Федеральный исследовательский центр «Информатика и управление»
Российской академии наук

На правах рукописи

Арлазаров Владимир Викторович

Мобильное распознавание и его применение к системе ввода идентификационных документов

Специальность 2.3.1 —
«Системный анализ, управление и обработка информации, статистика»

Диссертация на соискание учёной степени
доктора технических наук

Москва — 2023

Оглавление

Стр.

Стр.

Стр.

Стр.

Введение

Сегодня большинство процессов взаимодействия человека с организациями начинается с удостоверения личности и ввода персональных, а нередко и биометрических, данных в различные информационные системы. Кроме того, в результате мер, направленных на борьбу с эпидемией COVID-19, множество услуг стало предоставляться удаленно, при этом законами требуется проведение удаленной идентификации личности. Открытие счета в банке, получение кредита, перевод крупной денежной суммы, покупка страховых услуг, услуг связи, ювелирных украшений, получение посылки и многое другое требует физического или удаленного предъявления документов, удостоверяющих личность. При каждом таком действии необходимо ввести из предъявленного документа персональные данные, а также проверить, настоящий это документ или нет. Естественно, все эти процессы нуждаются в автоматизации.

Можно было бы сказать, что есть огромный опыт распознавания обычных документов, который можно использовать для этой цели. Но, хотя на первый взгляд процессы распознавания выглядят одинаково, есть ряд особенностей, которые существенно меняют задачу и не позволяют использовать подходы, разработанные для обычных документов.

Первой особенностью является сам документ, удостоверяющий личность. Он создан таким образом, чтобы максимально затруднить его фальсификацию и подделку. Для этого используются разные элементы защиты: сложные разноцветные фоны (в том числе и гильоширные), голографические элементы защиты. Кроме того, документы, удостоверяющие личность, часто ламинируют специальной пленкой или даже делают их из пластика. Сама информация наносится различными способами: эмбоссингом, гравировкой, термопечатью, спеканием. Зачастую используются уникальные секретные шрифты. Все эти особенности объекта распознавания делают многие методы, использованные для классического распознавания документов, не применимыми. Особую проблему представляют оптически-изменяемые элементы (англ. Optical Variable Devices, OVD), которые расположены поверх текста и при сканировании обычными планшетными сканерами могут в буквальном смысле засветить участок изображения, содержащий текст.

Вторую особенность принесло широкое распространение мобильных устройств, таких как смартфонов или планшетов, которые стали повседневным средством доступа к всевозможным услугам. Современные смартфоны имеют одну или несколько фото/видео камер, значительный объем памяти и вычислительные ресурсы для выполнения разнообразных задач. Например, для многих банков мобильное приложение стало основным способом взаимодействия с клиентом. Логичным шагом стало использование камеры для захвата изображения документов и вычислительных мощностей для распознавания. Однако, кроме высокой скорости захвата изображения документа, фотографирование принесло и новые проблемы, связанные с качеством камер и неконтролируемыми условиями съемки, не характерные для традиционного сканера: шум матрицы, переменные и неизвестные условия освещения, блики, расфокусированность, смаз. Появились также совершенно новые проблемы анализа документов: восстановление координатной системы сцены и поиск документа в этой трехмерной сцене. Несмотря на то что современные смартфоны обладают значительными вычислительными мощностями, в среднем они все еще уступают персональным компьютерам и, тем более, серверным системам, что налагает определенные рамки на вычислительную сложность алгоритмов распознавания.

Помимо вышеперечисленных новых особенностей, связанных со сменой источника изображений, сложностями самого объекта, стоит также отметить, что требования к качеству распознавания документов, удостоверяющих личность, превосходят требования по качеству для обычных документов ввиду важности удостоверяющих документов для краеугольных бизнес-процессов.

Таким образом, важная с практической точки зрения задача является **актуальной** и требует для своего решения новых научных и технических подходов.

Целью данной работы является исследование задач, связанных с построением систем распознавания документов, удостоверяющих личность, и создание архитектуры и алгоритмов, необходимых для построения кардинально новых систем этого типа.

Для реализации этой цели необходимо решить следующие **задачи**:

1. Предложить метод выделения документа на сложном фоне, не требующий бинаризации и учитывающий проективные искажения, возникающие при обработке фото и видео снимков.

2. Предложить метод выделения текстовых зон на документе, не требующий распознавания текстов и учитывающий возможные перепады яркостей и нарушения «прямолинейности» строк текста.
3. Разработать метод распознавания строки, не требующий базовых линий, учитывающий возможные блики и другие неконтролируемые условия освещения.
4. Предложить метод проверки подлинности удостоверяющих документов, включающий как проверки геометрического характера (подписи, печати и т. д.), так и логического характера.
5. Исследовать методы использования видеопотока для улучшения качества распознавания и предложить алгоритмы комбинирования результатов и ограничения числа рассматриваемых кадров.
6. Предложить методы построения коллекций «искусственных» удостоверяющих документов с нужными характеристиками для обучения и контроля программ распознавания, учитывая трудности получения реальных документов.
7. На базе предложенных методов разработать универсальный программный комплекс распознавания документов, удостоверяющих личность, который в каждом конкретном случае работал бы на уровне лучших мировых образцов.

Научная новизна диссертационного исследования заключается в следующих аспектах:

1. В диссертации впервые предложена универсальная архитектура системы распознавания документов, удостоверяющих личность, опирающаяся на современные методы обработки изображений. На основе методов математической морфологии, Виолы-Джонса и RANSAC разработаны алгоритмы, позволяющие эффективно выделять изображение документа на сложном фоне, осуществлять поиск образца и отображение документа на шаблон, выделять и распознавать отдельные поля документа, а также осуществлять проверки подлинности.
2. Впервые проведен широкий анализ методов использования видеопотока для распознавания документов, удостоверяющих личность. Построена новая вероятностная модель на основе распределения Дирихле, адекватно описывающая оценки результатов распознавания на кадрах видеопоследовательности и согласующаяся с эмпирическими

данными по критерию Андерсона-Дарлинга, пригодная к использованию как для получения комбинированного результата распознавания объектов в видеопоследовательности, так и для принятия решения об остановке процесса распознавания.

3. Впервые построены базы видеопотока изображений документов, удостоверяющих личность, в объемах, позволяющих проводить обучение алгоритмов и контроль программных комплексов. Для этого разработана оригинальная система аугментации, технология получения бумажных псевдодокументов и предъявления их в видеопотоке.
4. Разработан уникальный программный комплекс распознавания идентификационных документов. На базе него создан ряд прикладных систем, внедренных и внедряющихся в государственные организациях, банках, у операторов связи, в аэропортах, на железных дорогах и т. п., как в России, так и за рубежом.

Основные положения, выносимые на защиту:

1. При работе с идентификационными документами с помощью мобильных устройств традиционные методы распознавания документов, основанные на бинаризации изображения, поиске прямоугольников, идентификации полей текстами и т. п., оказываются недостаточными или даже неприменимыми. Необходимо опираться на современные методы обработки 2D и 3D изображений, особенности видеопотока и специфику самих документов, удостоверяющих личность.
2. Для идентификации документа на изображении и полей на документе можно использовать алгоритмы, основанные на выделении особых точек, но необходимы специальные механизмы для ускорения поиска, разработанные в диссертации.
3. Использование видеопотока является серьезным ресурсом для повышения качества распознавания документов, удостоверяющих личность. При этом механизмы, объединяющие выбор наилучшего кадра, комбинирование результатов распознавания и метод прогнозирования оценок для определения точки останова, разработанный в диссертации, в сумме обеспечивают близкие к оптимальным решения по критериям максимизации функции, включающей оценки скорости и качества распознавания.

4. Разработанные в диссертации методы построения коллекций видеоизображений и видеопотока, основанные на аугментации и специальных методах обработки, дают возможность построить базы данных для обучения системы распознавания и контроля результатов, даже имея единичные изображения заданного типа документов.
5. Ряд методов, разработанных для ускорения работы программ распознавания, как связанных с приближениями в нейронных сетях, так и с использованием особенностей архитектуры компьютеров, позволяют добиться на мобильных устройствах высокой скорости работы, сопоставимой или превышающей скорости захвата изображения. В том числе, удается добиться приемлемых результатов на отечественных платформах.
6. Важнейшей частью системы распознавания документов, удостоверяющих личность, является система проверки подлинности. Разработанные в диссертации методы (наличия обязательных объектов, таких как подписи и печати, контроля фактуры печати и шрифта и др.) позволяют охватить большой круг возможных фальсификаций и подделок.

Методология и методы исследования. Основой проводимых в диссертации исследований является методология системного анализа. В теоретической части используются методы теории вероятностей и математической статистики. В алгоритмической части работы используются современные методы Computer Science и, в частности, методы цифровой обработки изображений.

Публикации. Основные результаты по теме диссертации изложены в 40 печатных работах, 20 из которых изданы в журналах, рекомендованных ВАК, 30 работ индексируется Web of Science и Scopus (включая 10 работ, опубликованных в журналах Q1 и Q2).

Личный вклад. Все основные результаты диссертации получены автором самостоятельно. В большинстве совместных публикаций по теме диссертации автору принадлежат постановки задач и принципиальный подход к их решению. В то же время детальная проработка алгоритмов и их реализация чаще всего принадлежит соавторам. В работах [1-2; 35; 39] соискателем предложен общий подход к построению системы распознавания документов, удостоверяющих личность, на мобильных устройствах, проработана архитектура и определено место алгоритмов проверки подлинности документов. В работе [10] соискателем лично исследован вопрос использования проблемно-ориентиро-

ванных пакетов данных в научных работах по распознаванию документов, удостоверяющих личность, поставлена задача создания новых пакетов данных. В работах [3; 5; 20; 24; 30] соискателем лично разработана методика создания датасетов, подобран инструментарий и определены способы соблюдения законодательства. В работе [8] соискателем рассмотрены проблемы распознающих систем, обрабатывающих разнородные входные данные, а также сформированы подходы к построению системы распознавания. В работе [9] соискателем предложены основные этапы обработки шаблона документа ID-карт. В работе [11] соискателем предложены методы комбинирования множественных результатов распознавания текста при распознавании видеопотока. В работах [6; 21; 23; 28; 36] соискателем предложено использование семейства каскадных классификаторов для решения задачи локализации штампов и печатей на документах для определения их подлинности, а также предложен подход по аугментации обучающих данных для эффективного обучения таких классификаторов. В работах [13; 14; 19; 26; 29; 31; 32; 34] соискателем рассмотрена проблема останова распознавания документа, а также предложены и проанализированы методы останова процесса распознавания на основе анализа популяций результата распознавания, рассмотрена проблема комбинирования результатов распознавания видеопотока и предложена серия подходов для решения сформулированной задачи. В работах [33; 38-39] соискателем лично предложена модель распознавания документов с оценкой качества в видеопотоке.

Апробация работы. Основные результаты работы докладывались и обсуждались на следующих профильных международных конференциях (четыре из которых относятся к конференциям ранга A1 и A2 по системе ранжирования Qualis):

- 25th International Conference on Pattern Recognition (ICPR 2020), Милан, Италия – конференция **ранга A1**;
- 16th International Conference on Document Analysis and Recognition (ICDAR 2021), Лозанна, Швейцария – конференция **ранга A2**;
- 15th International Conference on Document Analysis and Recognition (ICDAR 2019), Сидней, Австралия – конференция **ранга A2**;
- 14th International Conference on Document Analysis and Recognition (ICDAR 2017), Киото, Япония – конференция **ранга A2**;
- 15th International Conference on Machine Vision (ICMV 2022), Рим, Италия;

- 14th International Conference on Machine Vision (ICMV 2021), виртуальная конференция;
- 13th International Conference on Machine Vision (ICMV 2020), виртуальная конференция;
- 12th International Conference on Machine Vision (ICMV 2019), Амстердам, Нидерланды;
- 11th International Conference on Machine Vision (ICMV 2018), Мюнхен, Германия;
- 10th International Conference on Machine Vision (ICMV 2017), Вена, Австрия;
- 9th International Conference on Machine Vision (ICMV 2016), Ницца, Франция.

