чигорин А. Автоматическое обнаружение 200 классов российских знаков дорожного движения // Digital Signal Processing and applications. 2013. V. 2. P. 187-190.
9. Чигорин А., Конушин А. Сборник тезисов конференции Ломоносовские чтения. 2013. P. 42-43.

Личный вклад автора

Все результаты, составляющие основное содержание диссертации, получены автором самостоятельно. Автору принадлежит идея использования искусственных данных для обучения классификаторов объектов. В работах [1-8] автору принадлежит идея алгоритма создания искусственных данных и способов оценки качества получаемых изображений в сравнении с реальными. В работе [2,6] автору принадлежит идея использования сверточной нейронной сети для обучения на большой выборке искусственно созданных данных. В работе [7,8] автором предложены модификации метода обнаружения объектов Виолы-Джонса, позволившие улучшить точность и скорость его работы при обучении на искусственной выборке. Алгоритм поиска параметров трансформаций объекта интереса и сегментации объекта от фона придуман автором и описан в работах [3,4].