

На правах рукописи

Усилин Сергей Александрович

**Алгоритмическое развитие Виола-Джонсовских
детекторов для решения прикладных задач
распознавания изображений**

Специальность 05.13.01 – «Системный анализ, управление и
обработка информации (информационно-вычислительное обеспечение)»

Автореферат
диссертации на соискание учёной степени
кандидата технических наук

Москва – 2017

Работа выполнена в Федеральном государственном учреждении «Федеральный исследовательский центр «Информатика и управление» Российской академии наук», лаборатория № 9-5 «Взаимодействие человека и компьютера»

Научный руководитель: чл.-кор. РАН, д-р техн. наук, профессор
Арлазаров Владимир Львович

Официальные оппоненты:

Мясников Владислав Валерьевич
д-р физ.-мат. наук, доцент, федеральное государственное автономное образовательное учреждение высшего образования «Самарский национальный исследовательский университет имени академика С.П. Королева», профессор кафедры геоинформатики и информационной безопасности

Конущин Антон Сергеевич
канд. физ.-мат. наук, Федеральное государственное бюджетное образовательное учреждение высшего образования «Московский государственный университет имени М.В. Ломоносова», доцент кафедры автоматизации систем вычислительных комплексов

Ведущая организация:

Федеральное государственное бюджетное учреждение науки «Институт проблем управления им. В.А. Трапезникова Российской академии наук»

Защита диссертации состоится «__» ____ 2018 года в __:__ на заседании диссертационного совета Д 002.073.04 при Федеральном государственном учреждении «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН) по адресу: 117312, г. Москва, проспект 60-летия Октября, 9 (конференц-зал, 1-й этаж).

С диссертацией можно ознакомиться в библиотеке ФИЦ ИУ РАН по адресу: г. Москва, ул. Вавилова, д. 40.

Электронные версии диссертации и автореферата размещены на официальном сайте ФИЦ ИУ РАН по адресу: <http://www.frccsc.ru>

Отзывы и замечания по автореферату в двух экземплярах, заверенные печатью, просьба высыпать по адресу: 117312, г. Москва, проспект 60-летия Октября, 9, ФИЦ ИУ РАН, диссертационный совет Д 002.073.04

Автореферат разослан «__» ____ 2018 г.

Телефон для справок: +7(499) 135-51-64

Ученый секретарь диссертационного совета
д-р техн. наук, профессор

В.Н. Крутько

ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

В настоящее время для обработки информации активно применяются подходы математической теории распознавания. Содержательная постановка задачи обучения машин распознаванию образов появилась в конце 50-х годов и заключалась в построении способа обучения машин решать классификационные задачи так же, как это делают живые существа. К концу 60-х годов уже были разработаны различные подходы для решения задач распознавания в рамках статистических и перцептронных моделей. Большой вклад в развитие теории распознавания внесли советские (в последующем российские) ученые: С.А. Айвазян, М.А. Айзerman, М.М. Бонгард, В.Н. Вапник, Ю.И. Журавлев, Н.Г. Загоруйко, А.Г. Ивахненко, В.А. Ковалевский, В.Д. Мазуров, В.А. Сойфер, А.Я. Червоненкис, В.А. Якубович

К концу XX века область, связанная с распознавания образов на изображениях, находилась на достаточно высоком научном уровне. Однако большинство алгоритмов были предназначены для решения задачи классификации, не возвращая каких-либо знаний о точном местоположении объекта. При этом локализация объекта является важной задачей в области автоматической обработки цифровых изображений и находит свое применение в таких областях, как сегментация изображений вида объект/фон, анализ пространственных взаимоотношений объектов на изображении, определение и восстановление траекторий движения и т.п.

Революционный прорыв в области детектирования объектов был сделан американскими учеными Полом Виолой и Майклом Джонсом в 2001 году, предложившими алгоритм поиска лица на изображениях. Суть предложенного алгоритма заключается в объединении в единую конструкцию четырех подходов: использование вычислительно легковесных и обладающих хорошей обобщающей способностью признаков Хаара, обучение классификаторов с помощью бустинга, обеспечение высокой производительности за счет использования каскада классификаторов, а также применение метода скользящего окна для определения локализации объектов. Несмотря на то, что данный метод изначально разрабатывался для поиска лиц, он оказался применим для детекции широкого класса объектов, обладающих жесткой геометрией.

На сегодняшний день **степень разработанности** метода Виолы и Джонса достаточно высока. Исследователями по всему миру предложено множество модификаций оригинального алгоритма и отдельных его частей. Развитие признакового пространства происходит как относительно формы, так и относительно самой природы вычисления признака. Немало работ посвящено

модификациям метода обучения сильных классификаторов. Предложены новые модели каскадных классификаторов (например, «связанный каскад» и «легкий каскад»).

Однако высокую степень разработанности алгоритма Виолы и Джонса и обширное количество модификаций, применение его для решения большого диапазона прикладных задач в индустриальных системах распознавания часто оказывается затруднительным. Во-первых, значения классических признаков Хаара, даже в известных нормализованных модификациях, оказываются неинвариантны к изменению освещенности, а модификации признакового пространства, оперирующие с граничными точками, оказываются либо вычислительно трудоемкими, либо чувствительны к шумовым выбросам, либо неприспособленными к масштабированию. Во-вторых, алгоритм Виолы и Джонса, как и представленные модификации, решает задачу поиска объектов в «лабораторной» постановке: наборы прецедентов известны, зафиксированы и не требуют дополнительной кластеризации на отдельные подтипы, детекция объектов выполняется на отдельных стационарных изображениях. Однако индустриальные системы распознавания предъявляют к алгоритмам поиска объектов дополнительные требования. Так, например, в качестве источника данных часто выступает видеокамера, обеспечивающая вместо отдельных стационарных изображений коррелированную последовательность кадров, которые могут быть использованы для повышения производительности алгоритма. В качестве обучающих примеров часто выступают «сырые» данные, требующие предварительного разделения на подтипы. И, наконец, хотя в соответствии с оригинальным методом обучение детектора ведется в «пакетном режиме» (обучающая выборка известна и зафиксирована), на практике регулярно возникает задача «дообучения» детектора в связи с появлением новых данных.

Таким образом, алгоритмическое развитие метода Виолы и Джонса, устраниющее указанные недостатки, является высоко востребованным, а тема исследований диссертационной работы – **актуальной**.

Основная **цель** диссертационной работы заключается в повышении эффективности метода Виолы и Джонса (повышение производительности и качества детектирования объектов) при использовании в индустриальных системах распознавания.