В 2020 году результаты диссертационной работы были представлены в пленарном докладе «Recognition of documents with fixed layout on a mobile device: coarse-to-fine Approach» на VI Международной конференции и молодежной школы «Информационные технологии и нанотехнологии» (ИТНТ-2020).

Грантовая поддержка. Ряд исследований по теме диссертационной работы поддержаны Российским фондом развития информационных технологий и Российским фондом фундаментальных исследований, в которых соискатель выступал непосредственно в роли руководителя проекта. Соответственно, часть результатов диссертации была получена в процессе выполнения работ по следующим грантам:

- Соглашение № 2021-550-80 от 29.06.2022, проект «Доработка программы для распознавания идентификационных карт личности «Smart ID Engine»: разработка модуля выявления признаков фальсификации документов и атак на предъявление документов в оптическом диапазоне на основе технологий искусственного интеллекта»;
- Проект РФФИ 17-29-03170 офи_м «Исследование быстродействующих методов и алгоритмов обработки изображений и оптического распознавания для использования в мобильных устройствах с ограниченной вычислительной производительностью»;
- Проект РФФИ 18-07-01384 «Исследование применимости методов нелинейных аппроксимаций для оптимизации быстродействия искусственных нейронных сетей на современных микропроцессорных архитектурах»;

– Проект РФФИ 15-07-06520 «Методы контроля подлинности документов и их фрагментов в гибридных системах обработки, передачи и хранения документов».

Теоретическая значимость работы состоит в том, что предложена новая постановка задачи и предложены новые подходы к построению систем распознавания документов, а также в возможности развития методов, разрабатываемых в диссертации, как в рамках систем рассматриваемого в работе класса, так и в рамках других активно развивающихся направлений распознавания документов.

Практическая значимость диссертационной работы состоит в применимости предлагаемых в диссертации концепций и методов к построению практических систем распознавания идентификационных документов. Она подтверждается созданием программного инструментария на базе решений, предложенных в работе, который применен в большом количестве прикладных систем, работающих в сотнях организаций, использующих распознавание паспортов РФ, пластиковых карт, машиночитаемых зон, водительских удостоверений и других документов. Акты о внедрении результатов диссертации приведены в Приложении А. Практическая значимость работы подтверждается также 2 патентами США, 5 патентами на изобретение РФ, 16 патентами на полезную модель и 4 свидетельствами о государственной регистрации программы для ЭВМ, в которые входят результаты, содержащиеся в диссертации.

Достоверность полученных результатов подтверждается многочисленными публикациями, многие из которых имеют высокий уровень цитируемости. Результаты также докладывались на ведущих по данной тематике международных конференциях. Программные комплексы, созданные на основе результатов, описанных в диссертации, успешно работают в большом числе организаций.

Результаты диссертационного исследования **соответствуют паспорту специальности** 2.3.1 «Системный анализ, управление и обработка информации, статистика», а именно пункту 1 «Теоретические основы и методы системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», пункту 2 «Формализация и постановка задач системного анализа, оптимизации, управления, принятия решений и обработки информации», пункту 4 «Разработка методов и алгоритмов решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта», пункту 5 «Разработка

специального математического и алгоритмического обеспечения для решения задач системного анализа, оптимизации, управления, принятия решений, обработки информации и искусственного интеллекта».

Объем и структура работы. Диссертация состоит из введения, 5 глав и заключения. Полный объём диссертации составляет 358 страниц, включая 104 рисунка и 27 таблиц. Список литературы содержит 350 наименований.

Глава 1. Автоматический анализ и распознавание изображений документов методами компьютерного зрения

1.1 Введение

Организация и ведение учета посредством документооборота сопровождает человеческую деятельность с древнейших времен. В начале 1980-х, следуя за развитием электронной вычислительной техники, начался процесс переноса процессов управления документами на бумажных носителях в электронную форму. Несмотря на повсеместные и постоянные прогнозы о том, что электронная форма неизбежно и очень скоро заменит бумажную, весь мир, спустя более сорока лет, продолжает использовать документы на бумажных носителях, и системы "смешанного" документооборота, в рамках которых полностью отказаться от традиционных бумажных документов все еще нельзя. В такой ситуации представляется важным обладать технологиями, которые бы позволяли легко и точно преобразовывать документы из одной формы в другую. Важнейшим спектром задач, решение которых необходимо для построения и функционирования таких технологий, являются задачи автоматического анализа и распознавания изображений документов (в англоязычной литературе - Document Image Analysis, DIA).

Для обсуждения задач анализа и распознавания изображений документов необходимо ввести несколько основных понятий. Мы будем понимать под "документом" совокупность неизменяемых (для зафиксированного класса документов) элементов и информационных атрибутов. Значения атрибутов интерпретируются информационной системой с целью проведения операций над документом. Примерами таких операций являются регистрация, аннулирование, контроль, синхронизация атрибутов документа и данных в электронном архиве и т. п., которые относятся к классу "деловых" документов. В отличие от произвольных документов атрибуты деловых документов являются параметрами для процессов документооборота. Один и тот же документ может считаться и произвольным, и деловым. Например, из изображения денежной банкноты, снабженной уникальным номером и обладающей признаками подлинности, можно извлечь признаки для проверки подлинности или происхождения

банкноты. В настоящее время такая регулярная обработка банкноты является исключительной. Примеры документов различных типовltdl представлены на рис. 1.1.

Рисунок 1.1 — Примеры идентификационных документов (сверху) и гибких форм деловых документов (снизу).

Документы содержат статические элементы и элементы заполнения: поля (атрибуты), подтверждающие элементы. Статическими элементами, прежде всего, являются слова статического текста. Статические слова группируются в строки, в заголовки, в абзацы и в параграфы. Другими статическими элементами являются разделяющие линии, бар-коды и QR-коды, рамки чек-боксов. Сложным статическим элементом является таблица. Поля могут определяться как тексты, ограниченные статическими элементами или разделяющими линиями и линиями таблиц. Возможны многострочные поля, части которых могут переноситься на другую страницу многостраничного документа. К подтверждающим элементам относятся подписи, печати и рукописные пометки. Дизайн

документов может содержать секции, объединяющие логически связанные группы элементов и полей. Существуют классы документов (к примеру, документы, удостоверяющие личность, либо другие документы, выполненные на специальных бланках), содержащие сложный фон, также относящийся к совокупности статических элементов документа, либо произвольный сложный фон (к примеру, фон пластиковой банковской карты), который, строго говоря, не является статическим элементом, но и, как правило, не несет информационной нагрузки.

Под распознаванием документа мы понимаем извлечение атрибутов из образа документа. При этом часто о распознаваемых документах заранее известна некоторая информация о структуре документа и характеристиках атрибутов. Среди разнообразия всевозможных документов можно рассмотреть следующие классы (группируя по структуре):

- "жесткие" формы, создаваемые единообразно полиграфическим способом, например, идентификационные документы (паспорт, удостоверение личности), водительские удостоверения, банкноты;
- "гибкие" формы, создаваемые по известным шаблонам, например, стандартные анкеты, уведомления, декларации, пластиковые банковские карты;
- документы без строгого шаблона, например, договоры, доверенности, формальные письма;
- документы, созданные без шаблона, например, деловые письма.

Документы могут быть одностраничными или многостраничными. Для жестких форм оформление и статические тексты каждой из страниц многостраничного документа не меняются. Каждую страницу многостраничной жесткой формы можно рассматривать как одностраничный документ. В других классах документов возможен перенос статического текста и заполнения текста с одной страницы на другую.

Возможна также классификация типов документов по их применению, например, рассматриваются идентификационные документы, регистрационные документы, финансовые и кредитные документы, договорные документы.

Как будет показано далее в этой главе, в области анализа и распознавания документов все еще существуют задачи, которые не решены на достаточном уровне, чтобы обеспечить полностью автоматическое функционирование процессов преобразования документов. В частности, технологический прогресс изменил акцент в системах электронного документооборота в сторону, не пред-

сказанную заранее: поскольку наиболее распространенными и повседневно используемыми персональными вычислительными устройствами стали мобильные устройства с оптическими камерами и доступом в интернет (смартфоны, планшетные компьютеры и т.п.), увеличился спрос на технологии электронного документооборота с использованием таких устройств. Несмотря на то что документы на бумажных носителях используются почти также широко, как и раньше, еще несколько лет назад стандартные планшетные сканеры стали менее распространены и более не воспринимаются как обязательный компонент для оцифровки документов. Переход от анализа сканированных изображений документов к анализу фотографий требует разработки и развития новых подходов, устойчивых к новым условиям, таким как проективные геометрические искажения или неравномерное освещение. Таким образом, в течение нескольких последних лет, область анализа и распознавания изображений документов расширилась задачами более широкой области компьютерного зрения.

Итак, основными способами получения изображения документа является сканирование и фотографирование, в том числе с помощью мобильных устройств. Физически документы обычно являются объектами плоской природы, поэтому наиболее подходящим способом их регистрации часто оказывается сканирование. Даже в обычных сканах документов возможны искажения, связанные с углом поворота, загрязнениями сканера, осветлением и затемнением, размытием. До сих пор встречаются сильно зашумленные копии страниц. В отличие от изображений, полученных с помощью сканеров, изображения, полученные с мобильной камеры, могут иметь различные дисторсии (аберрации), содержать сильные проективные искажения. Особенно часто эти трудности возникают при оцифровке в неконтролируемых условиях съемки. При любом способе оцифровки возможен дефект потери части изображения на границе кадра.

С помощью мобильных устройств можно получить не только изображение документа, но и видеопоток в виде набора кадров, снабженных метками времени. Видеопоток по сравнению с отдельным изображением содержит существенно больший объем информации о документе.

Упомянутые возможности оформления содержания документа и особенности современных устройств оцифровки приводят к большому разнообразию изображений документов. Разнообразие документов не позволяет утверждать, что в настоящее время решены все задачи распознавания. Актуальность иссле-

дований в области распознавания документов подтверждается значительным числом работ по данной тематике, опубликованных в последнее время. Можно упомянуть следующие соревнования и профильные конференции, посвященные задачам, моделям и алгоритмам распознавания документов:

- ICDAR – International Conference on Document Analysis and Recognition, Международная конференция по анализу и распознаванию документов (https://icdar2021.org, https://icdar2020.org, ...);
- ICPR – International Conference on Pattern Recognition, Международная конференция по распознаванию образов (https://icpr2020.net, ...);
- ICMV – International Conference on Machine Vision, Международная конференция по машинному зрению (http://icmv.org);
- ASPDAC – Asia and South Pacific Design Automation Conference, Азиатская и южно-тихоокеанская конференция по проектированию и автоматизации (https://aspdac2022.github.io/index.html);
- ICIP – International Conference on Image Processing, Международная конференция по обработке изображений (http://2020.ieeeicip.org, http://2019.ieeeicip.org, ...);
- CCVPR – Conference on Computer Vision and Pattern Recognition, Конференция по компьютерному зрению и распознаванию образов (http://www.ccvpr.org).

Растущий интерес к области анализа и распознавания документов можно наблюдать по росту цитируемости публикаций профильных конференций (см. рис. 1.2).

Богатая история исследований в области автоматического анализа и распознавания документов отражена в множестве сборников докладов международных конференций, обзорах [2—19] и книгах [20—22]. Однако более старые публикации не отражают технологического сдвига последних десяти лет, в рамках которого полный процесс анализа и распознавание документов на изображениях или в видео стал возможен напрямую на автономных мобильных устройствах. Большое количество публикаций в течение последних нескольких лет уделяют внимание лишь отдельным задачам, таким как классификация изображений документов [2], извлечение определенной информации из слабоструктурированных документов [3] и т. п., или развитию конкретных методов (в основном, на основе машинного обучения [14] или глубокого обучения [18; 19]).

