

На правах рукописи



**МАТВЕЕВ ДМИТРИЙ ВЯЧЕСЛАВОВИЧ**

**РАЗРАБОТКА АЛГОРИТМОВ АНАЛИЗА АУДИТОРИИ  
ДЛЯ СИСТЕМ ПРИКЛАДНОГО ТЕЛЕВИДЕНИЯ**

Специальность 05.12.04

Радиотехника, в том числе системы и устройства телевидения

**АВТОРЕФЕРАТ**

диссертации на соискание ученой степени  
кандидата технических наук

Владимир – 2016

Работа выполнена на кафедре динамики электронных систем ФГБОУ ВПО «Ярославский государственный университет им. П.Г. Демидова»

Научный руководитель: **Приоров Андрей Леонидович**  
доктор технических наук, доцент,  
доцент кафедры динамики электронных систем  
ФГБОУ ВПО «Ярославский государственный  
университет им. П.Г. Демидова», г. Ярославль

Официальные оппоненты: **Орлов Алексей Александрович**  
доктор технических наук, доцент,  
заведующий кафедрой физики и прикладной  
математики Муромского института (филиала)  
ФГБОУ ВО «Владимирский государственный  
университет имени Александра Григорьевича и  
Николая Григорьевича Столетовых», г. Муром,  
Владимирская область

**Голубев Максим Николаевич**  
кандидат технических наук,  
инженер-проектировщик систем безопасности  
ООО «Профстройсеть», г. Москва

Ведущая организация: ОАО «Ярославский радиозавод», г. Ярославль

Защита диссертации состоится 29 марта 2016 г. в 16.00 часов на заседании диссертационного совета Д 212.025.04 при ФГБОУ ВО Владимирском государственном университете имени Александра Григорьевича и Николая Григорьевича Столетовых по адресу: 600000, г. Владимир, ул. Горького, д. 87, ВлГУ, корп. 3, ФРЭМТ, ауд. 301.

С диссертацией можно ознакомиться в библиотеке ФГБОУ ВО Владимирского государственного университета имени Александра Григорьевича и Николая Григорьевича Столетовых и на сайте <http://diss.vlsu.ru>.

Автореферат разослан «18» января 2016 г.

Отзывы на автореферат, заверенные печатью, просим направлять по адресу: 600000, г. Владимир, ул. Горького, д. 87, ВлГУ, ФРЭМТ.

Ученый секретарь диссертационного совета  
доктор технических наук, профессор



А.Г. Самойлов

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

**Актуальность темы.** На современном этапе развития науки и техники задачи анализа видеоданных представляют важный теоретический и практический интерес, как один из важнейших механизмов обеспечения эффективного взаимодействия электронной техники с человеком. С каждым годом появляется все больше камер видеонаблюдения и, соответственно, возрастает значимость решения задачи автоматического анализа видеоизображений для систем охранного телевидения. Одновременно с этим растет спрос на интеллектуальные системы, способные использовать подобный анализ для решения различных прикладных задач. К таким задачам относится, например, задача анализа аудитории, под которой понимается количественный и, в случае технической возможности, качественный (по полу, возрасту, расе, эмоциям и др.) анализ потока людей по одному или нескольким видеоизображениям, производимый, как правило, в режиме реального времени.

Ключевой проблемой систем анализа аудитории является решение задачи обнаружения людей на видеоизображениях при различных ракурсах видеокамер. Целью решения такой задачи является определение наличия человека на видеопоследовательности и нахождение его положения. Алгоритмы, решающие задачу обнаружения людей, лежат в основе современных интерфейсов взаимодействия систем прикладного телевидения с человеком, находящихся применение в следующих областях: охранное видеонаблюдение, робототехника, следящие системы, системы помощи водителю, системы спортивной видеоаналитики и т. д. В зависимости от ракурса камеры объектом интереса систем анализа аудитории может служить лицо человека (задача детектирования лиц), фигура человека (задача детектирования пешеходов) или верхняя часть головы (задача анализа видеоизображений с купольных (потолочных) камер).

Проблема детектирования и локализации лиц на изображениях систем прикладного телевидения может быть сформулирована следующим образом: в данном кадре видеопоследовательности определить наличие или отсутствие лиц людей и при положительном ответе найти границы прямоугольных рамок, целиком включающих лица. Среди существующих на сегодняшний момент алгоритмов детектирования лиц классическим считается подход на основе метода Виолы-Джонса, ставший де-факто стандартом для многих современных исследований. В настоящее время для решения практических задач исследуется множество других подходов к обнаружению лиц, среди которых можно выделить алгоритмы на основе гистограмм направленных градиентов, модели деформируемых частей, сверточных нейронных сетей, глубокого обучения и др. Основной проблемой таких алгоритмов обычно является их высокая вычислительная сложность, поэтому исследования в этой области остаются актуальными.

Задача детектирования фигуры человека также является практически востребованной в широком спектре приложений. Одной из наиболее актуальных и в то же время наиболее требовательных к качеству и скорости работы подобных систем сфер применения является детектирование пешеходов на видеопотоке с камеры, установленной на движущемся автомобиле, в режиме реального времени. Существенные сложности при решении данной задачи связаны с многообразием

внешнего вида людей, окружающих предметов и их взаимного расположения, с тем, что детектирование должно производиться на видеоизображении с движущейся камеры. Следует отметить, что создание алгоритма детектирования пешеходов, обеспечивающего требуемый уровень качества, в настоящий момент остается открытой проблемой.

Особый интерес представляют системы автоматического распознавания движений спортсменов для игровых видов спорта, таких как футбол, баскетбол, волейбол для получения статистической информации, предназначенной для анализа поведения отдельных игроков, команд, проведенных матчей. Отличительной чертой таких систем является тот факт, что для получения корректных результатов необходимо использовать видеоданные, исключая сильные взаимные перекрытия объектов. Под данное условие подходят видеоданные, полученные с помощью купольных камер. Задача детектирования головы человека при таком ракурсе остается малоизученной и представляет широкий научный и практический интерес.

В разработку алгоритмов цифровой обработки изображений внесли вклад как отечественные ученые – Ю.Б. Зубарев, М.И. Кривошеев, В.П. Дворкович, А.В. Дворкович, М.К. Чобану, А.С. Крылов, М.Н. Рычагов, Ю.С. Бехтин, Ю.С. Радченко, А.К. Бернюков, Д.С. Ватолин, С.В. Умняшкин, Е.П. Петров, И.С. Трубин, В.Ю. Волков, Р.Г. Хафизов, А.Л. Приоров, В.В. Хрящев, так и зарубежные – R. Chellappa, Z. Wang, A. Bovik, Y. Neuvo, J. Astola, T. Ojala, K. Egiazaryan, M. Nikolova, R. Szeliski, R. Lukac, T. Kanade.

Первые работы по распознаванию динамических объектов опубликованы еще в конце 1980-х гг. Следует отметить значительный вклад советских и российских ученых, работающих в данной области, таких как Ю.И. Журавлев, В.А. Сойфер, А.И. Галушкин, В.Н. Вапник, А.Я. Червоненкис, Б.А. Алпатов, М.Н. Фаворская, Ю.В. Визильтер, Е.В. Медведева, А.С. Конушин.