Для ее достижения требуется решить следующие **задачи**:

1. Провести критический анализ оригинального метода Виолы и Джонса и существующих актуальных модификаций, а также исследовать индустриальные распознавающие системы на предмет дополнительных требований к методам детектирования объектов.

2. Разработать семейство вычислительно эффективных признаков, обеспечивающих высокий уровень обобщения объектов и устойчивых к различным параметрам освещенности, применимых для использования в индустриальных распознающих системах.

3. Построить математическую модель и разработать алгоритм обучения высокоуровневого классификатора, обеспечивающую функциональность дообучения при расширении обучающей выборки.

4. Построить математическую модель и разработать алгоритм многоклассовой детекции объектов на видеопоследовательности.

5. Разработать инструментарий, содержащий имплементацию полученных в рамках диссертации результатов и применимый для использования в индустриальных распознающих системах.

Методы исследования, используемые в диссертационной работе, включают в себя системный анализ, методы численного и натурного экспериментов, методы обработки цифровых изображений и распознавания образов, объектно-ориентированный анализ и проектирование.

Основные положения, выносимые на защиту:

1. Разработано семейство признаков для алгоритма Виолы и Джонса, устойчивых к различным параметрам освещенности и учитывающих геометрические особенности объектов.

2. Предложена математическая модель высокоуровневого классификатора Виолы и Джонса в виде решающего дерева сильных классификаторов, реализующая функциональность дообучения при расширении набора обучающих прецедентов, а также обеспечивающая лучшее по сравнению с каскадом качество детектирования за счет наличия нескольких положительных выходов.

3. Разработан алгоритм, сводящий обучение древовидного классификатора к последовательному обучению классических каскадных классификаторов, а также позволяющий выполнять дообучения высокоуровневого классификатора при расширении набора обучающих прецедентов.

4. Разработаны математическая модель и алгоритм адаптивного выбора распознающего классификатора Виолы и Джонса в задаче многоклассовой детекции объектов в видеопоследовательности.

5. Создан программный комплекс обучения классификаторов, реализующий представленные в диссертационной работе подходы и алгоритмы.

Научная новизна выносимых на защиту результатов состоит в следующем:

– в рамках диссертационной работы формулируется и решается задача адаптации метода Виолы и Джонса к использованию в индустриальных условиях, требующая разработки новых моделей и подходов;

– представлено новое семейство признаков, устойчивых к различным параметрам освещенности и учитывающих геометрические особенности объектов за счет использования в качестве исходных данных карты направленных границ;

– предложена новая модель высокоуровневого классификатора в виде решающего дерева, обеспечивающая несколько выходов с положительным исходом, а также представлен алгоритм его обучения и дообучения в случае расширения выборки прецедентов;

– впервые предложен алгоритм многоклассовой детекции объектов на видеопоследовательности, использующий методы обучения машин с подкреплением.

Теоретическая и практическая значимость работы. Диссертационная работа имеет как теоретическую, так и практическую значимость.

Теоретическая значимость работы заключается в первую очередь в постановке исследуемой задачи, предложенном семействе признаков, а также в разработанных моделях и алгоритмах для описания структуры высокоуровневого классификатора в виде решающего дерева и задачи многоклассовой детекции объектов на видеопоследовательности. Полученные результаты могут быть использованы для дальнейшего развития науки в данной области.

Практическая значимость работы подтверждается использованием полученных результатов в следующих индустриальных системах распознавания. Результаты диссертационной работы внедрены в систему классификации автомобилей по геометрическим характеристикам АКТС-4 (эксплуатируется на 200 полосах пунктов взимания платы автомагистралей России М1 «Беларусь», М4 «Дон» и Западного Скоростного Диаметра Санкт-Петербурга), в программные продукты распознавания документов Smart IDReader и Smart PassportReader (интегрированы в информационные решения крупнейших российских банков, страховых компаний и организаций различных государственных структур), а также в интеллектуальную систему автономного вождения C-Pilot (система содействия водителю, предназначена для установки на коммерческий транспорт).

Степень достоверности и апробация результатов. Основные положения и результаты диссертационной работы докладывались в виде очного доклада на следующих международных научных конференциях:

– Международной IEEE конференции «Image Processing» (ICIP'2010), Гонконг, 2010;

– Международной конференции «The 7th International Conference on Machine Vision» (ICMV'2014), Милан, Италия, 2014;

– Международной конференции «The 9th International Conference on Machine Vision» (ICMV'2016), Ницца, Франция, 2016;

- Международной конференции «The 10th International Conference on Machine Vision» (ICMV'2017), Вена, Австрия, 2017;
- Междисциплинарной школе-конференции Института проблем передачи информации им. А. А. Харкевича Российской академии наук (ИППИ РАН) «Информационные технологии и системы» (ИТИС) в 2015, 2014 и 2012 годах;
- Открытом германо-российском семинаре «8th Open German-Russian Workshop «Pattern Recognition and Image Understanding», г. Нижний Новгород, 2011.
- Научной конференции МФТИ «Современные проблемы фундаментальных и прикладных наук» в 2010, 2012 и 2013 годах.

Помимо научных конференций результаты докторской работы были обсуждены на международном научно-исследовательском семинаре Национального комитета при Президиуме РАН по распознаванию образов и анализу изображений «Анализ и понимание изображений».

Публикации. Основные научные результаты докторской работы изложены в 23 публикациях, в том числе: пять статей в изданиях, рекомендованных ВАК РФ, три публикации в изданиях, входящих в международные базы цитирования Scopus и Web of Science, патент на изобретение, два патента на полезную модель, а также три свидетельства на программу для электронных вычислительных машин (ЭВМ).

Структура и объем докторской работы. Докторская работа состоит из введения, пяти основных разделов, заключения, списка используемых по тексту работы сокращений и условных обозначений, списка использованной литературы, списка публикаций автора по теме докторской работы, а также трех приложений. Общий объем работы составляет 149 страниц (с учетом приложений), содержит 36 рисунков и 10 таблиц. Список литературы представлен 128 наименованиями.

ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

Во **введении** обоснована актуальность работы, показана степень разработанности данной научной области. Сформулированы цель и задачи исследования. Приведен перечень научных результатов и положения, выносимые на защиту.

В **первом разделе** содержится обзор оригинального метода поиска объектов Виолы и Джонса, представлены определения и теоремы, необходимые для подробного описания метода.