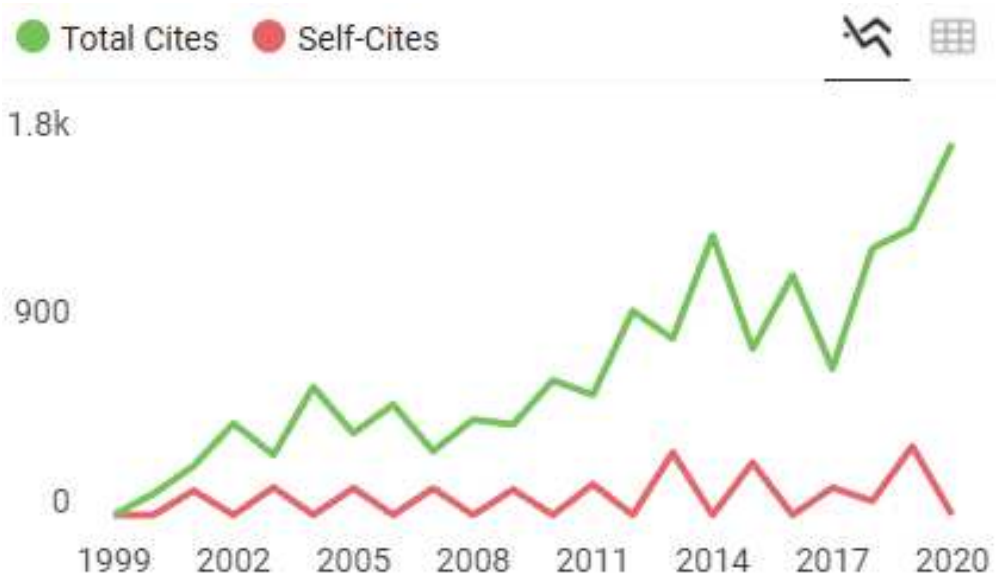


Рисунок 1.2 — Рост цитируемости публикаций международной конференции по анализу и распознаванию документов ICDAR [1].

Таким образом, в области анализа и распознавания изображений документов, некоторые давно поставленные задачи остаются нерешенными, при этом возникают новые проблемы и новые условия. Методология анализа данных и построения систем распознавания также претерпела значительное изменение за последние сорок лет, и методы на основе машинного обучения, главным образом основанные на искусственных нейронных сетях, постепенно заменяют классические алгоритмы. Данная глава будет посвящена обзору основных задач, связанных с анализом и распознаванием изображений документов и методов их решения.

1.2 Методы предварительной обработки изображения документа

В современной литературе, как в контексте систем обработки документов, так и в других областях применения компьютерного зрения, достаточно большое внимание уделяется методам и подходам к предварительной обработке изображения. Основной целью такой предварительной обработки является сужение области последующей обработки до области изображения, в котором находится целевой объект, включая предварительную геометрическую нормализацию этой области, разделение изображение на «объект» и «фон» и очистка «объекта» (фильтрация шума, минимизация искажений и т.п.). Концептуаль-

ная схема такого процесса, в достаточно общем виде, описана в работе [23] и представлена на рис. 1.3. Авторы определяют процесс «очистки» изображения как разделение изобрежения на «объект» и «фон», подавление шума на фоне и улучшение качества изображения на объекте.

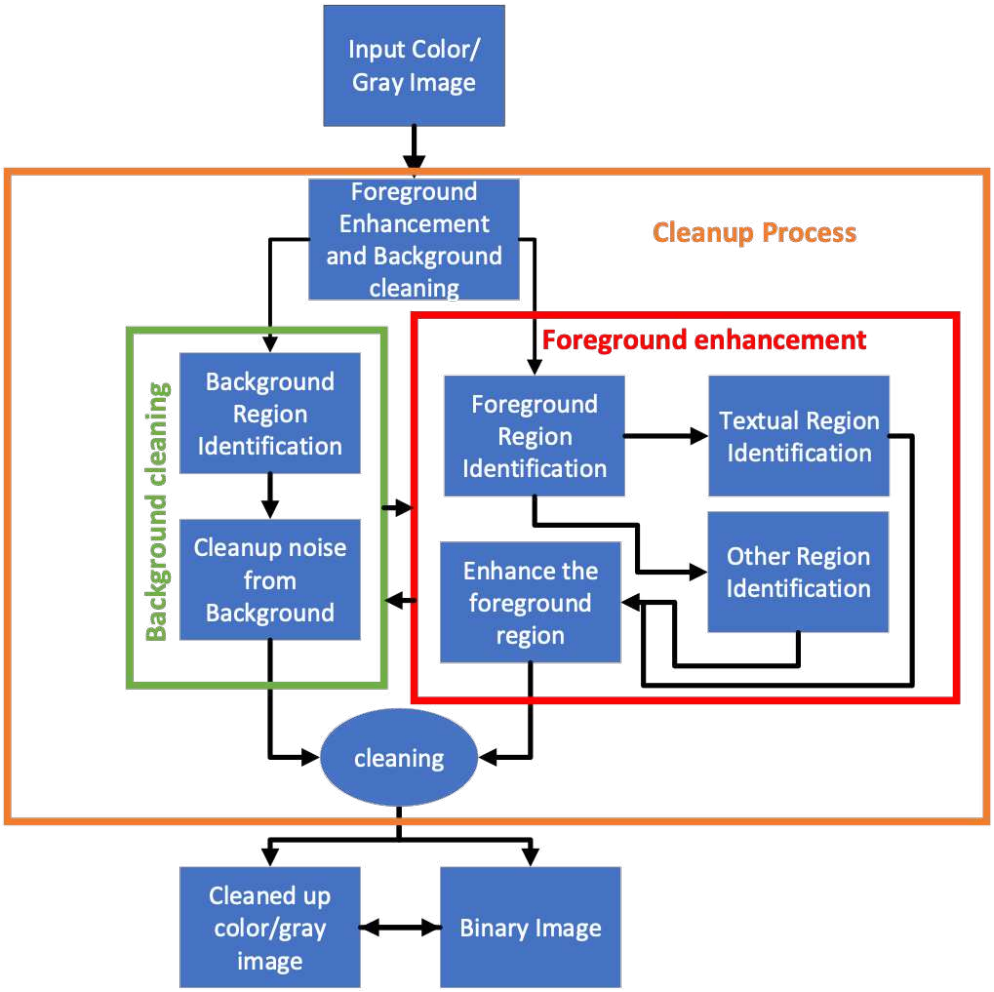


Рисунок 1.3 — Общая схема предварительной обработки изображения (на примере задачи анализа документа) [23].

Среди основных задач, которые возникают в контексте предварительной обработки и улучшения изображения, следует выделить и подробнее рассмотреть следующие:

– выравнивание и нормализация изображения целевого документа на изображении;
– цветокоррекция и улучшение качества изображения;
– бинаризация изображения;
– техника «супер-разрешения».

1.2.1 Нормализация изображений документов

Из-за человеческого и технического факторов при оцифровке изображения могут содержать искажения. Устранение искажений, или же нормализация, является классической задачей в области обработки изображений документов. Частным случаем таких искажений является наклон (поворот) зарегистрированного образа документа. Коррекция такого наклона считается одним из главных этапов подготовки документа для его дальнейшего анализа [24].

На данный момент разработано множество подходов к решению задачи определения угла наклона документа для его последующей коррекции. В фундаментальной работе 1996 года [25] было рассмотрено несколько десятков различных алгоритмов. С более современным обзором актуальных методов можно ознакомиться, например, в работах [4].

Первая группа методов базируется на анализе горизонтальных параллельных проекций изображения документа. Для изображения I размером $M \times N$ параллельная горизонтальная проекция определяется следующим соотношением:

$$\pi_I(j) = \sum_{i=0}^{M} I(i,j). \tag{1.1}$$

Задача определения наклона при этом сводится к последовательному повороту исходного изображения на предопределенное множество углов и выбору лучшей проекции. Очевидно, что для возможности ранжирования полученных проекций должен быть задан некоторый критерий $f(\pi)$. Угол φ, на котором достигается глобальный экстремум $f(\pi)$, выбирается в качестве результата. При этом в качестве критерия зачастую используется сумма квадратов значений элементов проекции, т.е.

$$f(\pi) = \sum_{j=0}^{N} \pi^2(j). \tag{1.2}$$

Вторая группа методов опирается на вычисление преобразования Хафа от изображения документа. Суть данного преобразования заключается в подсчете числа пикселей вдоль всевозможных прямых, проходящих через исходное изображение. Для задания прямых обычно используется нормальная (от слова «нормаль») параметризация (ρ, φ). При таком выборе задача сводится к

поиску в аккумуляторном пространстве строки, соответствующей прямым направления φ, на которой максимизируется некоторый заданный критерий. В качестве такого критерия обычно предлагается использовать сумму квадратов производной значений Хаф-образа [24]. Основной проблемой данной группы методов является высокая вычислительная сложность. Для преодоления этого недостатка в работе [26] было предложено для вычисления Хаф-образа использовать алгоритм быстрого дискретного преобразования Радона, что позволило снизить вычислительной сложность с $\Theta(n^3)$ до $\Theta(n^2 \log n)$.

Третья группа методов базируется на анализе компонент связности. Их идея состоит в том, чтобы провести кластеризацию выделенных компонент и выявить линейные кластеры, соответствующие строкам документа. После определения наклона в каждом из найденных кластеров выбирается единственный "усредненный" наклон.

Четвертая группа методов базируется на результате работы алгоритмов сегментации изображения документа. Вначале на нем выделяются различные примитивы, такие как строки текста, элементы разграфки, отдельные символы. После чего производится оценка угла наклона каждой из полученных компонент и вычисляется итоговый угол наклона.

К пятой группе можно отнести методы, в основе которых лежит использование искусственных нейронных сетей. Здесь интересно отметить подход, предложенный в работе [27]. В ней задача определения угла наклона рассматривается как задача классификации с 360 классами, каждый из которых соответствует определенному углу.

При наличии столь большого разнообразия методов естественным образом возникает вопрос о том, какой их них стоит выбрать. С этой целью в 2013 году был организован специальный конкурс, полностью посвященный проблеме детекции угла наклона документа [28]. В его рамках был подготовлен массив данных, моделирующий основные проблемы, встречающиеся при определении угла наклона документа. Для всех изображений вручную экспертами был указан ожидаемый ответ детектора угла. Были предложены критерии для оценки методов. По результатам конкурса лучше всего себя продемонстрировал метод, базирующийся на использовании преобразования Фурье с предварительной предобработкой изображения специального вида [29].

Определение угла наклона для последующей нормализации требуется не только для коррекции всего документа. При обработке изображений докумен-

тов требуется проводить распознавание их образов слов текста с помощью модулей OCR. Нормализация образов слов текста улучшает точность работы OCR.

Необходимо отметить, что задачи нормализации для печатных и рукописных атрибутов существенно различаются. В первом случае речь идет о поиске единственного доминирующего направления на всем фрагменте. Наклон образов слов рукописного текста может варьироваться в рамках фрагмента слова, что приводит к гораздо более сложным схемам обработки фрагментов. Примеры таких схем, опирающихся на принципы динамического программирования, приведены в работах [30].

1.2.2 Цветокоррекция и улучшение качества изображения

Под цветокоррекцией изображения понимают замену или изменение определенных параметров цветовой палитры изображения – тонов, их насыщенности, оттенков, с целью увеличения контрастности, осветления, убирания эффекта «дымки», и в целом для улучшения изображения для восприятия человеком или автоматической системой анализа.

Существует множество подходов, направленных на коррекцию цветовой палитры изображений. Существующие подходы можно условно разделить на две основные группы:

1. «композиционные» подходы, при которых выходное изображение создается путем применения нескольких заранее определенных преобразований, параметры которых либо заданы, либо предсказываются тем или иным методом;
2. «сквозные» подходы, при которых выходное изображение генерируется тем или иным алгоритмом (к примеру, специальной нейронной сетью), который получает исходное изображение на вход.

Один из подходов, который можно отнести группе композиционных, представлен в работе [31] и включает в себя применение различных методов ретуширования к входному изображению. Он включает в себя использование метода, основанного на глубоком обучении с подкреплением, чтобы найти оптимальную последовательность глобальных преобразований изображения.

Представленное решение, однако, требует значительных вычислительных ресурсов, поскольку оно основано на глубокой нейросетевой модели VGG-16 [32] с последующей оценкой гистограмм признаков для оценки параметров стратегии улучшения изображения.