Для решения практических задач детектирования человека при разных ракурсах камеры актуальным вопросом остается поиск алгоритмов, работающих в режиме реального времени и позволяющих идентифицировать человека в условиях наличия искажений и помех на телевизионных изображениях. Поэтому на современном этапе развития науки и техники разработка и анализ алгоритмов детектирования человека для систем прикладного телевидения представляет собой актуальную научно-техническую задачу [1–4, 7–10, 12].

**Целью работы** является разработка и исследование новых алгоритмов детектирования человека на видеоизображениях в условиях наличия искажений и помех, обусловленных практическими требованиями для функционирования систем прикладного телевидения.

**Объектом исследования** являются алгоритмы детектирования человека на видеоизображениях, применяемые в радиотехнических системах обработки и анализа многомерных цифровых сигналов и в современных системах прикладного телевидения.

**Предметом исследования** являются разработка и модификация алгоритмов детектирования человека с целью повышения эффективности функционирования систем прикладного телевидения.

### **Задачи диссертационной работы**

- разработка и исследование алгоритма детектирования лиц на видеоизображениях, характерных для систем прикладного телевидения, с дополнительными требованиями на скорость работы и робастность к основным видам искажений и помех;
- разработка и исследование алгоритма детектирования фигуры человека при различном качестве входного видеоизображения;
- разработка и исследование алгоритма детектирования головы человека на видеоизображениях, полученных с помощью купольных камер.

**Методы исследования.** При решении поставленных задач использовались современные методы цифровой обработки изображений, технического зрения, распознавания образов, машинного обучения, теории вероятностей и математической статистики. Для практической реализации алгоритмов применялись современные численные методы и методы программирования на языках C#, Python, MatLab.

### **Научная новизна полученных результатов**

В рамках диссертационной работы получены следующие новые научные результаты:

- разработан алгоритм детектирования лиц на видеоизображениях с использованием ансамбля решающих деревьев;
- разработана модификация алгоритма Далала-Триггса с использованием классификатора на базе бустинга и переобучения на сложных примерах для детектирования пешеходов на видеоизображениях;
- разработан алгоритм детектирования головы человека на видеоизображениях, полученных с помощью купольных камер, с дополнительным классификатором на основе гистограмм направленных градиентов.

### **Практическая значимость полученных результатов**

1. Разработана методика и программное обеспечение для исследования алгоритмов детектирования человека на телевизионных изображениях при различных ракурсах видеокамер.
2. Проведен анализ работы алгоритмов детектирования лиц на телевизионных изображениях и предложен новый алгоритм, с использованием ансамбля решающих деревьев, позволяющий улучшить комплексные характеристики системы в среднем на 13 %.
3. Предложены, реализованы на языках высокого уровня и протестированы робастные алгоритмы детектирования головы человека для диагонального и вертикального ракурса камер, позволяющие повысить точность обнаружения человека по сравнению с известными подходами.

Разработанные алгоритмы требуют для своей практической реализации относительно небольших вычислительных ресурсов, что позволяет использовать их для обработки телевизионных изображений в системах реального или близкого к реальному режимам времени.

Результаты работы внедрены в соответствующие разработки ООО «Пиклаб» г. Ярославль и ООО «БАЙТ-ПРО» г. Ярославль. Отдельные результаты диссертационной работы внедрены в учебный процесс ЯрГУ им. П.Г. Демидова в

рамках дисциплин «Цифровая обработка изображений», «Компьютерное зрение». Получено свидетельство о государственной регистрации программы для ЭВМ [18].

**Достоверность полученных научных результатов** обусловлена применением адекватного математического аппарата, подтверждается их согласованностью с результатами проведенного компьютерного моделирования и сопоставлением ряда полученных результатов с научными данными, известными из российской и зарубежной литературы.

**Апробация работы.** Результаты работы докладывались и обсуждались на следующих научно-технических конференциях:

- 13–17-я Международные конференции «Цифровая обработка сигналов и ее применение» (DSPA), Москва, 2011–2015;
- 2-я Всероссийская научная конференция молодых ученых с международным участием «Теория и практика системного анализа», Рыбинск, 2012;
- 4–5-я Всероссийские конференции «Радиоэлектронные средства получения, обработки и визуализации информации», Нижний Новгород, 2014; Москва, 2015;
- Международная конференция «Радиоэлектронные устройства и системы для инфокоммуникационных технологий – РЭУС-2015», Москва, 2015;
- 12-я Международная научно-техническая конференция «Распознавание – 2015», Курск, 2015;
- 11-я Международная научно-техническая конференция «Перспективные технологии в средствах передачи информации – ПТСПИ-2015», Суздаль, 2015;
- 21-я Международная научно-техническая конференция «Радиолокация, навигация, связь», Воронеж, 2015.

**Публикации.** По теме диссертации опубликовано 17 научных работ, из них 4 статьи в рецензируемых журналах из перечня ВАК и 13 докладов на научных конференциях различного уровня.

**Структура и объем работы.** Диссертация состоит из введения, четырех глав, заключения, списка литературы, содержащего 115 наименований, и приложения. Она изложена на 114 страницах машинописного текста, содержит 29 рисунков и 18 таблиц.

**Основные научные положения и результаты, выносимые на защиту**

- Алгоритм детектирования лиц на видеоизображениях с использованием ансамбля решающих деревьев, увеличивающий площадь под ROC-кривой в среднем на 13 % по сравнению со стандартным алгоритмом Виолы-Джонса.
- Модифицированный алгоритм Далала-Триггса с использованием классификатора на базе бустинга и переобучения на сложных примерах, позволяющий на 10–25 % повысить точность детектирования фигуры человека на видеоизображениях.
- Алгоритм детектирования головы человека на видеоизображениях, полученных с помощью купольных камер, на основе бустинга с дополнительным классификатором на основе гистограмм направленных градиентов.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**Во введении** обоснована актуальность выбранной темы, сформулированы цель и задачи исследования, изложены основные положения и результаты, выносимые на защиту, показаны научная новизна и практическая значимость работы.

**В первой главе** рассмотрены современные задачи, методы и алгоритмы в области систем анализа аудитории. Представлен обзор существующих алгоритмов обнаружения людей, метрики оценки качества работы соответствующих детекторов. Рассмотрены современные алгоритмы машинного обучения на основе бустинга, локальных бинарных шаблонов (ЛБШ), ансамблей решающих деревьев (АРД). Их применение в системах прикладного телевидения открывает новые практические возможности для построения систем анализа видеоизображений реального времени [5, 6, 15–17].