Метод Виолы и Джонса представляет из себя схему построения детекторов объектов с достаточно жесткой геометрией статистическим образом (опираясь на прецеденты). Изначально метод был применен для построения детектора человеческого лица анфас. Обучающая выборка состояла из 4916 изображений лиц,

приведенных к разрешению 24×24 пикселя (положительная выборка) и 9500 полноразмерных изображений, не содержащих лиц, из которых произвольным образом вырезались области, не содержащие лица (отрицательная выборка).

В качестве признакового метода Виолы и Джонса использует признаки Хаара (см. рисунок 1), значение которых вычисляется как разность сумм яркостей пикселов областей изображения внутри черных и белых прямоугольников. Для эффективного вычисления значения признаков Хаара используется интегральное представление изображения $I_f(y, x)$, которое для серого изображения $f(y, x)$ с размерами $M \times N$ определяется следующим образом:

$$I_f(y, x) = \sum_{i < y, j < x} f(i, j). \quad (1)$$

С каждым признаком Виолы и Джонс связывают бинарный «слабый» классификатор $h(x): \mathbb{X} \rightarrow \{-1, +1\}$, представленный в виде распознающего дерева с одним ветвлением. Такие классификаторы демонстрируют слабую распознающую силу, поэтому в методе Виолы и Джонса используется алгоритм AdaBoost для построения «сильного» классификатора в виде линейной комбинации наиболее мощных слабых классификаторов:

$$S(x) = \left[\sum_{t=1}^T \alpha_t \cdot h_t(x) > 0 \right], \quad (2)$$

где $[\cdot]$ - индикаторная функция.

Высокая производительность в методе Виолы и Джонса дополнительно обеспечивается за счет использования каскада классификаторов, который представляет собой произведение сильных классификаторов и позволяет быстро (на ранних этапах вычисления) распознавать «пустые» изображения (то есть изображения, не содержащие целевой объект):

$$Cascade(x) = \prod_{i=1}^N [S_i(x) > 0], \quad (3)$$

Поиск объекта на изображении выполняется с помощью построенного бинарного каскадного классификатора с помощью метода скользящего окна.



Рисунок 1 — Признаки Хаара и пример расположение признака относительно сканирующего окна

В разделах 1.4 и 1.5 диссертации рассмотрены существующие модификации метода в части используемого пространства признаков, обучающего алгоритма и каскадной структуры классификатора, а также проведен анализ применимости алгоритма и модификаций для решения задач детектирования объектов в индустриальных распознающих системах. В ходе анализа установлено, что несмотря на общность, декларированную в работах Виолы и Джонса, развитие данного метода для решения прикладных задач распознавания изображений является актуальной задачей.

Во втором разделе представлено новое семейство признаков, устойчивых к различным параметрам освещенности и не обладающих характерными яркостными особенностями. Данные признаки представляют собой прямоугольные признаки Хаара, вычисляемые поверх карты направленных границ, что позволяет существенно улучшить их обобщающую силу.

Карта направленных границ строится с помощью модифицированного метода Кэнни, который в качестве ответа возвращает множество пар $((y, x), d(y, x))$, где $d(y, x)$ – преимущественное направление границы в точке (y, x) , определяемое как дискретизация угла границы на горизонтальное, вертикальное, $+45^\circ$ и -45° направления. В работе предлагается использовать два типа карты направленных границ:

1. *Карта диагональных границ*, содержащая преимущественно границы вдоль диагоналей (см. рисунок 2б), вычисляемая следующим образом:

$$f_{de}(y, x) = \begin{cases} 0, & \text{если } (y, x) - \text{границчная точка и } d(y, x) = +45^\circ \\ 255, & \text{если } (y, x) - \text{границчная точка и } d(y, x) = -45^\circ \\ 128, & \text{иначе} \end{cases}. \quad (4)$$

2. *Карта прямых границ*, содержащая преимущественно горизонтальные и вертикальные границы (см. рисунок 2г), вычисляемая следующим образом:

$$f_{se}(y, x) = \begin{cases} 0, & \text{если } (y, x) - \text{границчная точка и } d(y, x) = 90^\circ \\ 255, & \text{если } (y, x) - \text{границчная точка и } d(y, x) = 0^\circ \\ 128, & \text{иначе} \end{cases}. \quad (5)$$

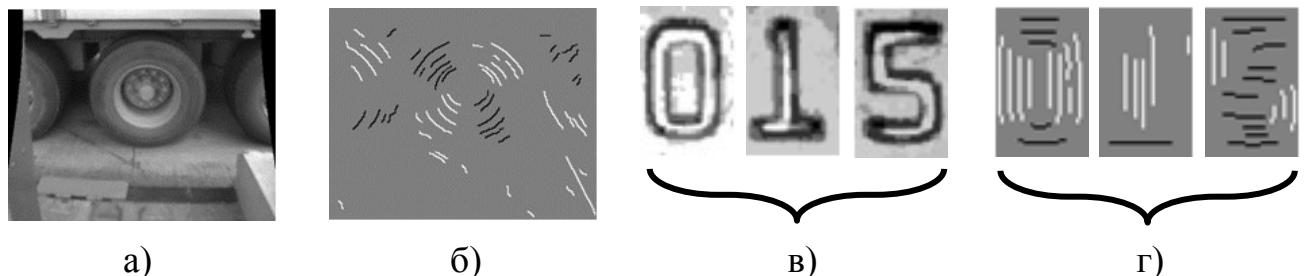


Рисунок 2 – Иллюстрация карты направленных границ автомобильного колеса и цифр номера банковской карты

Поверх построенной карты направленных границ (диагональной или прямой) вычисляются прямоугольные признаки Хаара.

Таким образом, разработанные признаки обладают следующими свойствами:

- учитывают геометрические особенности объектов;
- инвариантны к монотонному преобразованию пикселей изображения;
- вычислительно эффективны (сложность вычисления значения признака не зависит от геометрических размеров базовых прямоугольников);
- нечувствительны к незначительным геометрическим искажениям объектов;
- линейно масштабируемы (облегчает задачу поиска объектов при различных масштабах).

Эффективность описанных признаков была испытана в процессе решения двух задач распознавания: распознавание образов колес на фотографиях в боковом ракурсе и распознавание номера кредитной карты (рисунок 2). В обоих случаях на одной и той же обучающей выборке строились классификаторы на базе яркостных и контурных признаках Хаара. Для оценки качества работы построенных детекторов использовались метрики *точность* (*precision*) и *полнота* (*recall*), определяемые следующим образом:

$$Precision = \frac{TP}{TP + FP}, \quad Recall = \frac{TP}{TP + FN}, \quad (6)$$

где TP – число верно обнаруженных объектов, FP – число ложно обнаруженных объектов, FN – число ложно пропущенных (то есть не обнаруженных) объектов.