Аналогичный подход на основе фильтров представлен в работе [33] и использует интерпретируемые преобразования изображений (в работе также представлен подробный обзор возможных интерпретируемых преобразований). Реализация, предлагаемая в работе, также использует тяжеловесную вычислительную модель, склонна к неестественным искажениям цветов и обеспечивают низкую скорость вывода (по крайней мере, без дополнительных вычислительных оптимизаций).

В работе [34] представлен метод улучшения изображений, основанный на специально разработанных «самоинтерпретируемых» фильтрах, по методологии с учетом мнений экспертов. Предлагаемая модель обладает устойчивостью к искажениям, возникающим из-за субъективизма экспертов, однако в численных экспериментах метод показал более низкие результаты, чем альтернативные свозные нейросетевые подходы (к примеру, такие как метод Pix2Pix [35]).

Другой подход основан на комбинированной архитектуре, которая разделяет процесс улучшения на уточнение по каналам и по пикселям [36]. В этом решении используется остаточная модель опорной сети [37] в качестве подсистемы извлечения признаков, нелокальный блок внимания [38] для последующей агрегации информации и анализа признаков и глобального линейного отображения для финального улучшения характеристик визуального восприятия изображения. Похожий подход также предложен в работе [39], где используется параметризованное преобразование цвета на изображении.

Одной из передовых работ в области комбинированных моделей улучшения изображения является работа Татанова и Самарина [40], в которой авторы используют несколько известных фильтров и добиваются улучшения с помощью дополнительной регуляризации. Принципиальная схема работы данного подхода представлена на рис. 1.4.

Касаемо группы сквозных подходов, наиболее широко известными современными методами являются те, которые основанные на генеративно-состязательных нейронных сетях (Generative Adversarial Networks, GAN) [41], среди которых стоит выделить несколько хорошо зарекомендовавших себя методов, таких как Deep photo enhancer [42] и Pix2Pix [35]. Недостатком таких

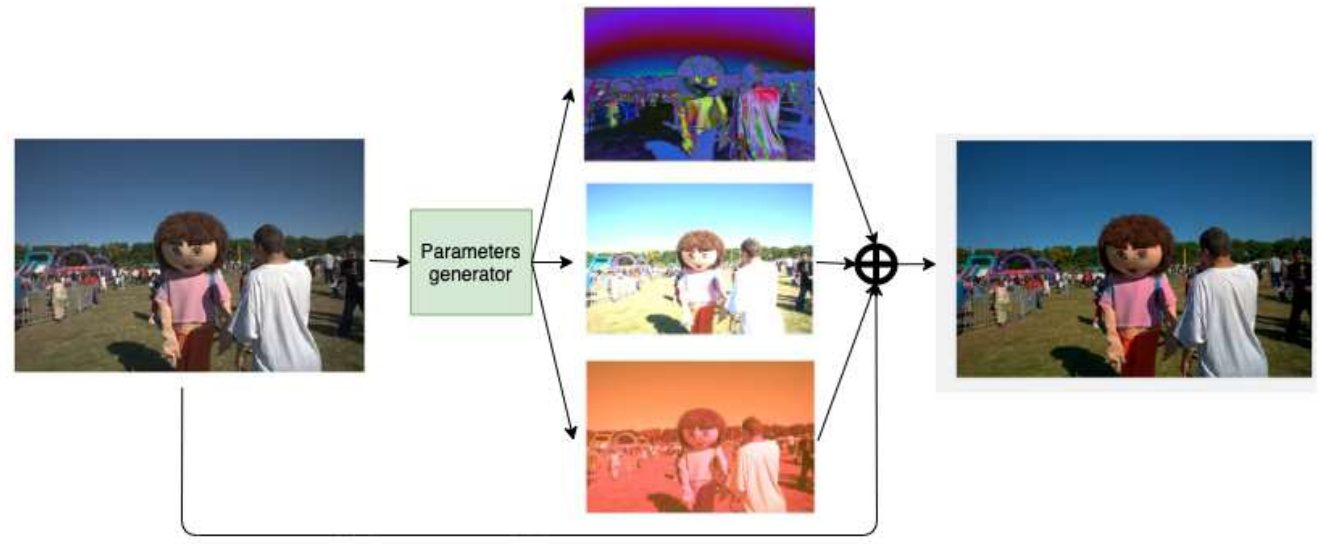


Рисунок 1.4 — Принципильная схема работы композиционного подхода к улучшению изображения LFIEF [40].

методов является их подверженность артефактам. Так, метод, описанный в работе [42] использует улучшение GAN с использованием метрики Вассерштейна и адаптивную схему взвешивания, однако конкурентноспособных результатов на наборе данных MIT Adobe FiveK, на котором проводились исследования, этот метод не демонстрирует.

Среди других методов, основанных на GAN, следует выделить метод EnhanceGAN [43], благодаря его конкурентноспособным количественным результатам по улучшению цвета (по отношению к экспертной оценке). К сожалению, предложенная авторами модель достаточно тяжеловесна, что делает использование данной архитектуры на мобильных устройствах практически невозможным. Детали устройства архитектуры EnhanceGAN представлены на рис. 1.5, на котором продемонстрирован подход по оценке разницы изображений до и после обработки путем применения дополнительных операций, таких как обрезка и анализ отдельных цветовых каналов.

Помимо коррекции цветовых характеристик изображения важной задачей является методы очистки фона (подавление шума, удаление теней и т.п.) для улучшение качества изображений документов. Так, авторы работы [44] предполагают, что постоянный цвет фона создает «карту теней», которая сопоставляет локальные цветовые характеристики фона с некоторым глобальным средним. Подобным образом проводится совместный анализ локальных и глобальных характеристик изображения документа в работах [45; 46]. В работе [47] предложен оригинальный метод коррекции освещенности для изображений

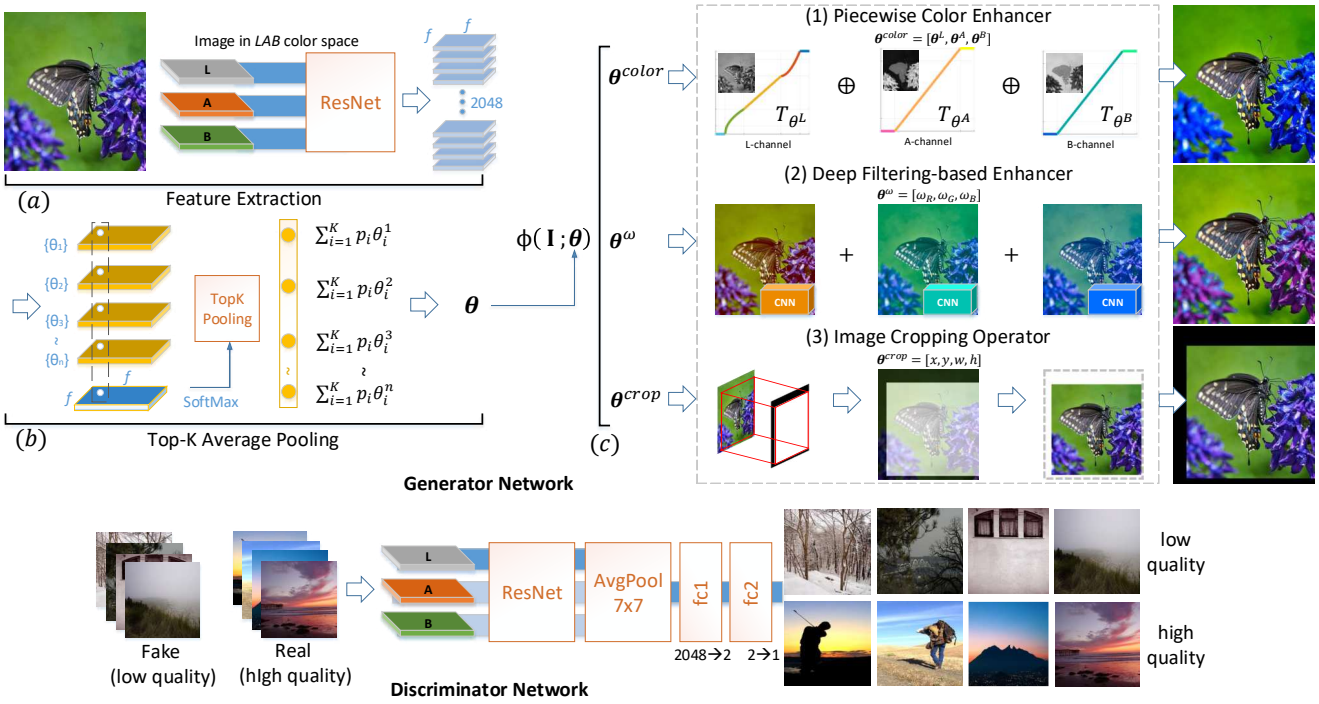


Рисунок 1.5 — Принципиальная схема работы архитектуры EnhanceGAN [43].

документа, вдохновленный топологическими поверхностями, заполненными водой. В работе [48] предложен подход к улучшению изображения документа путем представления входного изображения в виде трехмерного облака точек. Также была предложена глубокая нейросетевая архитектура [49] для оценки общего цвета фона документа и «карты внимания», которая вычисляет оценку вероятности того, что пиксель принадлежит бестеневому фону. В работе [50] предлагается метод коррекции освещенности и исправления изображения документа с использованием нейронней сети типа «кодировщик-декодировщик» на основе локальных регионов изображения.

В работе [51] предлагается глубой сверточный автоматический кодировщик на основе техники пропуска связей в нейронных сетях, предсказывающий изображение шума на документа, который можно убрать простым попиксельным вычитанием. В работе [52] предлагается сквозная структура улучшения документов с использованием условных генеративно-состязательных сетей (Conditional Generative Adversarial Networks, cGAN), где для генераторной сети используется архитектура на основе U-Net.

Особого внимания заслуживают методы, основанные на моделях с небольшим числом параметров. В работе [53] предложено облегченное решение на основе фильтров, учитывающее локальные особенности изображений, в рабо-

те [54] также представлена облегченная модель, основанная на оценке как глобальных, так и локальных характеристик изображения.

1.2.3 Бинаризация изображений документов

Важной схемой очистки изображения, часто применяемой в системах анализа и распознавания документа, является бинаризация – разделение пикселей изображения на пиксели переднего плана (как правило, помечающиеся черным цветом и включающие весь текст, присутствующий на документе) и пиксели фона (как правило, помечающийся белым цветом) [55]. При правильной бинаризации пиксели переднего плана сохраняются, а фон делается однородным. Такая классификации пикселей позволяет существенно уменьшить объем памяти для хранения изображения, а также упрощает все последующие этапы распознавания.

Методы бинаризации могут быть сгруппированы по принципу построения пороговой поверхности. Пороговой поверхностью $T(x,y)$ называется двумерная функция, заданная на области определения исходного изображения I, в каждом узле которой содержится значение порога бинаризации для соответствующего пикселя исходного изображения. Сама же процедура построения бинарного образа таким образом сводится к применению следующего простого правила: $B(x,y) = [I(x,y) > T(x,y)]$.

Простейшей группой методов являются так называемые глобальные методы бинаризации, в которой пороговая поверхность T является одинаковой в каждой своей точке. Такие алгоритмы характеризуются очень высокой производительностью, но при этом являются крайне чувствительными к содержанию образа изображения. Наибольшее распространение в этой группе методов получил метод Оцу, предложенный еще в 1979 году [56]. Метод Оцу определяет пороговое значение по гистограмме яркости изображения, минимизируя взвешенную внутриклассовую дисперсию. Глобальные методы не способны справляться с неравномерно освещенными документами, выбросовыми шумами и другими искажениями, были предложены модернизации методов. Применение специальной обработки изображения до глобальной бинаризации может существенно улучшить результат. Одной из популярных техник такой предобработки

является оценивание и нормализация фона изображения [57]. Поскольку оценка таких порогов может быть затруднена для испорченных изображений документов, существует адаптивное обобщение метода Оцу [58] для более устойчивой бинаризации.