**Во второй главе** предлагается новый алгоритм детектирования лиц на основе каскада ансамблей решающих деревьев. Предлагаемый подход является модификацией стандартного алгоритма детектирования Виолы-Джонса, с применением сканирования изображения каскадом бинарных классификаторов [1, 2, 7, 16]. Каждый бинарный классификатор состоит из ансамбля решающих деревьев, которые сравнивают интенсивность пикселей в качестве бинарного теста на своих внутренних узлах. Процесс обучения состоит из процедуры построения дерева регрессии на основе «жадного» алгоритма. Для решения задачи регрессии на основе изображения применяются оптимизированные бинарные решающие деревья, использующие сравнение интенсивности пикселей в качестве бинарных тестов в своих внутренних узлах. Бинарный тест по сравнению интенсивности для изображения  $I$  определяется как

$$\text{bintest}(I; l_1, l_2) = \begin{cases} 0, & I(l_1) \leq I(l_2) \\ 1, & \text{иначе} \end{cases}$$

где  $I(l_i)$  – интенсивность пикселя в точке  $l_i$ , а  $l_1$  и  $l_2$  являются нормированными координатами из набора  $[-1; +1] \times [-1; +1]$ .

Построение дерева контролируется. Обучающий набор данных – это набор  $\{(I_s, v_s, w_s): s = 1, 2, \dots, S\}$ , где  $v_s$  – контрольные данные, связанные с изображением  $I_s$ , а  $w_s$  – весовые коэффициенты. Бинарный тест в каждом внутреннем узле дерева выбирается так, чтобы минимизировать взвешенную среднеквадратическую ошибку, полученную, когда входные данные были разделены тестом. Это достигается минимизацией взвешенной среднеквадратической ошибки:

$$WMSE = \sum_{(I, v, w) \in C_0} w \cdot (v - \bar{v}_0)^2 + \sum_{(I, v, w) \in C_1} w \cdot (v - \bar{v}_1)^2,$$

где  $C_0$  и  $C_1$  – группы обучающих образцов, для которых результаты бинарного теста на изображении равны 0 и 1 соответственно. Скаляры  $\bar{v}_0$  и  $\bar{v}_1$  – взвешенное среднее контрольных данных в  $C_0$  и  $C_1$  соответственно. Глубина деревьев ограничена для уменьшения времени обучения, увеличения скорости обработки и в соответствии с требованиями памяти.

Одно решающее дерево, как правило, дает умеренную точность. С другой стороны, ансамбль деревьев может дать хорошие результаты. Для создания

дискриминантного ансамбля с помощью последовательной подстройки решающего дерева для соответствующей задачи наименьших квадратов используется алгоритм GentleBoost, являющийся модификацией хорошо известного алгоритма AdaBoost. Для того чтобы сгенерировать ансамбль из  $K$  деревьев из обучающего набора  $\{(I_s, c_s): s = 1, 2, \dots, S\}$  алгоритм проходит следующие шаги:

1. Задание начальных весов  $w_s$  для каждого изображения  $I_s$  и его отметки класса  $c_s \in \{-1; +1\}$  в виде:

$$w_s = \begin{cases} 1/P, & c_s = +1 \\ 1/N, & c_s = -1 \end{cases},$$

где  $P$  – общее количество положительных образцов, а  $N$  – общее количество отрицательных образцов.

2. Для  $k = 1, 2, \dots, K$ :

а) установить решающее дерево  $T_k$  с помощью взвешенных наименьших квадратов  $c_s$  для изображения  $I_s$  с весом  $w_s$ ;

б) обновление весов:

$$w_s = w_s \exp(-c_s T_k(I_s)),$$

где  $T_k(I_s)$  обозначает вещественный выход дерева  $T_k$  для изображения  $I_s$ ;

в) перенормировка весов так, чтобы их сумма была равна 1.

3. Выходной ансамбль  $\{T_k: k = 1, 2, \dots, K\}$ .

Так как получившийся детектор устойчив к малым изменениям в позиции и масштабе, вокруг каждой области интереса может быть несколько рамок. Эти перекрывающиеся рамки обнаружения объединяются в результате постобработки, если перекрытие между ними составляет более 30 %.

Для обучения фронтального детектора используется тестовая база GENKI, содержащая 3500 аннотированных лиц. С целью повышения обобщающей способности алгоритмов на классе лиц к исходным изображениям из базы применяется серия трансформаций. Из каждого изображения лица генерируется 15 положительных обучающих образцов с помощью небольших случайных изменений в позиции и масштабе лица, что делает детектор более устойчивым к шумам. Для обучения также использовалось 300 000 отрицательных образцов, не содержащих лица. Глубина каждого дерева фиксируется равной 6 и используется 20 ступеней классификации. Во время оптимизации для каждого внутреннего узла дерева проводится 256 бинарных тестов.

Для обучения детектора, выделяющего повернутые лица, составлен новый обучающий набор, состоящий из изображений баз лиц color FERET, CMU Multi-PIE, а также кадров, взятых из реальной видеозаписи. Этот набор содержит 2966 изображений с аннотированием рамкой и 4 ключевыми точками (центры глаз, кончик носа, центр рта), полученных посредством разметки в среде по аннотированию лиц. К данному набору также применены трансформации: поворот в плоскости на углы  $\pm 5^\circ, \pm 10^\circ$ , сдвиг изображения на  $\pm 2.5\%, \pm 5\%$ , изменение в масштабе  $\pm 5\%$ . В результате трансформирования на основе одного изображения получено 10 положительных обучающих образцов, т.е. для обучения использовалось 29660 изображений.

Тестирование и анализ характеристик различных детекторов лиц проводился на тестовой базе изображений Robotics, содержащей 6623 изображения 90 людей,



головы которых повернуты относительно вертикальной оси на углы в интервале  $[-90^\circ; +90^\circ]$  с шагом  $5^\circ$ .

На рис. 1 приводятся: блок-схема итогового алгоритма детектирования на основе АРД и результаты его сопоставления со стандартным алгоритмом на основе бустинга (Виолы-Джонса) и с фронтальной модификацией в виде ROC-кривых. На рис. 1б площадь под ROC-кривой равна соответственно: фронтальный АРД – 0,932, бустинг – 0,83, итоговый алгоритм АРД (фронтальный+поворот) – 0,956. Таким образом, показано, что предложенный итоговый алгоритм АРД увеличивает площадь под ROC-кривой почти на 13 % по сравнению с алгоритмом на основе бустинга.

Для оценки робастности алгоритма детектирования лиц на основе АРД смоделированы ситуации искажений изображения, характерных для систем прикладного телевидения: размытие, присутствие аддитивного белого гауссовского шума (АБГШ), импульсного шума, мультипликативного шума, сжатие алгоритмами JPEG и JPEG2000. Результаты исследования приведены на рис. 2(а-е). Их анализ показывает, что размытие и сжатие алгоритмами JPEG и JPEG2000 оказывают слабое влияние на работу алгоритма (уменьшение площади под ROC-кривой на 3,5 % при снижении метрики PSNR до 21,58 дБ для размытия, на 7,5 % при снижении метрики PSNR до 24,73 дБ для JPEG и на 0,3 % при снижении метрики PSNR до 31,79 дБ для JPEG2000). Искажения типа АБГШ, мультипликативный и импульсный шум вида «соль-и-перец» оказывают сильное влияние на работу алгоритма (уменьшение площади под ROC-кривой на 21,2 % при снижении метрики PSNR до 17,99 дБ для АБГШ, на 8,8 % при снижении метрики PSNR до 15,71 дБ для импульсного шума «соль-и-перец» и на 18,4 % при снижении метрики PSNR до 18,75 дБ для мультипликативного шума).