Анализ задачи детектирования колес показал, что на изображениях автомобилей в боковом ракурсе имеется большое количество вертикальных и горизонтальных границ (днище автомобиля, бензобак грузовика и т.д.), в то время как границ, направленных по диагоналям мало – и значительное количество из них как раз относится к колесам. Следовательно, классификатор, обученный поверх карты диагональных границ должен обеспечивать высокое качество. В процессе эксперимента оказалось, что классификатор, построенный поверх контурных признаков обеспечивает большую полноту ($recall = 0.80$), чем аналогичный классификатор, построенный поверх яркостных признаков ($recall = 0.64$) при одном и том же уровне точности ($precision = 0.95$).

Аналогичный результат был получен при тестировании разработанных признаков в задаче распознавания номера банковской карты. В отличии от задачи распознавания колес, использовалась карта прямых границ (см. рисунок 1г). В таблице 1 приведено сравнение полноты «контурного» и «яркостного» классификаторов цифр при практически идеальной точности. Как и в первом

эксперименте «контурные» классификаторы для каждой цифры продемонстрировали более высокое качество распознавания, чем соответствующие «яркостные» классификаторы.

Таблица 1 — Значение полноты (*recall*) контурного и яркостного классификаторов при значении точности *precision* = 0.99

Цифра	Контурный классификатор	Яркостной классификатор
0	0,91	0,45
1	0,89	0,57
2	0,72	0,49
3	0,64	0,52
4	0,96	0,54
5	0,83	0,52
6	0,86	0,54
7	0,78	0,68
8	0,86	0,48
9	0,75	0,41

В третьем разделе представлена модель классификатора Виолы и Джонса в виде решающего дерева, а также разработан алгоритм обучения такого классификатора.

Решающее дерево сильных классификаторов представляет собой вид бинарного решающего дерева: узел дерева – это сильный классификатор, на правое ребро которого попадают подокна, предположительно содержащие объект, а на левое – те, которые не распознались как объект, соответственно. Окончательный ответ дается только в листьях. Классический каскадный классификатор, описанный в оригинальной работе Виолы и Джонса – это, по сути, древовидный классификатор, содержащий лишь один «положительный» выход (лист) и множество «отрицательных» выходов.

В диссертации показывается, что любой путь от корня до самого нижнего узла древовидного классификатора может быть представлен как каскад, в котором отдельные сильные классификаторы входят с инвертированным ответом.

Под *взвешенным каскадом* будем подразумевать классический линейный каскад, у которого для каждого уровня (сильного классификатора) $S_i(x)$ дополнительно задан вес (знаковый множитель) $w_i = \pm 1$ с которым данный уровень участвует в произведении при вычислении значения каскада.

Исходя из определения, формула вычисления взвешенного каскада записывается следующим образом:

$$WeightedCascade(x) = \prod_{i=1}^N [w_i \cdot S_i(x) > 0]. \quad (7)$$

Аналогично классическому каскаду, взвешенный каскад строится итерационно, уровень за уровнем. Но в отличии от классического каскада, где при обучении очередного уровня обновляется только отрицательная выборка, состоящая из ложных срабатываний уже обученных уровней каскада, при построении очередного уровня взвешенного каскада обновляется так же и положительная выборка, состоящая из успешно пройденных сквозь уже обученный взвешенный каскад положительных прецедентов.

Таким образом, разработанный алгоритм построения древовидного классификатора представляет собой последовательное обучение отдельных путей (взвешенных каскадов). Формальное описание алгоритма представлено в диссертации как в виде последовательности шагов, так и в виде блок-схемы.

Разработанный алгоритм построения высокоуровневого классификатора в виде решающего дерева позволяет обучать классификаторы объектов, обладающих высокой вариативностью. Благодаря возможности периодически «перебрасывать» положительные образцы на отрицательную ветку, представленный алгоритм выполняет естественную кластеризацию положительной выборки.

Кроме того, представленный в данном разделе алгоритм позволяет выполнять дообучение существующего классификатора (при расширении обучающей выборки). За счет появления новых листьев с положительным исходом повышается полнота классификатора. Зная долю верных ответов (*true positive rate*, *TPR*) отдельных уровней (сильных классификаторов), можно оценить *TPR* для всего древовидного классификатора. Действительно, для некоторого взвешенного каскада *WC* уровней доля верных ответов определяется следующим образом:

$$TPR_{WC} = \prod_{i:w_i>0} tpr_i \cdot \prod_{i:w_i<0} (1 - tpr_i), \quad (8)$$

где tpr_i – доля верных ответов отдельного i -го уровня. Тогда *TPR* древовидного классификатора в целом определяется как сумма долей верных ответов всех входящих в дерево взвешенных каскадов:

$$TPR = \sum_{WC} TPR_{WC} = \sum_{WC} \left(\prod_{i:w_i>0} tpr_i \cdot \prod_{i:w_i<0} (1 - tpr_i) \right). \quad (9)$$

Эффективность разработанного алгоритма оценивалась при решении двух задач: распознавание образов колес на фотографиях в боковом ракурсе и в задаче поиска логотипа платежной системы VISA на изображениях банковских карт.

Как упоминалось ранее, автомобильное колесо в боковом ракурсе – достаточно вариативный объект (см. рисунок 3а), следовательно для качественного детектирования такого объекта необходим сложный классификатор. В процессе решения задачи на одном и том же обучающем наборе было проведено два эксперимента. Первый эксперимент представлял собой обучение классического каскада Виолы и Джонса, в результате которого был построен каскад, состоящий из 20 сильных классификаторов. Второй эксперимент заключался в обучении древовидного классификатора с помощью разработанного алгоритма. В результате был построен древовидный классификатор, содержащий в общей сложности 17 сильных классификаторов (см. рисунок 3б). Качество работы обученных классификаторов, посчитанное на тестовом наборе, состоящем из 12437 изображений и содержащем 2621 образов колес, представлено в таблице 2.