В группе алгоритмов локальной (или адаптивной) бинаризации [59] значения пороговой поверхности зависят не только от яркостного значения самого пикселя, но и от значений его соседей в некоторой заданной окрестности. Такие алгоритмы значительно более устойчивы и к условиям оцифровки образа, и к наличию шума на полученном изображении. С другой стороны, локальные методы требовательны к подходящему выбору их настроечных параметров, а сам подбор этих значений параметров является нетривиальной задачей. Поэтому предлагаются локальные методы с автоматически определяемыми значениями параметров, в том числе – с использованием обобщения критерия Отсу [60]. В работе [61] Саувола и Пиетикайнен предложили локальный адаптивный метод пороговой обработки для задачи бинаризации изображения. Алгоритм Сауволы вычисляет локальный порог для каждого пикселя, что позволяет ему улучшать изображения с определенными типами ухудшения качества, но неспособными хорошо работать в условиях низкой контрастности. Чтобы улучшить работу алгоритма Сауволы в условиях низкой контрастности, существуют также его многомасштабные (multi-scale) обобщения [62]. Широко используемым методом, также основанным на оконных статистиках, является метод Ниблака [63].

Классические методы бинаризации обладают высокой производительностью, но дают стабильно высокую точность не во всех приложениях. Поскольку они используют информацию, основанную на интенсивности пикселей, для выполнения локального или глобального порогового отсечения для получения бинарного изображения, и зачастую оказываются малоэффективными в случае изображений документов, с неоднородным или зашумленным фоном [64]. Для увеличения точности в некоторых работах предлагается использовать несколько различных результатов бинаризации и уже на их основе принимать итоговое решение о классификации каждого из пикселей, но это явный паллиатив. Не удивительно, что самой популярной в данный момент группа методов основывается на попиксельной классификации с использованием методологии обучения машин. В основном речь здесь идет об искусственных нейронных сетях различной архитектуры [65; 66]. Такие алгоритмы обеспечивают высокую точность бинаризации в ряде сценариев, чаще всего, не требуют никаких предварительь-

ной обработки изображения документа, но при этом являются вычислительно сложными.

Существуют также ряд альтернативных методов, такикх как метод, описанный в [67], который предполагает расчет и использование ориентации текстового штриха на основе фильтра Габора. В работе [68] предлагается быстрый метод бинаризации документов на основе нечеткой кластеризации методом K-средних. В этом подходе пиксели изображения группируются в три класса: текстовые пиксели, фоновые пиксели и сомнительные пиксели. Затем сомнительные пиксели дополниительно классифицируются на текстовые и фоновые на основе структурной симметрии. В работе [69] предлагается метод, преобразующий входное изображение в матрицу подобия Бхаттачарьи, затем для получения бинарного изображения используется классификатор максимальной энтропии.

Выбор подходящего метода бинаризации в нынешних условиях является нетривиальной задачей. Для отслеживания успехов в области бинаризации документов была запущена платформа DIB [70]. В ее рамках проводится консолидация всех имеющихся знаний и артефактов в области бинаризации документов. На платформе представлены датасеты, такие как DIBCO (Document Image Binarization Contest), Nabuko, LiveMemory, синтетические данные и многие другие. В рамках данной платформы также стали проводиться соответствующие конкурсы, отличительной особенностью которых стал учет времени работы предлагаемых решений. Так, в течение нескольких последних лет лидирующие позиции в конкурсе DIBCO, объектом которого ялвются изображения документов различной природы, занимают нейросетевые подходы, в том числе методы, основанные на архитектуре U-Net, к примеру, метод, описанный в работе [66].

Стоит также отметить, что предварительная бинаризация изображения документа не обязательно приводит к увеличению ожидаемой точности распознавания. Так, в аналитической работе [71] показано на примере задачи анализа и распознавания документов, удостоверяющих личность, что использование даже лидирующих алгоритмов бинаризации изображений документов может приводить к уменьшению точности распознавания (по крайней мере, при использовании нескольких наиболее известных открытых систем распознавания текста) по сравнению с распознаванием исходных изображений. При этом, согласно полученным результатам, точность распознавание все же улуч-

шается, при использовании идеальной маски переднего плана, что может свидетельствовать о том, что ухудшение точности распознавание при бинаризации обусловлено неустойчивостью методов к определенным классам искажений исходных изображений, либо отсутствии какого-либо универсального подхода, который бы приводил к улучшенным изображениям, пригодным для максимально качественного распознавания в любых условиях.

1.3 Методы классификации, поиска и извлечения информации на изображениях

1.3.1 Классификация и локализация документа по особым точкам

В анализе изображений особыми точками называются точки изображения, локальные окрестности которых обладают отличительными особенностями по сравнению с другими окрестностями. В последние несколько лет методы, основанные на поиске, анализе и сопоставлении особых точек, демонстрируют хорошие результаты в таких задачах, как детектирование объектов, классификация изображений, склейка панорамы, распознавание лиц и других.

Детекторы особых точек – методы поиска области интереса (Region of Interest, ROI), которые рассматриваются, как опорные точки/области для локальных дескрипторов, содержащих описание особой точки и особенности ее окрестности. Впоследствии, набор дескрипторов позволяет охарактеризовать локальные участки изображения. Известны различные методы детектирования и вычисления дескрипторов. Чаще всего эти методы предлагаются парами и носят одно название, но это не обязательно так.

В задаче распознавания документов механизм особых точек используется для классификации и локализации документа или его части путем сравнения с эталоном. Локализация документов с бланками фиксированной геометрии может быть осуществлена этим методом в том числе при съемке камерой, в присутствии проективных искажений. В таком случае для сопоставления созвездий особых точек обычно используются алгоритмы семейства RANSAC, это позволяет полностью определить внутреннюю систему координат документа на его

изображении. При таком сценарии использования важным свойством особых точек оказывается аффинная инвариантность (проективное преобразование локально аффинно). При большом числе эталонов классификация осуществляется в модели «мешок признаков» (Bag of Features) с помощью быстрых поисковых деревьев [72].

Простейшими примерами особых точек являются углы, концы отрезков и другие топологические особенности морфологического скелета изображений. Особые точки имеет преимущество перед другими особенностями изображения, такими как края и области, поскольку лучше локализованы, и при этом имеют высокоинформативное содержание [73]. Особые точки, как правило, стабильны при преобразованиях изображений, а также при преобразованиях, меняющих угол обзора. К недостаткам использования особых точек можно отнести снижение вероятности правильной классификации при увеличении объема возможных классов, неустойчивость к сильной расфокусировке. Среди быстрых методов детектирования особых точек можно выделить детектор углов Харриса [74], FAST [75], Difference-of-Gaussian (DOG) [76] и YAPE [77].

Алгоритм детектирования Scale Invariant Feature Transform (SIFT) не чувствителен к масштабированию и повороту изображения и частично инвариантен к изменению освещенности и точки обзора камеры [76]. Алгоритм SIFT обладает высокой степенью отличительности, то есть позволяет с высокой вероятностью правильно сопоставить один признак с большой базой данных признаков, обеспечивая основу для распознавания объектов и сцен. Затраты на извлечение особых точек в нем оптимизированы с использованием каскадной фильтрации, при которой более дорогостоящие операции применяются только в местах, прошедших начальную проверку. Несмотря на это, основным недостатком SIFT считается высокая вычислительная сложность.

В работе [78] в 2008 году был представлен метод Speeded Up Robust Features (SURF), который основан на гауссовом многомасштабном анализе изображений. Детектор SURF основан на детерминанте матрицы Гессе и использует интегральные изображения для повышения скорости обнаружения признаков. 64-битный дескриптор SURF описывает каждую обнаруженную особую точку с распределением вейвлет-откликов Хаара в определенной окрестности. Для каждого из 44 субрегионов, каждый вектор признаков содержит четыре части:

$$v = \left(\sum dx, \sum dy, \sum |dx|, \sum |dy| \right), \tag{1.3}$$

где вейвлет-отклики dx и dy суммируются для каждого субрегиона, а абсолютное значение откликов $|dx|$ и $|dy|$ обеспечивает полярность изменения интенсивности. Алгоритм SURF инвариантен к вращению и масштабу, но имеет сравнительно слабую аффинную инвариантность. Однако дескриптор может быть расширен до 128 бит для того, чтобы иметь дело с большими изменениями угла обзора. Главным преимуществом SURF перед SIFT является его низкая вычислительная стоимость.

Алгоритм Oriented FAST and Rotated BRIEF (ORB) был предложен в 2011 году. ORB представляет собой комбинацию модифицированного детектора FAST (Features from Accelerated Segment Test) [75] и направленно-нормализованного дескриптора BRIEF (Binary Robust Independent Elementary Features). FAST-особенности детектируются в каждом слое многомасштабной пирамиды, а качество найденных точек оценивается с помощью детектора углов Харриса. Поскольку метод BRIEF крайне чувствителен к поворотам, то была использована модифицированная версия BRIEF-дескриптора. Ориентация FAST-особенностей оценивается центроидом интенсивности, который представляет собой смещение между интенсивностью определенного угла и его центром. Это смещение оценивает ориентацию, которая является вектором между местоположением объекта и центроидом [79]. В [79] центроид определяется как:

$$C = (m_{10}/m_{00}, m_{01}/m_{00}), \quad m_{pq} = \sum x^p y^p I(x, y). \tag{1.4}$$

Метод ORB устойчив к масштабированию, вращению и ограниченным аффинным искажениям.

В работе [80] описан алгоритм Binary Robust Invariant Scalable Keypoints (BRISK), который детектирует углы, используя метод AGAST [81], и фильтрует их с помощью оценки угла Харриса при поиске максимумов в многомасштабной пространственной пирамиде. BRISK-дескриптор основан на определении характерного направления каждого признака для достижения инвариантности ко вращению. Чтобы вычислить ориентацию ключевой точки k, BRISK использует локальные градиенты между парами выборок, которые определяются следующим образом:

$$g(p_i, p_j) = (p_j - p_i) \cdot \frac{I(p_j, \sigma_j) - I(p_i, \sigma_i)}{|p_j - p_i|^2}, \tag{1.5}$$

Таблица 1 — Качество детектирования и скорость работы методов поиска особых точек

Дескриптор	Детектор	Качество (для разных поворотов) (%)						Среднее время для 15 кадров (с)
		0	30	60	90	180	Среднее	
SIFT	SIFT	88,4	84,6	83,9	87,7	87,7	86,4	1,21
SIFT	SURF	98,6	93,5	93,2	98,6	98,5	96,4	2,76
SIFT	BRISK	98,7	98,8	98,4	98,7	98,5	98,6	1,40
SIFT	ORB	98,3	97,7	97,1	98,5	98,3	97,9	1,50
SIFT	AKAZE	64,7	61,9	61,6	65,0	65,0	63,6	0,70
SURF	SURF	92,5	10,0	10,0	92,6	92,6	59,5	3,06
SURF	SIFT	55,2	05,0	05,0	46,0	43,8	31,0	1,20
SURF	BRISK	94,0	69,8	69,9	93,2	93,2	84,2	1,80
SURF	ORB	96,9	83,0	80,0	96,1	96,0	90,4	1,30
SURF	AKAZE	01,0	00,3	00,4	01,0	01,0	00,7	0,80
ORB	ORB	81,2	79,5	79,0	80,0	80,0	79,9	6,81
AKAZE	AKAZE	73,1	70,0	68,7	67,2	72,9	70,4	9,90
KAZE	KAZE	80,7	00,4	00,5	14,9	65,1	32,3	2,10

где (p_j, p_i) – одна из пар точек выборки. Сглаженные значения интенсивности в этих точках равняются $I(p_j, \sigma_j)$ и $I(p_i, \sigma_i)$ соответственно. Для обеспечения инвариантности к освещению результаты простых проверок яркости также объединяются, и дескриптор строится в виде двоичной строки. Метод BRISK инвариантен к масштабу, вращению и ограниченным аффинным искажениям.

Разработка новых методов детектирования и описания особых точек продолжается, и помимо методов, которые уже себя хорошо зарекомендовали, появляются экспериментальные методы, использующие нейронные сети [82].