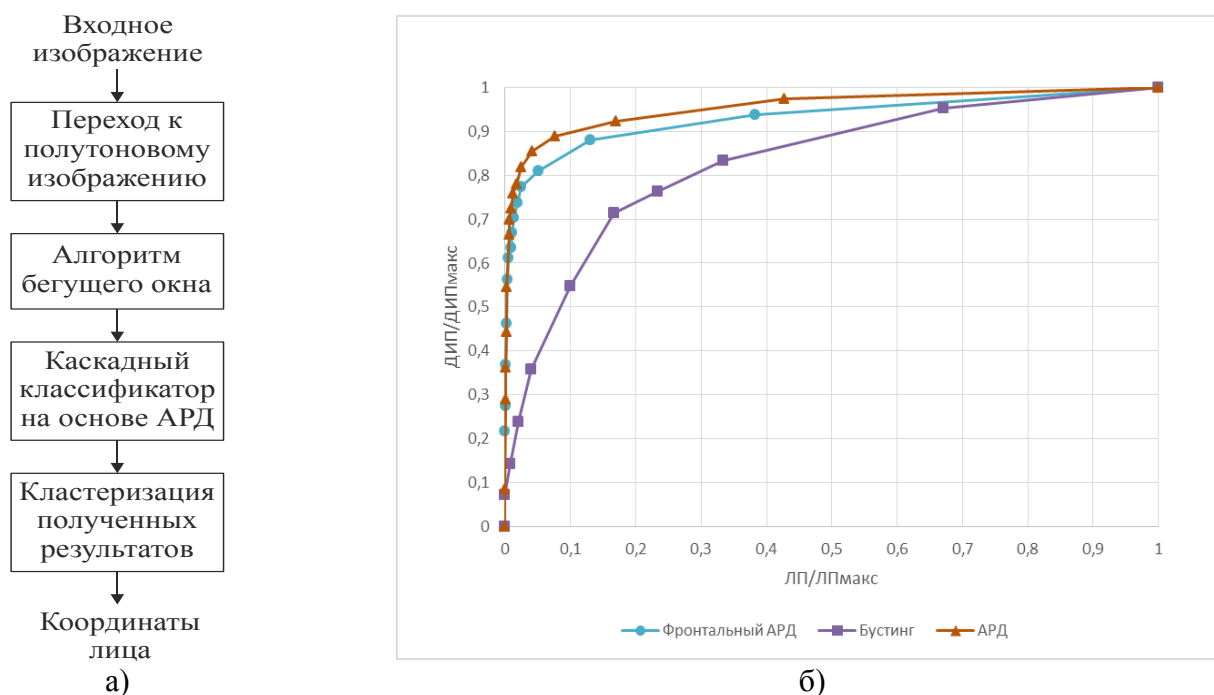
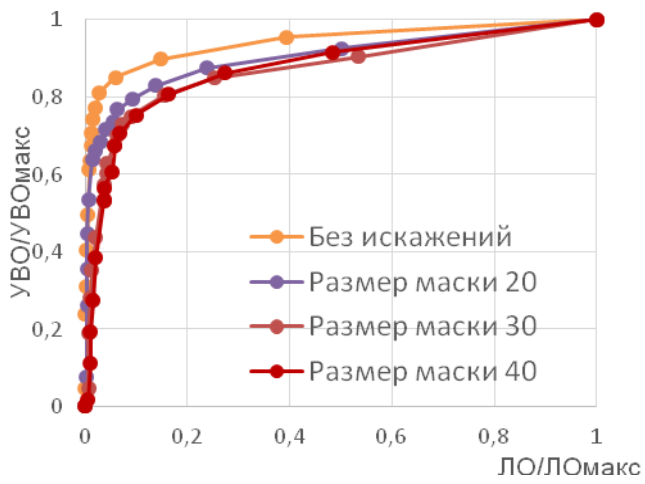
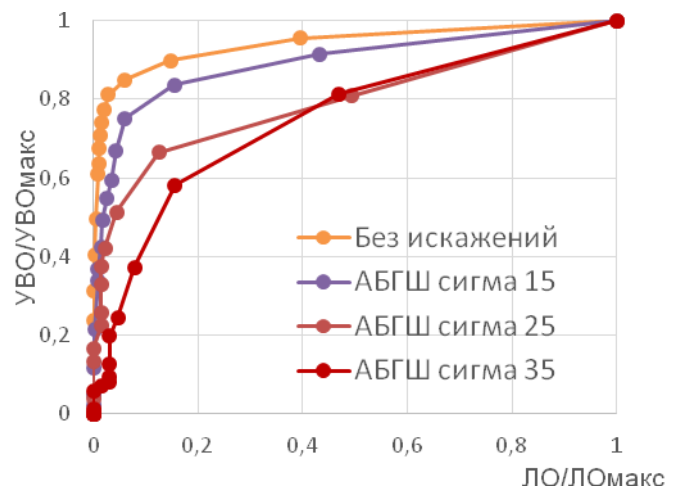


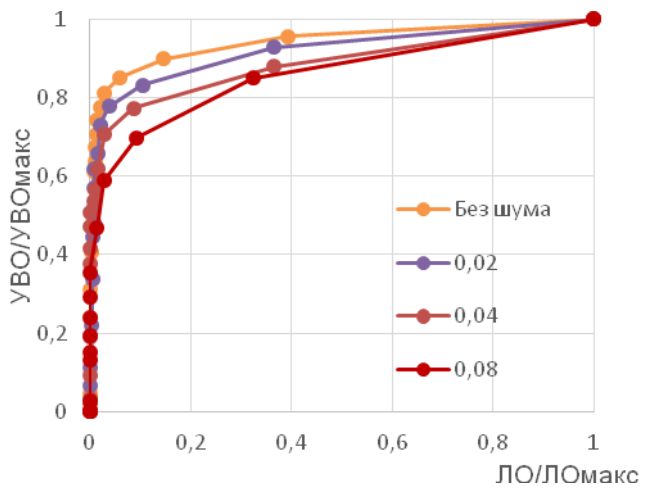
Рис. 1. Алгоритм детектирования на основе АРД:  
а) блок-схема; б) ROC-кривые для сравнения с алгоритмом бустинга (Виолы-Джонса)



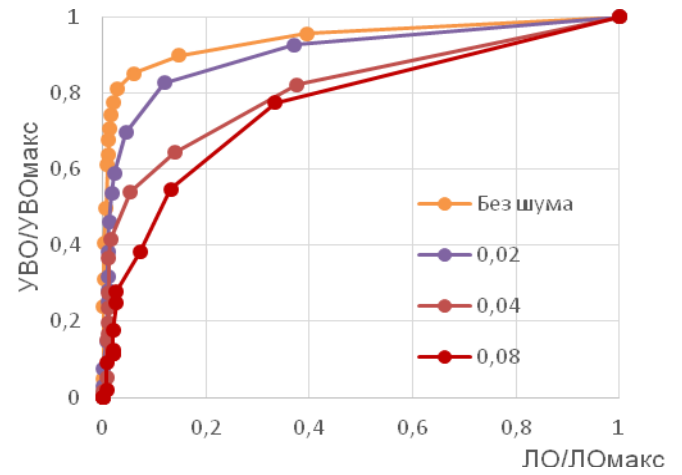
а)