Таблица 2 — Качество детектирования каскадного и древовидного классификаторов автомобильного колеса в боковом ракурсе

Классификатор	TP	FP	FN	Precision	Recall
Каскадный	2452	57	169	0,9773	0,9355
Древовидный	2587	61	34	0,9770	0,9870

Несмотря на кажущуюся однотипность логотипов платежной системы VISA, с точки зрения обучения классификаторов они достаточно разнообразны (см. рисунок 4а). Аналогично, было проведено два эксперимента. Первый эксперимент состоял в обучении классического линейного каскада, в результате которого был построен каскад из 10 уровней. Второй эксперимент заключался в дообучении с помощью разработанного в диссертации алгоритма имеющегося каскада до древовидного классификатора с целью повышения полноты. Полученный в результате древовидный классификатор состоял в общей сложности

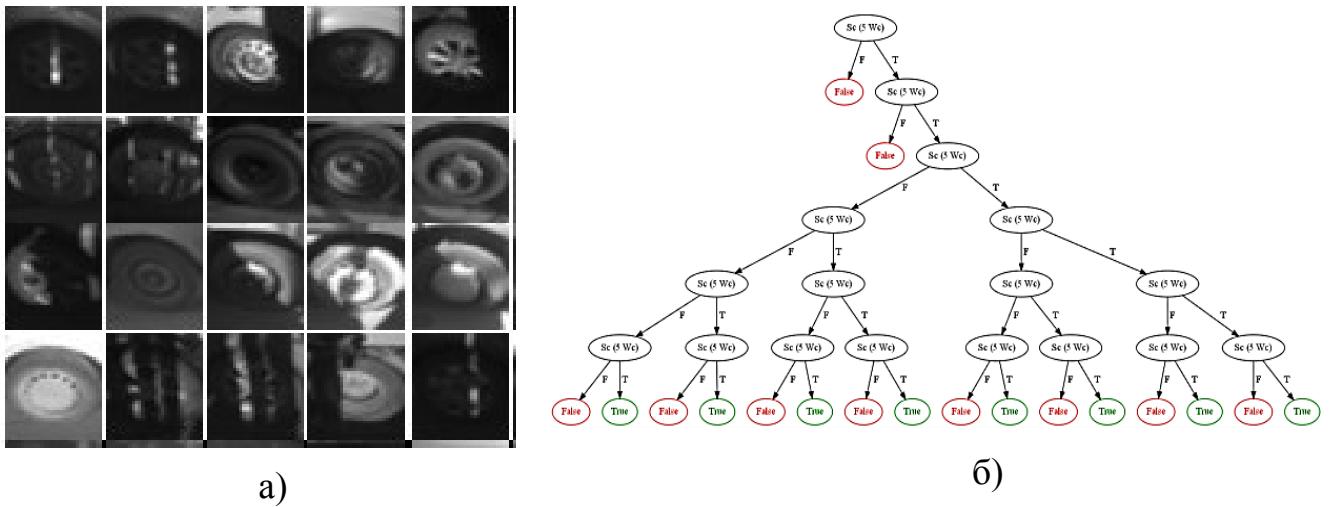


Рисунок 3 — Примеры изображений колес в боковом ракурсе (а) и визуализация обученного древовидного классификатора (б)

из 21 сильного классификатора (см. рисунок 4б). Качество работы обученных классификаторов логотипа платежной системы VISA представлено в таблице 3.

Таким образом, анализируя полученные результаты, отметим следующие особенности разработанного алгоритма:

- разработанный алгоритм позволяет обучать эффективные классификаторы для вариативных объектов, не проводя предварительную кластеризацию обучающей выборки;

- дообучение линейного каскада до древовидной структуры позволяет повысить полноту детектора, немного потеряв при этом в точности.

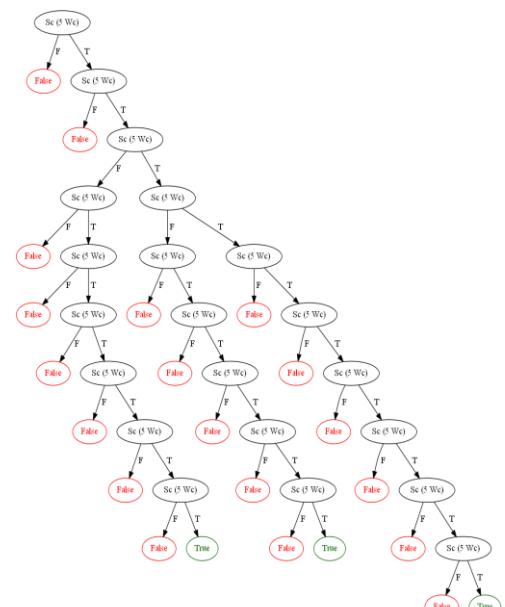
Таблица 3 — Качество детектирования каскадного и древовидного классификаторов логотипа платежной системы VISA

Классификатор	TP	FP	FN	Precision	Recall
Каскадный	1080	111	96	0,9068	0,9184
Древовидный	1090	115	86	0,9046	0,9269

В четвертом разделе описан оригинальный метод адаптивного выбора распознающих классификаторов Виолы и Джонса в задаче многоклассовой детекции объектов в видеопотоке для систем распознавания, работающих в режиме реального времени, в случае, когда в каждый момент времени на кадре может присутствовать не более одного объекта. Метод использует информацию, накопленную по уже обработанным видеокадрам, с целью выбора наилучшего классификатора для текущего видеокадра.



a)



б)

Рисунок 4 — Примеры логотипов платежной системы VISA (а) и визуализация обученного древовидного классификатора (б)

Рассматривается следующая задача. Пусть $\mathbb{V} = \{F_i\}_{i=1}^T$ – видеофрагмент, содержащий T кадров. На каждом кадре может присутствовать не более одного объекта, то есть каждому кадру F_i можно поставить в соответствие число $j_i^* \in \{0, 1, \dots, N\}$, где j_i^* – индекс типа объекта, причем $j_i^* = 0$ соответствует отсутствию какого-либо объекта на рассматриваемом кадре. Пусть $\mathbb{C} = \{C^j(F) : F \rightarrow \{0, j\}\}_{j=1}^N$ – множество обученных классификаторов Виолы и Джонса. Требуется построить правило адаптивного выбора классификатора для очередного кадра из видеофрагмента $a : \mathbb{V} \rightarrow \mathbb{C}$ так, чтобы максимизировать количество правильно найденных объектов.

Предположим, на очередном кадре F_i в качестве действия выбран классификатор $C_i = C^j$. Будем считать, что данный кадр распознан правильно, если $C_i(F_i) = j_i^*$. Тогда функционал качества для правила адаптивного выбора классификатора a определяется следующим образом:

$$Q(a, \mathbb{V}) = \sum_{i=1}^T L(a, F_i), \quad (10)$$

где $L(a, F_i)$ – индикатор ошибки, определяемый следующим образом:

$$L(a, F_i) = \begin{cases} 1, & a(F_i) = j_i^* \\ 0, & a(F_i) \neq j_i^*. \end{cases} \quad (11)$$

В такой постановке задача выбора «правильного» классификатора аналогична задаче выбора «лучшего» рычага n -ру�ого бандита, если заметить, что успешное детектирование объекта на очередном кадре аналогично хорошему вознаграждению в игре, а видеофрагмент – последовательной серии игр. Предлагаемый алгоритм адаптивного выбора распознающих классификаторов реализует идею алгоритма выбора лучшего действия в задаче об n -руком бандите с учетом следующих особенностей.