В обзоре методов детектирования и описания особых точек [83] приводятся результаты замеров качества и скорости работы разных методов (см. таблицу 1). Авторы статьи [83] сгенерировали 1630 видео. Для замера качества был использован открытый датасет WikiBook [84], который содержит 700 изображений листов А4 с текстом. Качество замерялось по точности правильного поиска области интереса на изображении. Для замера качества работы методов также могут быть использованы открытые датасеты, например, Tobacco dataset [85].

1.3.2 Локализация границ документа

Локализация методом поиска особых точек предполагает, что бланк документа имеет уникальные особенности, причем они заранее известны. Во многих случаях это не так. У офисных документов (договоров и пр.) бланк может отсутствовать, а у банковских карт бланки чрезвычайно разнообразны и заранее, как правило, неизвестны. В таком случае на этапе геометрической нормализации в качестве модели документа обычно рассматривают прямоугольник. Модель может быть дополнена известным соотношением сторон документа. При сканировании в некоторых случаях осмысленно дополнить модель абсолютными размерами. Образом документа на сканированных изображениях является прямоугольник, отличающийся от прообраза сдвигом, поворотом и изотропным сжатием, причем коэффициент сжатия, как правило, известен (он задается разрешением сканирования). На изображениях, полученных с мобильных камер, образ документа – почти произвольный выпуклый четырехугольник, поскольку изображение подвергается проективной дисторсии.

Работы, посвященные локализации границ документа, рассматривают как сканированные изображения, так и фотографии. Второй случай более общий, так как на изображении может появиться неоднородный фон, который увеличивает вероятность ложных срабатываний, кроме того, границы документа могут быть частично заслонены. Мы будем рассматривать задачу детекции документов на изображениях, полученных с камеры. Описанные ниже методы могут быть применены и для сканированных изображения при введении дополнительных геометрических ограничений на результирующий четырехугольник.

Существует три основных подхода для решения задачи детекции границ документа: анализ контуров, сегментация изображения и детекция углов документа.

Самым широко распространенным подходом является анализ контуров на изображении документа. Одной из классических работ, использующих этот подход, является [86], в которой представлен метод детекции лекционной доски на изображении. Предложенная авторами схема состоит из следующих этапов:

– выделение контуров на изображении, приведенном к градациям серого;
– детекция прямых линий на карте контуров с помощью преобразования Хафа;

– составление всех возможных альтернатив четырехугольников, удовлетворяющих геометрическим ограничениям;
– оценка консистентности каждого четырехугольника B: $\text{consistency}(B) = \sum_{b \in B} \text{consistency}(b)$, где b – сторона четырехугольника, а $\text{consistency}(b)$ вычисляется как доля пикселей стороны, которые имеют соответствие на карте контуров;
– выбор лучшего четырехугольника и его уточнение.

Данная схема легла в основу многих алгоритмов детекции границ документов. Авторы [87], придерживаясь в остальном описанного метода, предлагают отказаться от приведения к градациям серого перед выделением границ из-за возможной потери информации. Вместо этого изображение переводится в цветовое пространство CIELAB. Для каждого канала строится отдельная карта контуров, после чего результаты объединяются с помощью логической операции «или».

Существуют также работы, которые предлагают дополнить описанную выше схему методами фильтрации высокочастотного шума, что позволяет снизить количество ложных срабатываний детектора прямых. В статье [88] описан метод фильтрации текста: на карте контуров выделяются компоненты связности, вычисляются их окаймляющие прямоугольники, и каждая компонента оценивается по соотношению сторон прямоугольника, его размеру относительно региона интереса и количеству пикселей.

Для построения альтернатив четырехугольников классическая схема предполагает использование переборов с отсечениями, этого же подхода придерживается большинство описываемых здесь работ.

В некоторых работах также описаны новые способы оценки консистентности четырехугольника. В статье [89] соотношение сторон документа P_d считается известным, и оценка четырехугольника учитывает отклонение оцененного соотношения сторон P_e от заданного. Также используется вес соответствующих прямых в пространстве Хафа W_l, и для каждого угла вычисляется штраф P_c, равный сумме интенсивности пикселей на карте контуров вдоль соответствующих сторон и вне их. Это позволяет отсеять четырехугольники, одна из сторон которых образована контуром, выходящим за пределы четырехугольника:

$$\text{consistency}(B) = |1 - P_d/P_e| \left(\sum_{i=1}^{4} W_{l_i} - \sum_{i=1}^{4} P_{c_i} \right). \tag{1.6}$$

Стоит отметить, что основным предположением подхода, основанного на анализе контуров, является то, что все стороны документа видны на изображении и имеют сильный контраст по отношению к фону. Однако при использовании в реальных системах распознавания документов одна из сторон документа может находиться за пределами кадра или быть заслонена, что делает практически невозможным формирование четырехугольника по 4 сегментам.

В рамках второго подхода поиск границ документа рассматривается как задача сегментации изображения. Такой подход накладывает более слабые ограничения на видимость границ документа, однако в большинстве случаев алгоритмы обладают большей вычислительной сложностью. Авторы работы [90] используют функционал «local-global variational energy», основанный на вероятностных распределениях векторов координат и цветовых характеристик пикселя для регионов документа и фона. Задача сегментации решается путем оптимизации этого функционала.

Отдельная группа работ (например, [91]) посвящена использованию дерева форм, которая является иерархическим представлением изображения по принципу включения линий уровня изображения для детекции объектов на изображении. Стоит отметить, что алгоритм, описанный в работе [91], показал лучший результат в первом испытании тематического конкурса ICDAR 2015 Smartdoc.

В работе [92] на изображении выбирается несколько точек фона, сегментируются области, похожие на эти точки, а четырехугольник документа находится итеративно, минимизируя количество похожих на фон пикселей внутри четырехугольника, и максимизируя их количество вне его.

Для решения задачи сегментации области документа используют также нейросетевой подход: архитектуру U-Net [93], OctHU-PageScan – Fully Octave Convolutional Neural Network, основанный на этой же архитектуре, HoughEncoder [94] – автоэнкодер, использующий прямое и транспонированное преобразования Хафа.

Третий подход предполагает детекцию углов документа. В работе [95] для этого используется двухэтапная нейросетевая схема. Первая нейронная сеть грубо оценивает положение углов документа на входном изображении; после чего каждый из углов итеративно уточняется второй нейронной сетью, которая принимает на вход локальную окрестность каждого угла документа. Ограничениями данного подхода является требование видимости и высокого контраста с

фоном всех углов документа, что зачастую не выполняется при съемке с мобильной камеры.

Существует несколько открытых датасетов, которые позволяют измерить качество детекции границ документа: датасет, который использовался в соревновании ICDAR 2015 SmartDoc [72], датасет SEECS-NUSF [95], CDPhotoDataset, описанный в работе [96], будет доступен после подведения итогов конкурса соревнование ICDAR2021. Наиболее часто используемой метрикой для оценки качества детекции границ документа является индекс Жаккарда. Он, в частности, использовался в конкурсе ICDAR 2015 SmartDoc.

1.3.3 Классификация и локализация по общему виду

Существует еще один подход к локализации и классификации документов. Он применяется в случае, если бланк документа не обладает уникальными стабильными локальными особенностями, но документ все же легко опознается по общему виду (visual appearance). В таком случае для классификации документа могут быть применены те или иные методы машинного обучения. Методы классификации, не обеспечивающие одновременной локализации документа, применяются после рассмотренных ранее алгоритмов геометрической нормализации путем локализации границ. Достаточно подробный обзор таких методов можно найти в разделе 2.2.2 работы [2].

Кроме того, существует как минимум один метод машинного обучения, решающий обе задачи. Метод детектирования объектов, основанный на быстром вычислении локальных контрастов (так называемых признаков Хаара), был предложен Виола и Джонсом для поиска лиц на фотографиях [97]. При использовании ими понятия интегрального изображения вычисления контраста для произвольной прямоугольной подобласти в среднем выполнялось за несколько тактов центрального процессора. Это позволило быстро применять детекторы, основанные на вычислении признаков Хаара, к каждой прямоугольной подобласти изображения, взятой со всевозможными масштабами из заранее заданного множества. В результате быстродействие метода Виолы и Джонса было равно 15 кадрам в секунду на персональном компьютере 2000 года без использования вычислений на видеокарте.

Другим преимуществом метода Виолы и Джонса по сравнению с нейросетевыми методами глубинного обучения является меньшая требовательность к объему обучающих данных. Это важно в задачах, связанные с распознаванием документов, удостоверяющими личность, поскольку для них задача сбора обучающих данных сопряжена с серьезными правовыми ограничениями. В работе [98] были обучены несколько классификаторов Виолы и Джонса для локализации нескольких типов документов, удостоверяющих личность, на изображениях, полученных с помощью сканера, причем положительная выборка для каждого из классов состояла всего из нескольких сотен примеров. Метод обладает устойчивостью к шумовым элементам, таким как гильоширные рисунки, защитные волокна, рукописные пометки и т.п., поскольку строится на признаках низкого разрешения, формирующих визуальный образ «в целом», игнорируя мелкие детали.

Метод Виолы и Джонса применим также для поиска и классификации опорных элементов бланка или заполнения в случаях, когда документ не содержит достаточного числа особых точек. Еще одним интересным объектом поиска являются круглые печати, поиск и снятие которых играет особую роль при распознавании документов. Во-первых, проверка наличия печати необходима для проверки подлинности документа. Во-вторых, печати содержат в себе значимую информацию, распознавание и контроль которой сам по себе представляет ценность [99].

В настоящее время для решение задачи детекции объектов, в том числе документов, также широко используются нейросетевые методы. Наверное, самым известным и широко используемым нейросетевым подходом к этой задаче, можно считать метод YOLO (You Only Look Once, дословно «ты смотришь только один раз») [100]. Метод YOLO предсказывает ограничивающие рамки искомых объектов и одновременно производит их классификацию практически для каждой области изображения (границы размеров проверяемых областей являются гиперпараметрами метода).

Авторы работы [101] провели анализ комбинаций различных функций, улучшающих работу стандартных сверточных сетей в задаче детекции. Оказалось, что некоторые функции работают исключительно с определенными моделями и исключительно с определенными задачами или только с небольшими наборами данных; в то время как некоторые функции, такие как функция нормализация пакетов и остаточные (residual) соединения, применимы к боль-

шинству моделей, задач и наборов данных. Авторы утверждают, что такие универсальные функции включают взвешенные остаточные соединения (WRC), межэтапные частичные соединения (CSP), перекрестную мини-пакетную нормализацию (CmBN), самосостязательное обучение (SAT) и Mish-активацию. Авторы представили новую, 4-ую по счету версию архитектуры YOLO, включающую WRC, CSP, CmBN, SAT, активацию Mish, а также дополнительный набор техник для аугментации обучающей выборки, регуляризации при обучении, и др.

1.4 Методы анализа содержания документов

1.4.1 Методы анализа структуры документа

После предварительной обработки изображения и локализации границ документа в изображении важнейшим этапом является анализ его структуры и сегментация документа на составные части. Задачи, которые входят в блок анализа структуры документа, можно разделить на следующие группы:

- детектирование и локализация текстовых элементов (строк, слов) документа;
- анализ структуры текстовых блоков документа и сегментация на отдельные текстовые блоки, колонки, параграфы, или текстовые поля (см. рис. 1.6);
- определение последовательности прочтения текстовых блоков документа (колонок, параграфов, полей);
- детектирование и локализация графических элементов документа – фигур, рисунков, формул, печатей, штампов и т.п.

Поскольку сегментация изображения документа на информационные фрагменты в значительной степени зависит как от структуры документа, так и от специфики конкретных систем распознавания, в литературе наблюдается широкий спектр методов и подходов, предлагаемых для решения такого рода задач, группируемых различным способом.

Рисунок 1.6 — Пример структуры текстовых блоков документа [2].

Со стороны особенностей документов методы анализа их структуры подразделяются на методы анализа структуры жестких форм (т.е. документов, переменные элементы которых расположены на различных экземплярах в одних и тех же местах), анализа документов с одноколоночной и многоколоночной манхэттенской структурой (т.е. документов, которые могут быть декомпозированы на непересекающиеся ортотропные прямоугольные регионы, каждый из которых содержит либо текстовый блок, либо рисунок, таблицу и т.п.), и анализ гибких документов произвольной структуры (т.е. документов, графические элементы и текстовые блоки которых могут быть произвольно формы и с произвольным направлением текста). Методы разбора различных структур документов также подразделяются на методы, предполагающие априорные знания о шаблоне («грамматике») структуры компонентов документа, и на методы, применимые к документам с заранее неизвестной структурой.