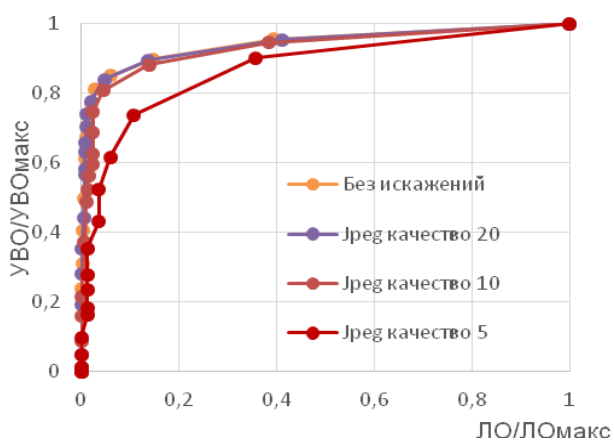
б)



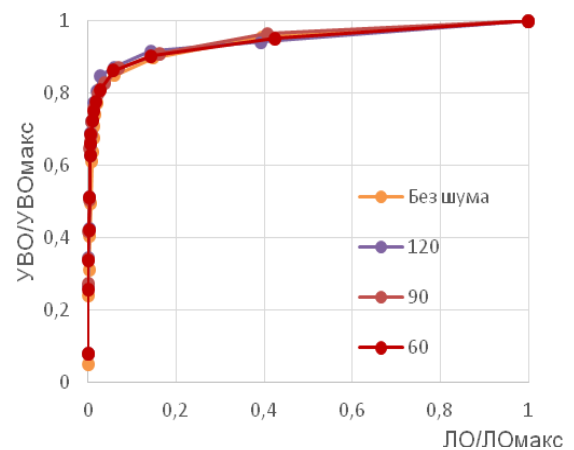
в)



г)



д)



е)

Рис. 2. ROC-кривые для моделей искажения тестового набора изображений:  
 а) размытие; б) гауссовский шум; в) импульсный шум; г) мультипликативный шум;  
 д) сжатие алгоритмом JPEG; е) сжатие алгоритмом JPEG2000

**В третьей главе** рассматривается модификация и исследование алгоритма детектирования фигур людей (пешеходов), основанного на НОГ-дескрипторах. Многие исследователи пытаются улучшить качество и скорость работы стандартного алгоритма Далала-Триггса путем замены алгоритма классификации на основе метода опорных векторов алгоритмом бустинга. При этом подходы к выбору слабых классификаторов, используемых в цепочке формирования сильного, могут значительно отличаться. В данной работе в качестве слабых классификаторов выбраны линейные машины опорных векторов, которые переобучались на сложных примерах [4, 8, 10, 11, 13]. Для этого в тестовой выборке отбирались те части изображений, на которых детектор на основе алгоритма Далала-Триггса дал неверные ответы. Эти сложные примеры приводились к стандартному разрешению окна детектора 64x128 пикселей. В процессе обучения финального классификатора модифицированного алгоритма акцент делается именно на эти сложные эталоны, которые распознаются хуже.

Исходная выборка, использованная для обучения детектора Далала-Триггса, состояла из  $M$  изображений  $(x_i, y_i)$ , характеризуемых вектором НОГ-признаков  $x_i$  и принадлежностью к классу «пешеходов»  $y_i \in \{-1; 1\}$ . К обучающей выборке добавлялось  $N$  изображений, на которых детектор Далала-Триггса дал некорректные ответы. Выбранные из базы для контрольной выборки изображения в последующем исключались из тестовой выборки. Таким образом, сформирована база из  $M + N$  изображений для обучения классификатора с помощью алгоритма на основе бустинга с начальными весами, инициализированными следующим образом:

$$D_1(i) = \frac{1}{M + kN} \text{ при } i = 1 \dots M \text{ и } D_1(i) = \frac{k}{M + kN} \text{ при } i = M + 1 \dots M + N,$$

где  $k$  – коэффициент усиления сложных примеров, выбираемый эмпирически.

Затем производится поиск слабого классификатора  $h_1(x)$ , разделяющего объекты классов «пешеход» и «не пешеход» согласно минимуму взвешенной ошибки  $\varepsilon_j$ :

$$h_1(x) = \arg \min_{h_j \in H} \varepsilon_j = \sum_{i=1}^{M+N} D_1(i) [y_i \neq h_j(x_i)].$$

Аналогичным образом производится поиск еще  $T - 1$  классификаторов. При этом на каждой стадии  $t = 1 \dots T$  поиска слабого классификатора происходит изменение весов объектов  $D_t(i)$  согласно соотношению

$$D_{t+1}(i) = \frac{D_t(i) \exp(-\alpha_t y_i h_t(x_i))}{Z_t},$$

где  $\alpha_t$  – коэффициент точности слабого классификатора  $h_t(x)$ , вычисляемый на основе ошибки, выдаваемой этим классификатором, и равный

$$\alpha_t = \frac{1}{2} \ln \left( \frac{1 - \varepsilon_t}{\varepsilon_t} \right),$$

а  $Z_t$  – нормировочный коэффициент, выбранный таким образом, чтобы совокупность весов  $D_{t+1}(i)$  удовлетворяла соотношению

$$\sum_{i=1}^{M+N} D_{t+1}(i) = 1.$$

Изменение весов объектов происходит таким образом, что ошибки слабого классификатора предыдущей стадии получают больший вес на новой стадии, правильно обнаруженные объекты – меньший. Итоговый сильный классификатор  $H(x)$ , который будет формировать решающую поверхность в пространстве HOG-признаков, строится в виде линейной комбинации всех слабых классификаторов  $h_i(x)$ :

$$H(x) = \text{sign}\left(\sum_{i=1}^T \alpha_i h_i(x)\right).$$

Качество работы детектора оценивается на тестовой выборке, не пересекающейся с обучающей. Для оценки качества детекторов используются стандартные характеристики: полнота (recall, R), точность (precision, P) и среднее гармоническое точности и полноты (F-мера, F).

Проведем сравнение алгоритма Далала-Триггса, алгоритм Виолы-Джонса и модифицированного алгоритма на задаче детектирования фигур людей. Для анализа работы алгоритмов использовались две базы изображений: INRIA Person Dataset – набор тестовых изображений со средним разрешением  $640 \times 480$  пикселей, снятых статичной камерой и CVC-02 Pedestrian Dataset – набор тестовых изображений, полученных из видеопоследовательности с разрешением  $640 \times 480$  пикселей, снятой с помощью видеорегистратора автомобиля.