Во-первых, в задаче поиска объектов понятие «вознаграждение» требует отдельного определения. Интуитивно понятно, что вознаграждение за выбор действия (распознающего классификатора) должно быть положительным, если удалось правильно найти объект, и нулевым в противном случае. Однако в момент игры достоверной информации об имеющихся на кадрах объектах нет. Поэтому примем, что вознаграждение в очередной игре r_{k+1} равно 1 при нахождении объекта на кадре и 0 в противном случае. Во-вторых, будем обновлять ценность действия $V_{k+1}(a)$ используя формулу вычисления экспоненциального среднего:

$$V_{k+1}(a) = \alpha r_{k+1} + (1 - \alpha)V_k(a), \quad (12)$$

где $\alpha \in [0; 1]$ – размер шага (чем больше значение α , тем больший вес в ценности действия имеет новое вознаграждение).

Суть алгоритма выбора распознающего классификатора заключается в выборе в каждый момент времени лучшего (с точки зрения ценности) классификатора, применении его для поиска объекта на текущем кадре и обновлении значения ценности в зависимости от результата детектирования.

Эффективность описанного алгоритма оценивалась в рамках решения задачи определения типа банковской карточки в видеопотоке. Оператор фотографирует с помощью web-камеры банковскую карту, а система распознавания должна определить тип карты по имеющемуся логотипу. Всего тестовый видеофрагмент содержал 29 представлений банковских карт следующих пяти типов: VISA, MasterCard, American Express, Discover и UnionPay. Общее количество кадров видеофрагмента составило 1116, из которых 529 кадров содержали изображение карты. Детектирование карт выполнялось путем использования классификаторов логотипов соответствующих платежных систем.

Качество работы представленного алгоритма выбора распознающего классификатора (далее – АВРК) со значением параметра $\alpha = 0.9$ было сравнено с качеством работы следующих трех методов: поочередным методом (обученные классификаторы применяются к кадрам по очереди), ϵ -жадным методом при $\epsilon = 0.1$ и softmax-методом при $\tau = 0.1$. Рост функционала качества Q для каждого исследуемого метода выбора классификатора представлен на рисунке 5. Помимо значения Q большой интерес представляет доля верно распознанных изображений, содержащих банковскую карточку (true positive rate, TPR). Как видно из таблицы 4 разработанный алгоритм АВРК не только опережает все исследуемые методы по значению функционала качества, но и намного точнее распознает изображения, непосредственно содержащие банковскую карточку.

Таблица 4 — Качество работы исследуемых методов выбора классификатора

Метод	Q	TPR
поочередный	676	0,18
ϵ -жадный ($\epsilon = 0.1$)	773	0,52
softmax ($\tau = 0.1$)	724	0,44
АВРК ($\alpha = 0.9$)	852	0,80

Пятый раздел посвящен проектированию и реализации программного комплекса детекции объектов методом Виолы и Джонса, содержащего имплементацию полученных в рамках диссертационной работы научных результатов, обеспечивающего полный цикл обучения классификаторов, обладающего гибкостью настройки, легкостью добавления новых структурных

элементов, кроссплатформенностью и легкостью подключения к индустриальным распознающим системам.

В результате был разработан программный комплекс *objed*, состоящий из двух модулей: *модуля обучения* (обеспечивает полный цикл обучения и тестирования классификаторов) и модуля воспроизведения (обеспечивает применение обученного классификатора к изображению). Разработанный программный комплекс содержит имплементацию всех изложенных в диссертационной работе результатов, а также классические основы оригинального алгоритма Виолы и Джонса.

В настоящий момент программный комплекс *objed* используется в следующих индустриальных распознающих системах:

1. Автоматический классификатор транспортных средств «АКТС-4»: обучены детекторы колес в ракурсе «вид сбоку» для использования в подсистеме определения колесных пар проезжающего транспортного средства. Система АКТС-4 не имеет мировых аналогов, работает на данный момент на 200 полосах российских дорог с ежедневной пропускной способностью около 500 тысяч автомобилей.

2. Системы распознавания документов, удостоверяющих личность, Smart PassportReader и Smart IDReader: обучены детекторы паспорта гражданина РФ, водительского удостоверения гражданина РФ, СНИЛС, миграционной карты в рамках подсистемы локализации, ориентации и идентификации документов на сканах и отдельных кадрах видеопоследовательности. Системы распознавания

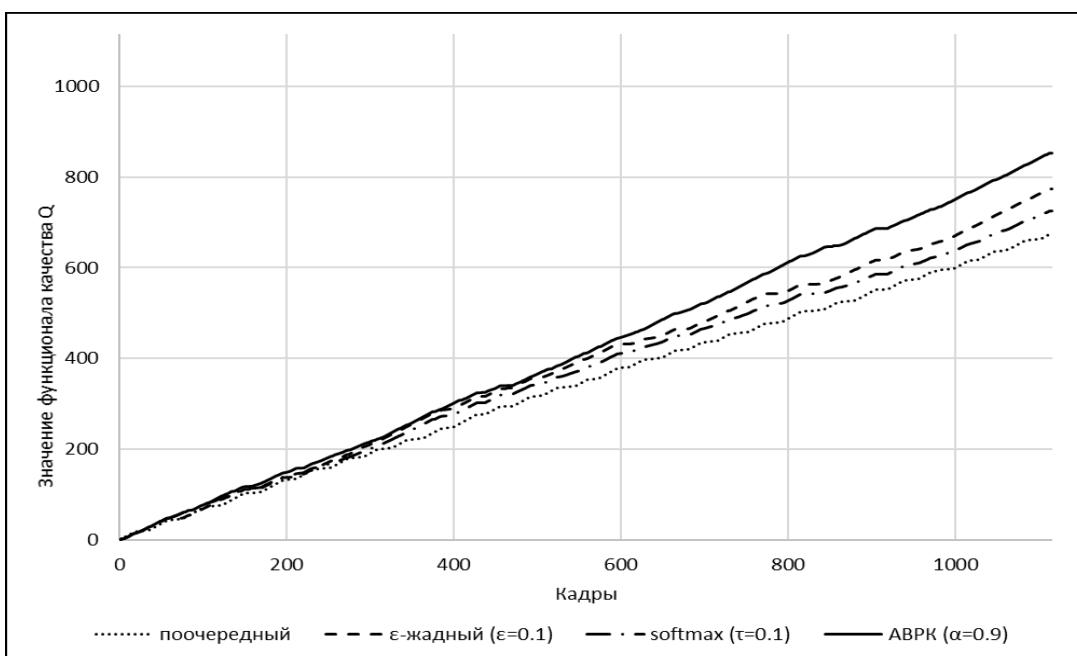


Рисунок 5 — График зависимости значения функционала качества Q от номера кадра

Smart PassportReader и Smart IDReader интегрированы в информационные системы крупных российских банков (Почта Банк, Тинькофф Кредитные Системы, БинБанк), страховых компаний (АльфаСтрахование, Ренессанс Страхование, УралСиб Страхование, Сбербанк Страхование), авиакомпаний (Аэрофлот), за счет чего общее количество пользователей систем распознавания Smart PassportReader и Smart IDReader превышает 10 миллионов человек.