Алгоритмы анализа структуры документа с точки зрения подхода можно условно разделить на три группы: подходы «bottom-up», подходы «top-down» и гибридные.

Подходы группы «bottom-up» опираются на предварительное выделение и анализ отдельных компонентов документов, таких как рисунки, текстовые строки и слова, статические тексты, особые точки, маркеры и т.п. и на их основе восстанавливают общую структуру. К этой группе относятся методы на основе анализа групп компонент связности, идентификации прямолинейных элемен-

тов [102], ключевых текстовых компонентов и их совместных связей [103], а также методы на основе диаграмм Вороного и триангуляции Делоне.

Подходы группы «top-down» предполагают анализ изображения документа в целом и предварительной сегментации структуры документа на блоки, каждый из которых впоследствии разбирается независимо. К таким методам относятся анализ проекций изображения на горизонтальную и вертикальную оси и анализ просветов [104], декомпозиция документа на блоки при помощи оптимального накладывания множества Гауссовых ядер [105] или других типов функций.

К третьей группе относятся гибридные подходы, состоящие в совместном применении нескольких техник, объединение подходов «top-down» и «bottom-up» с дополнительными модификациями, определяемыми спецификой задачи или спецификой документа. К этой группе также можно отнести методы на основе машинного обучения и применения глубоких нейронных сетей, наиболее активно развивающиеся в последние годы. Так, модели на основе сверточных и рекуррентных нейронных сетей, в том числе глубокие архитектуры VGG-16 и YOLOv2/v3 используются для детектирования, поиска и классификации отдельных компонентов документов, таких как фигуры и рисунки [106] и формулы, а также текстовых элементов документов и связей между ними. Глубокие сверточные модели используется для классификации блоков документов [107], полносверточные нейронные сети, сверточные нейронные сети с мультипликативными слоями (Trainable Multiplication Layers, TML) и «сиамские» сети используются для решения задачи семантической сегментации документов [108].

Отдельно стоит выделить задачу детектирования и локализации произвольных текстовых строк и отдельных слов, которая широко обсуждается в литературе вкупе с методами анализа структуры документов (хотя, строго говоря, является более общей задачей, которая возникает также вне систем обработки документов). К этой задаче применяются термины «text in the wild», «word spotting» или «text spotting». Для ее решения создаются методы, позволяющие выделять участки входных изображений, содержащие символы, графемы, слова и текстовые строки целиком, в условиях пиксельных (яркостных, цветовых) и геометрических искажений. Методы включают как структурный анализ участков изображения с последующим комбинаторным выбором участков, обладающих признаками текста, использование техник обучения локальных

признаков текста, а также глобальное использование глубоких сверточных нейронных сетей для выделения пикселей текста на произвольных изображениях [109].

С обширностью темы анализа структуры документов, и таких подтем как сегментация изображения документа и поиск графических или текстовых элементов, связан интерес научного сообщества к разработке новых алгоритмов и методов решения связанных задач. Регулярно проводятся научные конкурсы, такие как конкурс распознавания документов со сложной структурой (ICDAR Recognition of Documents with Complex Layouts) и другие. Для оценки вновь публикуемых алгоритмов используются ставшие традиционными открытые наборы данных, такие как PRImA [110] (анализ структуры документов), COCO-text [111] (поиск и распознавание текста на естественных изображениях) и наборы данных международного проекта развития систем анализа документов MAURDOR [112]. Помимо них, международные коллективы создают новые наборы данных, отражающие специфику и характеристики отдельных типов документов, например, набор BID [113] для анализа документов, удостоверяющих личность.

1.4.2 Применение искусственных нейронных сетей для распознавания символов и слов

С конца 90-х годов прошлого века большинство методов и подходов к оптическому распознаванию символов, предлагаемых для решения данной задачи основано на искусственные нейронные сети (ИНС). Все методы, используемые для распознавания строк можно поделить на две большие группы: методы с явной сегментацией строки на символы и без явной сегментацией строки.

При сегментации строки для каждого символа строится окаймляющий прямоугольник, содержимое которого передается классификатору. В данной группе методов сегментация считается задачей более сложной, чем классификация. Все стандартные алгоритмы сегментации испытывают трудности при распознавании касающихся или пересекающихся символов (при слиянии букв в некоторых шрифтах в лигатуры, например, ff) или же при обработке символов состоящих из более, чем одного компоненты связности (примитива). Особенно

эти проблемы проявляются при наличии сложного фона у документа. На этом, в частности, базируются современные методы создания CAPTCHA. Поэтому часто алгоритмы сегментации содержат эвристики, которые затрудняют их использование для различных языков и алфавитов.

Алгоритмы сегментации строки делятся на две группы:

– обработка связных компонент;

– анализ проекции на горизонтальную ось.

На основе анализа проекций разработан подход эвристической излишней сегментации. В данном подходе строятся несколько путей сегментации (несколько вариантов разбиения строки на символы) и из них выбирается лучший [114]. При этом строка может быть заведомо разделена на излишнее количество частей (с делением символов), после чего каждая часть классифицируется, и на основе оценок сегментации классификатора строится лучший путь [115]. Такие методы иногда используют эвристики, например, признак моноширинности шрифта [114] или отсутствие символов из нескольких компонент связности [115]. В алгоритмы сегментации может быть встроена ИНС (например, применение классификатора, который умеет отличать символ от его части [116]).

После сегментации строки проводится классификация образов символов. Для нее в основном используются сверточные нейронные сети (СНС). Также существуют подходы на основе каскада ИНС, когда итоговый ответ строится на основе оценок всех входящих в каскад сетей. Для достижения state-of-the-art результатов сейчас строятся каскады сетей с очень большим количеством параметров. Например, нынешняя наименьшая ошибка на наборе данных MNIST (0,14%) достигнута каскадом из 15 сетей по 35.4×106 каждая [117], в то время как ошибка, допускаемая человеком, составляет примерно 0,20%. Таким образом, очень многие современные архитектуры содержат излишнее число параметров и склонны к переобучению [118].

При решении задачи мобильного распознавания зачастую разработчики систем используют готовые программные пакеты (такие, как открытый пакет Tesseract [119]), решающие задачу распознавания текста общим, и далеко не самым эффективным, образом, в контексте конкретной системы. Авторы работ по методам распознавания текстов на мобильных устройств отмечают, что даже если задача локализации текста на изображении решается достаточно успешно, то задача распознавания самих текстовых символов на мобильных устройствах все еще требует существенного улучшения, принимая во внимание проблемы огра-

ниченных вычислительных ресурсов, необходимость распознавать изображени, захваченные камерами низкого качества или в сложных условиях, независимо от целевого алфавита, языка или стиля написания и т.п.

С появлением утверждений о неустойчивости алгоритмов сегментации на искаженных изображениях, стали появляться подходы без явной посимвольной сегментации. В наши дни большинство таких подходов основано на классификаторах со скользящим окном [120] и на рекуррентных нейронных сетях (РНС) [121], в частности – на LSTM (сети с блоками долгой краткосрочной памяти). Главное преимущество РНС – их возможность обрабатывать и запоминать последовательности, но оно же может стать недостатком, например, на строках без языковых моделей или же при применении сети к другому языку с той же письменностью. Другой подход к распознаванию без сегментации – распознавание слова целиком – имеет два существенных недостатка: число классов становится гигантским (по числу необходимых слов), а также его невозможно применить на текст с заранее неизвестным набором слов.

Методы распознавания строк зачастую подстраивают под особенности письменности. При этом, большинство исследований проводилось и все еще проводится для печатного текста исключительно с латинским алфавитом. В работе [122] изучили проблему распознавания нескольких алфавитов для бангла, арабского, телугу, непальского, ассамского, курумукхи и деванагари. Они использовали набор простых классификаторов, включая гистограмму ориентированных положений пикселей, случайные леса, k-ближайших соседей и несколько других подходов. Авторы работы [123] предложили метод DevNet, включающий в себя нейросетевую архитектуру на основе сверточной нейронной сети с пятью сверточными слоями для задачи распознавания рукописного деванагари. Работа [124] описывает высокопроизводительную архитектуру на основе сверточной нейронной сети, использующую глобальное средневзвешенное объединение выходных сигналов сети для расчета карт активаций классов, применяющуюся к задаче распознавания китайского рукописного текста. В работе [125] использовали сверточную нейронную сеть с самоадаптирующимся алгоритмом точной настройки выходного полносвязного слоя для распознавания рукописного тамильского письма.

Языки, основанные на арабской письменности, представляют отдельные трудности для распознавания из-за связного написания, а также разнообразия форм [126]. В частности, одно и то же слово, в зависимости от шрифта, может

быть написано как с лигатурами, так и без них. Одним из наиболее трудных для распознавания современных языков является урду в почерке насталик, используемый в Пакистане, в том числе – в документах, удостоверяющих личность.

Отдельной группой стоят языки, где количество классов для распознавания приводит либо к появлению очень тяжелых ИНС, либо к обработке неполного алфавита. К таким языкам, прежде всего, относятся китайский, японский и корейский. Для них предлагаются нейросетевые методы embedded learning [127]. Они, по-видимому, применимы и к алфавитно-слоговым письменностям Индии и Юго-Восточной Азии, в которых часто образуются лигатуры из нескольких символов.

В подобных задачах важной проблемой является недостаточный объем доступных данных для обучения глубоких нейросетевых моделей. Для экспериментов часто используются закрытые наборы данных [114]. Важной особенностью подходов к распознаванию сложных письменностей является перенос знаний (transfer learning), позволяющий уменьшить требуемые объемы обучающих выборок, повторно используя обученные слои в задачах классификации или сегментации изображений [123; 128; 129]. К примеру, в работе [129] использовли дообучение модели VGG-16 [32] в два этапа для распознавания рукописных деванагари и бангла.

Аугментация изображений, генеративно-состязательные сети (GAN) и автокодировщики также помогают работать с ограниченными наборами данных [129; 130]. Аугментация позволяет искусственно расширить наборы данных, используя такие операции, как транспозиция, отражение, вращение, сдвиг, масштабирование и т.п. для ограниченного набора входных изображений. Это помогает в обучении надежных классификаторов с ограниченным набором обучающих данных. Нейронные сети типа GAN используются для генерации новых синтетических данных, подобным реальным данным. Так, в работе [131] использовали GAN для генерации рукописных символов деванагари. Автокодировщики также представляют собой глубокие нейронные сети, которые используются для обучения компактному представлению данных, а также для создания синтетических данных. Подобный подход использовался в работе [130] для обучения сверточной нейронной сети для распознавания рукописных символов урду. Для повышения надежности и устойвости, в особенности в задачах распознавания рукописного текста, также используются гибридные подходы традиционного машинного обучения и глубокого обучения, как, к примеру,

совместное использование метода опорных векторов (Support Vector Machine, SVM) и сверточной нейронной сети для распознавания рукописного малаялламского текста в работе [132]. В работе [133] была предложена комплексная нейросетевая архитектура для распознавания рукописного текста HCR-Net (см. рис. 1.7), объединяющая в себе подходы с аугментацией и перенос знаний для задачи распознавания рукописных вариантом множества языков (бангла, пенджаби, хинди, урду, фарси, тибетского, каннада, малаялам, телугу, маратхи, непальского, арабского и др.).