Проведенный эксперимент на изображениях из базы INRIA показал значительный перевес алгоритмов, которые используют вектор HOG-признаков для описания изображения, над алгоритмом, использующим хааровские признаки. Модификация алгоритма Далала-Триггса позволила увеличить полноту на 9 %, а точность на 10 % в сравнении с оригинальным алгоритмом. Результаты детектирования, полученные по итогам эксперимента, приведены в табл. 1.

Таблица 1. Анализ работы детекторов на изображениях из базы INRIA Person

Детектор	R, %	P, %	F, %
Алгоритм Виолы-Джонса	27	62	38
Алгоритм Далала-Триггса	57	83	68
Модифицированный алгоритм	<b>66</b>	<b>93</b>	<b>77</b>

В случае с кадрами из видеопоследовательностей базы CVC-02 Pedestrian модифицированный алгоритм демонстрирует значительно более высокую точность работы (91 % против 65 %), чем алгоритм Далала-Триггса за счет переобучения исходного классификатора на ложных отрицательных срабатываниях детектора в области фона (табл. 2), но при этом требует больше времени для обучения и настройки.

Таблица 2. Анализ работы детекторов на изображениях из базы CVC-02 Pedestrian

Детектор	R, %	P, %	F, %
Алгоритм Виолы-Джонса	23	50	31
Алгоритм Далала-Триггса	25	65	36
Модифицированный алгоритм	<b>26</b>	<b>91</b>	<b>40</b>

Результаты работы алгоритма Далала-Триггса и модифицированного алгоритма на изображениях базы INRIA Person в условиях присутствия на

изображении АБГШ с нулевым математическим ожиданием и различной дисперсией по параметрам полноты и F-меры приведены на рис. 3.

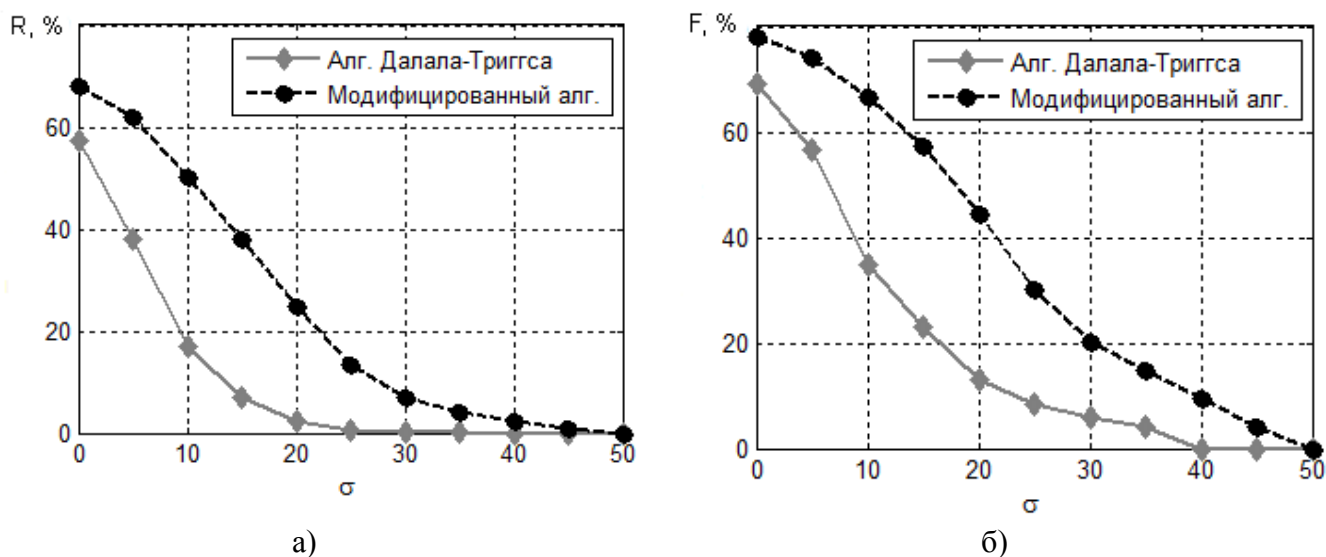


Рис. 3. Зависимость уровня детектирования от СКО гауссовского шума для алгоритма Далала-Триггса и модифицированного алгоритма: а) полнота; б) F-мера

Проведенное моделирование показывает, что модифицированный алгоритм детектирования фигур людей оказывается более робастным к гауссовскому шуму, что особенно заметно на средних степенях зашумления  $\sigma = 10-20$ , а это соответствует изображениям, получаемым в практических приложениях прикладного телевидения.

**В четвертой главе** проводится разработка алгоритмов детектирования голов на видеоизображениях с купольных камер. Анализ видеопотока с вертикально направленных камер – достаточно новое направление исследований, поэтому для обучения и тестирования алгоритмов использовался ряд видеопоследовательностей, подготовленных автором. Их разрешение варьировалось от  $640 \times 480$  до  $1920 \times 1080$  пикселей, высота подвеса камеры менялась от 2,5 до 7 метров. Изменялись также параметры плотности потока людей, скорости их перемещения, уровень освещенности в здании, тип используемой камеры (от недорогих IP-камер до известной экшн-камеры GoPro Hero3+) [9, 14, 15].

На основе записанных видеопоследовательностей составлены обучающие и тестовые наборы для детектирования людей (рис. 4). При этом часть разрабатываемых алгоритмов в качестве объекта детектирования использовала широкую область интереса (голова+плечи, рис. 4 а-в), а другая часть – узкую (только голова, рис. 4 г-е). К первой группе относятся алгоритмы на основе: детектирования движения, контекста формы и HOG-признаков. Усредненные результаты тестирования этих алгоритмов по метрикам полноты, точности и F-меры на наборе тестовых видеопоследовательностей приведены в табл. 3. Установлено, что наилучшие результаты получены для алгоритма на основе HOG-признаков. Однако следует отметить, что перечисленные выше алгоритмы слабо устойчивы к ситуациям плотного размещения людей внутри сцены, что ограничивает область их практического использования в современных системах прикладного телевидения.



Рис. 4. Примеры обучающих и тестовых изображений для детектирования голов: (а)-(в) с широким охватом области; (г)-(е) с узким охватом области

Таблица 3. Результаты тестирования с широким охватом области

Алгоритмы на основе	R, %	P, %	F, %
детектирования движения	88,0	88,5	88,2
контекста формы	78,6	66,0	71,8
НОГ-признаков	<b>92,0</b>	<b>97,0</b>	<b>95,8</b>

Алгоритмы второй группы объектом детектирования выбирают не широкий контур (голова+плечи), обладающий большой вариативностью, а только область головы. К подобным алгоритмам относятся разработанные алгоритм детектирования голов на основе бустинга и алгоритм на основе ЛБШ-признаков. Для улучшения качества их работы также реализован дополнительный постклассификатор на основе НОГ-признаков (рис. 5, где МОВ – метод опорных векторов). На вход алгоритма постклассификации подаются истинно положительные и ложно положительные результаты с алгоритмов классификации на основе бустинга и ЛБШ-признаков. Дополнительному классификатору необходимо выделить из всей выборки только истинно положительные результаты (рис. 6).