3. Интеллектуальная система автономного вождения Cognitive C-Pilot: обучены детекторы попутных и двигающихся навстречу транспортных средств, а также детекторы основных дорожных знаков;

Программный комплекс вместе с исходным кодом размещен в репозитории GitHub и доступен для ознакомления в свободном режиме по адресу <https://github.com/usilinsergey/objed>.

В **заключении** приведены основные результаты диссертации.

ОСНОВНЫЕ РЕЗУЛЬТАТЫ ДИССЕРТАЦИИ

1. Представлено семейство признаков для алгоритма Виолы и Джонса, устойчивых к различным параметрам освещенности и учитывающих геометрические особенности объектов. Вычислительная сложность представленных признаков сопоставима со сложностью яркостных признаков Хаара. Эффективность контурных признаков продемонстрирована на примере распознавания образов колес в рамках задачи распознавания типа транспортного средства, а также на примере распознавания номера банковской карты в рамках задачи распознавания документов. В обоих случаях качество детектирования с помощью контурных признаков оказалось выше более, чем на 15% по сравнению с яркостными признаками.

2. Разработан алгоритм обучения древовидного классификатора Виолы и Джонса, который автоматически выполняет кластеризацию данных в процессе обучения, а также обеспечивает быстрое дообучение классификаторов при расширении обучающих наборов. Алгоритм был испытан для решения задачи обучения детектора образов автомобильных колес и детектора логотипа платежной системы VISA. Построенный древовидный классификатор в обоих задачах показывал лучшее качество детектирования по сравнению с классическим линейным каскадом.

3. Представлен алгоритм адаптивного выбора распознающего классификатора в задаче многоклассовой детекции объектов на видеопоследовательности, отвечающей модели поочередной демонстрации целевых объектов. В основе алгоритма лежит стратегия жадного выбора действия в задаче n-руского бандита. Эффективность описанного алгоритма была

продемонстрирована на примере поиска и распознавания логотипов банковских карт в видеопотоке путем сравнения с классическими для задачи n-руского бандита стратегиями: ε-жадным методом и softmax-методом.

4. Разработан программный комплекс детекции объектов objed, содержащий программную имплементацию результатов диссертационной работы. Программный комплекс архитектурно состоит из двух частей: модуль обучения классификаторов и модуль воспроизведения (применения) классификаторов и позволяет выполнять полный цикл обучения классификаторов. За счет абстрагирования сущностей и активного применения различных паттернов программирования программный комплекс objed обеспечивает возможность легкого алгоритмического расширения (добавления новой функциональности). Программный комплекс вместе с исходным кодом размещен в репозитории GitHub и доступен для ознакомления в свободном режиме.

5. Практическая ценность полученных результатов подтверждается их использованием при решении ряда задач в индустриальных системах распознавания. Результаты диссертационной работы внедрены в систему классификации автомобилей по геометрическим характеристикам АКТС-4 (эксплуатируется на 200 полосах пунктов взимания платы автомагистралей России М1 «Беларусь», М4 «Дон» и Западного Скоростного Диаметра Санкт-Петербурга), в программные продукты распознавания документов Smart IDReader и Smart PassportReader (интегрированы в информационные решения крупнейших российских банков, страховых компаний и организаций различных государственных структур), а также в интеллектуальную систему автономного вождения C-Pilot (система содействия водителю, предназначена для установки на коммерческий транспорт).

СПИСОК РАБОТ, ОПУБЛИКОВАННЫХ АВТОРОМ ПО ТЕМЕ ДИССЕРТАЦИИ

Публикации в изданиях, входящих в Перечень ВАК РФ

1. Усилин, С.А. Использование метода преследования для повышения быстродействия алгоритма многоклассовой детекции объектов в видеопотоке каскадами Виолы-Джонса / С.А. Усилин // Труды Института системного анализа Российской академии наук (ИСА РАН). — М.: Поли Принт Сервис, 2017. — Т. 67. — №1. — С. 75-82.

2. Поляков, И.В. Построение оптимальных каскадов Виолы–Джонса при помощи «жадных» алгоритмов перебора управляемых параметров с промежуточным контролем по валидационной выборке / И.В. Поляков, Е.Г.

Кузнецов, С.А. Усилин, Д.П. Николаев // Сенсорные системы. — М.: Наука, 2016. — Т. 30. — №3. — С. 241-248.

3. Усилин, С.А. Распознавание гильоширных элементов: определение страниц паспорта РФ / С.А. Усилин, Д.П. Николаев, Д.Л. Шоломов, В.В. Арлазаров // Труды Института системного анализа РАН. Обработка информационных и графических ресурсов. — М.: URSS, 2013. — Т. 63. — № 3. — С. 106-110.

4. Усилин, С.А. Локализация, ориентация и идентификация документов с фиксированной геометрией на изображении / С.А. Усилин, Д.П. Николаев, В.В. Постников // Труды Института системного анализа РАН. Обработка информационных и графических ресурсов. — М.: URSS, 2010. — Т. 58. — С. 248-261.

5. Гладилин, С.А. Построение устойчивых признаков детекции и классификации объектов, не обладающих характерными яркостными контрастами / С.А. Усилин, Д.П. Николаев, А.А. Котов, С.А. Гладилин // Информационные технологии и вычислительные системы. — М.: ФИЦ ИУ РАН, 2014. — Т. 1. — С. 61-72.