Рис. 5. Схема алгоритма с постклассификацией на основе НОГ-признаков



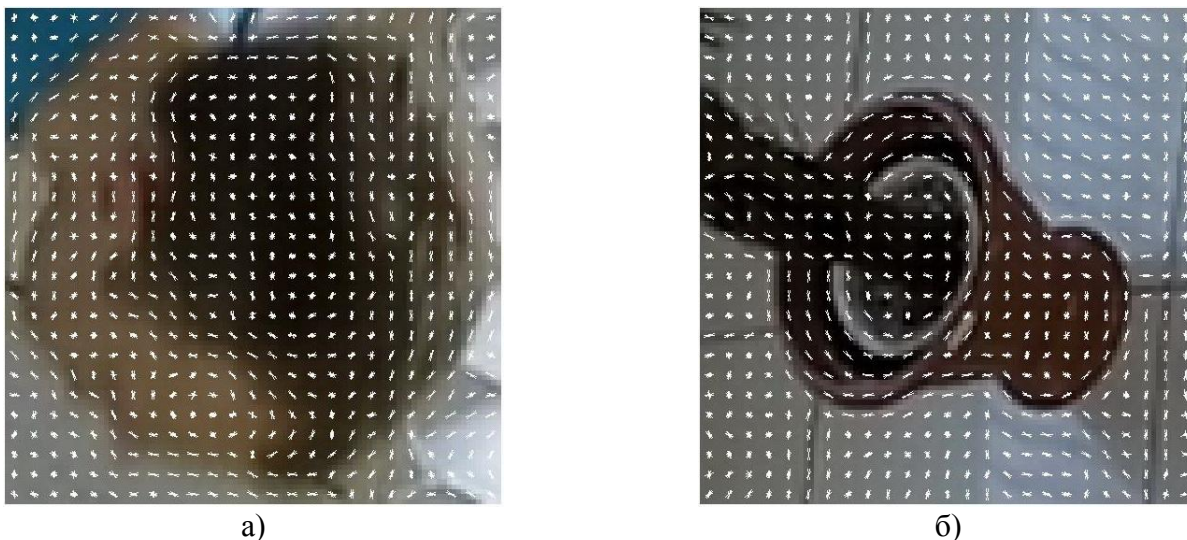


Рис. 6. Визуализация HOG-признаков для истинно положительных и ложно положительных результатов

Результаты тестирования алгоритмов на основе бустинга и ЛБШ-признаков с применением и без применения дополнительного классификатора приведены в табл. 4.

Таблица 4. Результаты тестирования алгоритмов с узким охватом области

Алгоритм	R, %	P, %	F, %
Бустинг	<b>84,3</b>	86,8	85,5
ЛБШ	75,6	75,4	75,5
Бустинг+HOG	<b>84,3</b>	<b>98,5</b>	<b>91,5</b>
ЛБШ+HOG	75,6	98,0	85,5

Проведенные исследования показывают, что алгоритм на основе бустинга имеет преимущество по исследуемым параметрам над алгоритмом ЛБШ, проигрывая ему по скорости работы в 2–2,5 раза. Применение дополнительного классификатора позволило увеличить показатель F-меры для алгоритма на основе бустинга на 6 %, а для алгоритма на основе ЛБШ на 10 %.

Для комплексной оценки качества работы системы к алгоритму детектирования голов на основе бустинга с дополнительным HOG-классификатором добавлен алгоритм трекинга на основе метода Лукаса-Канаде. Производился подсчет числа верно детектированных людей от их общего количества при различной высоте подвеса камеры и разной скорости перемещения людей. Исследования показали, что высота подвеса видеокамеры не оказывает существенного влияния на работу предложенной системы (доля правильно подсчитанных людей варьируется от 76 % до 83 %). При этом увеличение высоты подвеса видеокамеры приводит к росту вычислительной сложности работы алгоритмов. Увеличение скорости передвижения людей (с шага на легкий бег) снижает долю правильно подсчитанных объектов на 13 %.

Таким образом, предлагаемая система способна проводить количественный анализ аудитории при вертикальном подвесе видеокамеры с долей правильно подсчитанных людей от 80 % до 93 % в зависимости от параметров расположения камеры, условий освещения и скорости движения объектов внутри сцены.

## ОСНОВНЫЕ РЕЗУЛЬТАТЫ РАБОТЫ

Основные выводы и результаты диссертационной работы можно сформулировать в следующем виде.

1. Разработанный алгоритм детектирования лиц на основе АД увеличивает площадь под ROC-кривой на 13 % по сравнению со стандартным алгоритмом Виолы-Джонса. Предлагаемый подход позволяет обрабатывать видеоизображения с высокой скоростью, предоставляя возможность переобучения алгоритма на необходимых наборах данных, а также возможность построения на его основе каскада классификаторов для детектирования лиц, повернутых под разными углами относительно вертикальной оси. Алгоритм инвариантен к вращению изображения в плоскости экрана за счет использования при обучении множества копий оригинального изображения, повернутых на различные углы, а также небольших сдвигов.
2. Исследования робастности алгоритма детектирования на основе АД показывают, что АБГШ, мультипликативный и импульсный шум вида «соль-и-перец» оказывают сильное влияние на работу алгоритма (уменьшение площади под ROC-кривой на 21,2 % при снижении метрики PSNR до 17,99 дБ для АБГШ, на 8,8 % при снижении метрики PSNR до 15,71 дБ для импульсного шума «соль-и-перец» и на 18,4 % при снижении метрики PSNR до 18,75 дБ для мультипликативного шума).
3. Искажения типа размытие и сжатие алгоритмами JPEG и JPEG2000 оказывают слабое влияние на работу алгоритма детектирования на основе АД (уменьшение площади под ROC-кривой на 3,5 % при снижении метрики PSNR до 21,58 дБ для размытия, на 7,5 % при снижении метрики PSNR до 24,73 дБ для JPEG и на 0,3 % при снижении метрики PSNR до 31,79 дБ для JPEG2000).
4. Модификация алгоритма детектирования фигуры человека на основе метода Далала-Триггса с использованием классификатора на базе бустинга и переобучения на сложных примерах позволила значительно повысить результаты по метрике полноты (с 83 % до 93 % для базы INRIA Person и с 65 % до 91 % для базы CVC-02 Pedestrian). Использование модифицированного алгоритма позволяет достичь также увеличения метрики точности (с 57 % до 67 % для базы INRIA Person и с 25 % до 26 % для базы CVC-02 Pedestrian).
5. Модифицированный алгоритм детектирования фигуры человека показывает большую робастность к присутствию на видеоизображении гауссовского шума. Это особенно заметно на средних степенях зашумления ( $\sigma=10-20$ ), что соответствует изображениям, получаемым в практических приложениях прикладного телевидения.
6. Разработаны и исследованы пять алгоритмов детектирования людей на видеоизображениях с купольных камер (на основе: детектирования движения, контекста формы, НОГ-признаков, бустинга и ЛБШ). Установлено, что наилучшие результаты получены для алгоритмов на основе НОГ-признаков (большая область детектирования) и бустинга (малая область детектирования).
7. Разработан и исследован алгоритм постклассификации истинно положительных и ложно положительных результатов на основе вычисления НОГ-признаков. Применение дополнительного классификатора позволило увеличить показатель



- F-меры для алгоритма на основе бустинга на 6 %, а для алгоритма на основе ЛБШ на 10 %.
8. Установлено, что уровень подвеса камеры при моделировании спортивной съемки не оказывает существенного влияния на работу предложенного алгоритма детектирования на основе бустинга, а увеличение скорости передвижения людей (с ходьбы на легкий бег) снижает количество правильно подсчитанных людей на 13 %.
  9. Цель и задачи диссертационной работы успешно выполнены. Разработанные алгоритмы анализа аудитории интегрированы в программу VIStat 1.0 для автоматизированного расчета и визуализации статистики. Получено свидетельство о государственной регистрации программы для ЭВМ.

## **ОСНОВНЫЕ ПУБЛИКАЦИИ ПО ТЕМЕ ДИССЕРТАЦИИ**

### **Статьи в журналах из перечня ВАК**

1. Шмаглит Л.А., Хрящев В.В., Ганин А.Н., Матвеев Д.В. Распознавание пола человека по выделенной области лица на изображениях // Проектирование и технология электронных средств. 2011. № 4. С. 36–43.
2. Шмаглит Л.А., Приоров А.Л., Хрящев В.В., Матвеев Д.В. Детектирование лиц на изображениях в условиях аддитивного белого гауссовского шума // Электромагнитные волны и электронные системы. 2014. Т. 19, № 5. С. 62–70.
3. Никитин А.Е., Хрящев В.В., Приоров А.Л., Матвеев Д.В. Разработка и анализ алгоритма распознавания лиц на основе локальных квантованных шаблонов // Нелинейный мир. 2014. Т. 12, № 8. С. 35–42.
4. Приоров А.Л., Хрящев В.В., Никитин А.Е., Матвеев Д.В. Анализ и модификация алгоритмов детектирования пешеходов на цифровых изображениях // Радиотехнические и телекоммуникационные системы. 2014. № 4 (16). С. 40–47.

### **Материалы всероссийских и международных конференций**

5. Матвеев Д.В., Аминова Е.А., Ганин А.Н. Алгоритм отслеживания движения точечных объектов // Докл. 13-й Междунар. конф. «Цифровая обработка сигналов и ее применение» (DSPA–2011). Москва. 2011. Т. 1. С. 275–277.
6. Ганин А.Н., Вдовченко П.В., Матвеев Д.В. Разработка и анализ алгоритмов гендерной классификации по изображениям лиц // Матер. Междунар. науч.-практ. конф. «Молодежь и наука: модернизация и инновационное развитие страны». Пенза. 2011. Ч. 1. С. 221–223.
7. Ганин А.Н., Кисельникова О.А., Матвеев Д.В., Носков А.А. Применение метода цифровых гистограмм для детектирования повторного появления объекта на видеоданных // Докл. 15-й Междунар. конф. «Цифровая обработка сигналов и ее применение» (DSPA–2013). Москва. 2013. Т. 2. С. 96–100.
8. Лебедев А.А., Матвеев Д.В., Никитин А.Е., Хрящев В.В., Шемяков А.М. Анализ и модификация алгоритмов детектирования пешеходов на цифровых изображениях // Докл. 16-й Междунар. конф. «Цифровая обработка сигналов и ее применение» (DSPA–2014). Москва. 2014. Т. 2. С. 560–564.

9. Лебедев А.А., Матвеев Д.В., Хрящев В.В. Обработка видеопотока с потолочных камер // Докл. 17-й Междунар. конф. «Цифровая обработка сигналов и ее применение» (DSPA–2015). Москва. 2015. Т. 2. С. 703–707.
10. Матвеев Д.В. Разработка и анализ алгоритмов детектирования людей на видеоданных // Сб. докл. Междунар. конф. «Радиоэлектронные устройства и системы для инфокоммуникационных технологий – РЭУС–2015». Москва. 2015. С. 338–341.
11. Хрящев В.В., Приоров А.Л., Матвеев Д.В. Модификация алгоритма Далала-Триггса для обнаружения пешеходов на видеоданных // Сб. матер. XII Междунар. науч.-техн. конф. «Опτικο-электронные приборы и устройства в системах распознавания образов, обработки изображений и символьной информации. Распознавание – 2015». Курск. 2015. С. 383–385.
12. Хрящев В.В., Ганин А.Н., Манов И.А., Матвеев Д.В. Система анализа аудитории на основе алгоритмов детектирования, слежения и гендерной классификации // Сб. тр. II Всерос. науч. конф. молодых ученых с международным участием. Рыбинск. 2012. Т. 2. С. 136–142.
13. Матвеев Д.В., Никитин А.Е. Алгоритм Далала-Триггса и его модификация для детектирования человека на видеоданных // Докл. 4-й Всерос. конф. «Радиоэлектронные средства получения, обработки и визуализации информации». Нижний Новгород. 2014. С. 196–201.
14. Матвеев Д.В., Лебедев А.А., Хрящев В.В. Разработка алгоритмов детектирования объектов с потолочных камер видеонаблюдения // Докл. 5-й Всерос. конф. «Радиоэлектронные средства получения, обработки и визуализации информации». Москва. 2015. С. 135–138.
15. Матвеев Д.В. Разработка алгоритмов детектирования людей на видеоданных с купольных камер // Докл. 11-й Междунар. науч.-техн. конф. «Перспективные технологии в средствах передачи информации – ПТСПИ–2015». Суздаль. 2015. С. 64–166.
16. Хрящев В.В., Приоров А.Л., Никитин А.Е., Матвеев Д.В. Алгоритм распознавания лиц с использованием информации о расположении центра глаз // Сб. тр. XXI Междунар. науч.-техн. конф. «Радиолокация, Навигация, Связь». Воронеж. 2015. Т. 1. С. 177–187.
17. Матвеев Д.В. Об одном алгоритме распознавания движения на последовательности кадров // Сб. докл. 14-й Всерос. конф. «Математические методы распознавания образов». Москва. Макс-Пресс, 2009. С. 408–410.

**Свидетельство о государственной регистрации программы для ЭВМ**

18. Хрящев В.В., Матвеев Д.В., Ганин А.Н., Казина Е.М. VIStat 1.0 – автоматизированная система расчета и визуализации статистики // Свидетельство о государственной регистрации программы для ЭВМ № 2013612188 от 18 февраля 2013.

Матвеев Дмитрий Вячеславович

**РАЗРАБОТКА АЛГОРИТМОВ АНАЛИЗА АУДИТОРИИ  
ДЛЯ СИСТЕМ ПРИКЛАДНОГО ТЕЛЕВИДЕНИЯ**

Автореферат диссертации на соискание ученой степени  
кандидата технических наук

Подписано в печать 15.01.2016

Формат 60x84/16. Усл. печ. л. 1. Тираж 100 экз.

ИП Платонова И.В. 150000, Ярославль, ул. Б. Октябрьская 37/1