Публикации в трудах профильных конференций

6. Fedorenko, F. Real-time object-to-features vectorisation via Siamese neural networks / F. Fedorenko, S. Usilin // Proceedings of the Ninth International Conference on Machine Vision (ICMV 2016). — SPIE, 2017. — Vol. 10341. — 103411R. (**SCOPUS**)

7. Minkina, A. Generalization of the Viola-Jones method as a Decision Tree of Strong Classifiers for Real-time Object Recognition in Video Stream / A. Minkina, D. Nikolaev, S. Usilin, V. Kozyrev // Proceedings of the Seventh International Conference on Machine Vision (ICMV 2014). — SPIE, 2015. — Vol. 9445. — 944517. (**SCOPUS**)

8. Usilin, S. Visual Appearance Based Document Image Classification / S. Usilin, D. Nikolaev, V. Postnikov, G. Schaefer // Proceedings of the 17th IEEE International Conference on Image Processing (ICIP 2010). — IEEE, 2010. — Р. 2133-2136. (**SCOPUS, WoS**)

9. Кузнецова, Е.Г. Модификация слабых классификаторов машины Виолы-Джонса для мультиспектральных изображений / Е.Г. Кузнецова, С.А. Усилин, А. Г. Минкина, Д.П. Николаев // Информационные технологии и системы (ИТиС'15): сборник трудов конференции. — М.: ИППИ, 2015. — С. 329-337.

10. Минкина, А.Г. Обобщение метода Виолы и Джонса в виде решающего дерева сильных классификаторов для распознавания объектов в видеопотоке в режиме реального времени / А.Г. Минкина, А.С. Григорьев, С.А. Усилин, Д.В. Полевой, Д.П. Николаев // Информационные технологии и системы (ИТиС'14): сборник трудов конференции. — М.: ИППИ, 2014. — С. 158-163.

11. Григорьев, А.С. Ускорение поиска объектов в видеопотоке методом Виолы–Джонса путем адаптивного выбора распознающих каскадов / А.С. Григорьев, С.А. Усилин, Н.Д. Петрович // Труды 55-й научной конференции МФТИ. — Москва-Долгопрудный-Жуковский : МФТИ, 2012. — С. 11-12.
12. Жуковский, А.Е. Синтез обучающей выборки на основе реальных данных в задачах распознавания изображений / А.Е. Жуковский, Н.А. Тарасова, С.А. Усилин, Д.П. Nicolaev // Информационные технологии и системы (ИТиС'12): сборник трудов конференции. — М.: ИППИ, 2012. — С. 377-382.
13. Котов, А.А. Построение устойчивых признаков для алгоритма Виолы и Джонса в задаче классификации транспортных средств / А.А. Котов, С.А. Усилин, Д.П. Nicolaev // Информационные технологии и системы (ИТиС'12): сборник трудов конференции. — М.: ИППИ, 2012. — С. 383-388.
14. Usilin, S. Guilloche Elements Recognition Applied to Passport Page Processing / S. Usilin, D. Nikolaev, D. Sholomov // Proceedings of the 8th Open German-Russian Workshop «Pattern Recognition and Image Understanding» OGRW-8-2011. — Nizhny Novgorod : Lobachevsky State University of Nizhny Novgorod, 2011. — Р. 303-306.
15. Жуковский, А.Е. Применение активного обучения к методу Виола-Джонса поиска объектов на изображении / А.Е. Жуковский, Д.П. Nicolaev, С.А. Усилин // Информационные технологии и системы (ИТиС'11): сборник трудов конференции. — М.: ИППИ, 2011. — С. 234-240.
16. Усилин, С.А. Поиск объектов в видеопотоке при известных кинематике и геометрической модели сцены / С.А. Усилин, Д.П. Nicolaev, В.В. Постников // Труды 53-й научной конференции МФТИ. — Москва-Долгопрудный-Жуковский : МФТИ, 2010. — С. 67–69.
17. Усилин, С.А. Идентификация изображений документов методами, предназначенными для поиска лиц / С.А. Усилин, Д.П. Nicolaev, В.В. Постников // Информационные технологии и системы (ИТиС'10): сборник трудов конференции. — М.: ИППИ, 2010. — С. 132–137.

Патенты, свидетельства о регистрации программ для ЭВМ

18. Автономное автоматизированное рабочее место контроля паспортных документов : Патент на полезную модель № 166152 Российская Федерация / В.В. Арлазаров, А.П. Гладков, Д.П. Nicolaev, С.А. Усилин ; патентообладатель Общество с ограниченной ответственностью «СМАРТ ЭНДЖИНС СЕРВИС». — № 2016122432; заявл. 07.06.2016 ; опубл. 26.10.2016.
19. Автоматизированное рабочее место контроля паспортных документов : Патент на полезную модель № 166038 Российская Федерация / В.В. Арлазаров, А.П. Гладков, Д.П. Nicolaев, С.А. Усилин ; патентообладатель Общество с

ограниченной ответственностью «СМАРТ ЭНДЖИНС СЕРВИС». — № 2016106183; заявл. 25.02.2016 ; опубл. 24.10.2016.

20. Программа для распознавания идентификационных карт личности «Smart IDReader» : свидетельство о государственной регистрации программы для ЭВМ № 2016616961 / В.В. Арлазаров, Д.П. Николаев, С.А. Усилин, К.Б. Булатов, Т.С. Чернов, Д.Г. Слугин, Д.А. Ильин, П.В. Безматерных, А.А. Муковозов, Е.Е. Лимонова. — № 2016612014 ; заявл. 10.03.2016 ; опубл. 22.06.2016.

21. Библиотека для распознавания в видеопотоке паспорта гражданина Российской Федерации «Smart PassportReader» : свидетельство о государственной регистрации программы для ЭВМ № 2015616071 / В.В. Арлазаров, К.Б. Булатов, Д.А. Ильин, А.В. Куроптев, Д.П. Николаев, Д.В. Полевой, С.А. Усилин, И.А. Фараджев, Т.С. Чернов. — № 2015612880 ; заявл. 10.04.2015 ; опубл. 29.05.2015.

22. Программа для распознавания и детектирования автомобилей, с учетом их ракурса и направления движения : свидетельство о государственной регистрации программы для ЭВМ № 2015610891 / Д.В. Полевой, С.А. Усилин, П.В. Безматерных. — № 2014662041 ; заявл. 26.11.2014 ; опубл. 20.01.2015.

23. Способ автоматической классификации транспортных средств : Пат. 2486597 С1 Российская Федерация / Д.П. Николаев, В.В. Постников, Т.М. Ханипов, С.А. Усилин, А.С. Григорьев ; патентообладатель Общество с ограниченной ответственностью «Техно-траффик». — № 2012105146/11 ; заявл. 09.02.2012 ; опубл. 27.06.2013, Бюл. № 18. — 15 с.