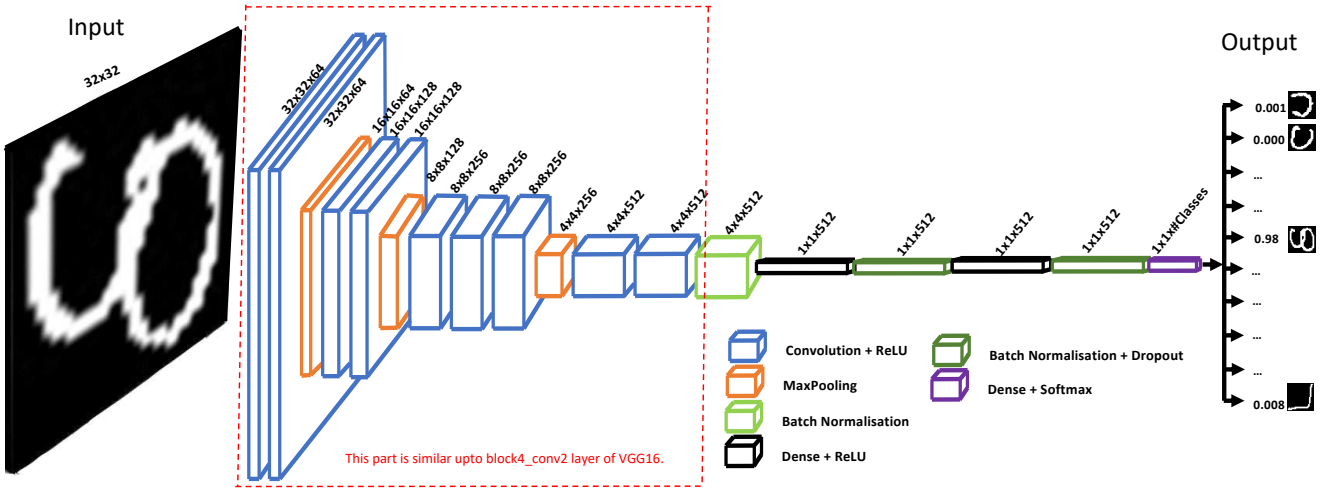


Рисунок 1.7 — Архитектура HCR-Net [133]

Существуют, однако некоторые известные общедоступные тестовые наборы для замера качества распознавания строк. Для точного замера качества желательно, чтобы либо датасет состоял из вырезанных изображений строк [134], либо в разметке присутствовала информация об окаймляющих прямоугольниках (или четырехугольниках) строк, иначе в замеры качества распознавания строк вмешиваются ошибки предыдущих подсистем, как при экспериментах с SmartDoc-2015. Вышеупомянутые наборы данных содержат примеры для латиницы. Для арабской письменности существует ряд наборов данных для распознавания строк, например – [135]. Для языков Индии и Юго-Восточной Азии также существуют отдельные наборы данных, но не все они выложены в открытый доступ [136]. В рамках International Conference on Document Analysis and Recognition (ICDAR) регулярно проходят различные соревнования, включающие в себя распознавания строк. В Таиланде регулярно проходит National Software Contest, где предоставляются наборы данных для тайского языка.

В заполнении печатных документов могут быть рукописные или рукопечатные строки. Подходы к их распознаванию в общем случае делятся на две

принципиально разные группы: онлайн и оффлайн подходы. В первом случае, текст распознается в момент его написания не только по изображению, но и по движению руки и порядку написания строки. Но такой режим не характерен для задач распознавания документов. Вторая же группа принципиально не отличается от подходов к распознаванию печатных строк. В ней для распознавания используются те же подходы, что были описаны выше. При этом, в распознавании рукописных текстов сегментация оказывается еще более сложной, а иногда и практически нерешаемой задачей.

1.4.3 Постобработка результатов распознавания

Результат распознавания текстового документа может содержать различные ошибки: неверное распознавание одного или нескольких символов слова, ошибки сегментации границ слов, пунктуационные ошибки. Точность распознавания может быть повышена при помощи пост-обработки результатов распознавания. Обычно алгоритмы пост-обработки включают два этапа: детектирование ошибок и дальнейшее их исправление. Чаще всего эти этапы тесно связаны между собой: детектирование предполагает обнаружение мест, в которых результат распознавания не удовлетворяет заложенным в алгоритм правилам, исправление – выбор наилучшего варианта, удовлетворяющего правилам.

Заложенные в алгоритмы пост-обработки правила обычно используют информацию о синтаксической и семантической структуре распознаваемых данных. Синтаксические правила описывают допустимые с точки зрения языка распознавания конструкции. Семантические правила основываются на смысловой интерпретации данных. Таким образом, с точки зрения языка ошибки можно разделить на те, которые приводят к появлению недопустимых для языка слов, и те, когда неверный результат распознавания допустим для языка, но противоречит грамматическим правилам или контексту.

Подходы к автоматической пост-обработке результатов распознавания могут работать как на уровне символов, основываясь на синтаксисе и семантике слов языка распознавания, так и на уровне слов, учитывая семантику связей распознаваемых данных. Также могут быть использованы смешанные подходы.

К методам автоматической коррекции, использующим семантические правила, можно отнести словарные методы [137]. Использующие контекст подходы чаще всего основываются на статистических языковых моделях, модели шумного канала [138], N-граммах [139]. Также могут применяться методы машинного обучения [139], использоваться внешние ресурсы, такие как, например, вызов поисковой системы Google с функцией проверки орфографии [140] или лексической базы данных WordNet [141]. Кроме того, существуют подходы, комбинирующие несколько методов пост-обработки для более точного учёта специфики распознаваемых данных.

Детекция ошибок распознавания при словарном подходе заключается в выяснении, входит ли распознанное слово в словарь допустимых значений. На этапе коррекции формируется список вариантов правильного распознавания, из которого потом выбирается вариант с наибольшей оценкой. Построение оценки соответствия распознанного слова s_r и словарного слова s_d может производиться, например, на основе расстояния Левенштейна или длины наибольшей общей подпоследовательности.

Словарные методы достаточно просты в реализации, однако имеют ряд ограничений. Во-первых, требуется составить словарь, покрывающий все возможные слова и формы слов, которые могут встретиться в распознаваемом тексте. Это существенно ограничивает применимость словарных методов для коррекции текстов естественного языка, особенно для языков со сложной морфологической структурой. Также при составлении словарей требуется иметь в виду временной аспект – возможность возникновения в тексте устаревших слов и словарных форм. Кроме того, словарные методы не подходят для исправления ошибок, когда неверный результат распознавания присутствует в словаре, однако не соответствует синтаксическим правилам или контексту.

Редакционное расстояние между словами может использоваться в сочетании с оценкой контекстного сходства двух именованных сущностей, рассчитываемого согласно некоторой языковой модели. Данный подход используется для кластеризации неверно распознанных слов и словосочетаний с целью дальнейшей коррекции ошибок распознавания [142].

Статистические N-граммные модели пост-обработки могут работать как на уровне символов [143], так и на уровне слов [139]. N-граммные модели основываются на предположении, что вероятность возникновения различных последовательностей символов или слов в распознаваемых данных разная.

Основные сложности применения N-грамм заключаются, во-первых, в составлении словаря N-грамм, то есть статистической модели, адекватно отражающей особенности распознаваемых данных. Для улучшения статистической модели могут применяться подходы машинного обучения, такие как метод опорных векторов [144]. Во-вторых, при больших значениях N сильно увеличивается трудоёмкость использования N-граммного подхода. На практике чаще всего используются N-граммы со значением N не превышающим 4.

К статистическим методам коррекции также относятся подходы, основанные на применении скрытых марковских моделей [145]. При этом используется предположение о том, что вероятность появления символа в слове или слова в тексте зависит от предыдущих символов или слов. Ограничивающим факторов использования статистических моделей является, во-первых то, что они зависимы от языка, во-вторых, не очень хорошо работают с текстами, которые могут содержать неканоническую орфографию, пунктуацию и т.п. Для решения задачи коррекции ошибок в подобных условиях могут применяться рекуррентные нейронные сети, в частности с использованием LSTM моделей [146].

Предложенный в [147] подход, основывающийся на взвешенных конечных преобразователях (Weighted Finite-State Transducers, WFST), соединяет в себе языковую модель, модель ошибок и модель гипотез. Подход на основе WFST позволяет построить алгоритм коррекции результатов распознавания, не заточенный изначально под специфику распознаваемых данных, а адаптируемый путём изменения семантических и синтаксических правил. Однако сложность использования этого подхода может заключаться в необходимости иметь информацию о распределении оценок принадлежности символов к классам распознавания (для модели гипотез), но зачастую этой информации нет на выходе распознавания. Ещё одним недостатком подхода является высокая трудоёмкость поиска оптимального пути в преобразователе в случае если кодируемый язык сложный. Кроме того, построение общей языковой модели в виде взвешенного конечного преобразователя в некоторых случаях также может быть достаточно сложным.

Более простым для расширения и имплементации, хоть и менее общим, является подход, при котором языковая модель представляется в виде проверяющей грамматики. В этом случае задача пост-обработки результатов распознавания сводится к задаче дискретной оптимизации. В [148] предлагается эффективный алгоритм её решения.

В настоящее время наблюдается тенденция к применению комплексных подходов к детекции и коррекции ошибок распознавания, многие из которых используют нейронные сети и методы машинного обучения для подстройки алгоритмов пост-обработки под распознаваемые данные. Вместе с тем остаётся актуальным использование больших корпусов данных, таких как Google Web 1T, Google Book и других, для построения языковых моделей. В силу широкого распространения технологий автоматического распознавания текстовых документов и, как результат, высокой вариативности данных большой интерес представляют подходы, обладающие свойством адаптивности, то есть дающие возможность легко подстроить алгоритм детекции и коррекции результатов распознавания без изменения модели целиком [148].

1.5 Применение методов анализа и распознавания изображений документов

Рассмотрим несколько сценариев применения технологий распознавания изображения документов.

1.5.1 Извлечение атрибутов

Наиболее востребованным сценарием обработки результатов распознания документов является извлечение атрибутов (текстовых или графических полей). Извлеченные данные передаются в систему документооборота. Например, излеченные атрибуты вместе с изображением могут сохраняться в электронном архиве.

Для поиска границ полей документов с известной геометрической структурой могут быть использованы методы сегментации текстов на основе описаний (шаблонов). Так, в работе [149] описано применение трехстрочного шаблона для выделения текстовых полей при распознавании банковских карт.

Для документов с более гибкой структурой можно применяются алгоритмы класса «text in the wild», не использующие априорной информации

о геометрии документа. Такие алгоритмы могут быть реализованы как с помощью искусственных нейронных сетей [150], так и классических методов обработки изображений [151], а также их комбинации.

В работе [152] поставлена задача сегментации составного объекта с известными ограничениями на взаимное расположение его элементов. В работе [153] рассматривается случай, когда граф ограничений является простой цепью, и с помощью динамического программирования проводится поиск границ полей для ID-документов и автомобильных номеров.

После нахождения границ поля проводится распознавание содержащихся в поле строк. Распознавание зависит от атрибутов поле, таких как алфавит, тип постобработки, признак печатного или рукописного поля. Для всех типов документов при нахождении таблицы из каждой ячейки таблицы извлекаются текстовое значение.

Для извлечения полей из изображений документа со сложной структурой, например, писем в произвольной форме, могут быть применены алгоритмы извлечения данных в распознанных с помощью OCR текстах. В качестве извлекаемых из текстов данных обычно выступают следующие объекты:

- значимый объект: имя персоналии, название компании и пр. для новостных сообщений, термин предметной области специального текста, ссылка на литературу для научно-технических документов и т. д.;
- атрибуты объекта, дополнительно характеризующие его, например, для компании это юридический адрес, телефон, имя руководителя и т.п.;
- отношение между объектами: к примеру, отношение «быть владельцем» связывает компанию и персону-владельца, «быть частью» соединяет факультет и университет;
- событие/факт, связывающее несколько объектов, например, событие «прошла встреча» включает участников встречи, а также место и время ее проведения.

Известны следующие основные способы извлечения данных:

- распознавание и извлечение именованных сущностей (named entities);
- выделение атрибутов (attributes) объектов и семантических отношений (relations) между ними;
- извлечение фактов и событий (events), охватывающих несколько их параметров (атрибутов).

1.5.2 Сравнение и проверка документов

Задача сравнения изображений документов становится актуальной при проверке корректности подписанных бумажных документов, например, при подписании двумя сторонами договоров и соглашений [154]. В этом случае требуется подробный анализ содержимого документа, так как изменение даже одного символа может стать критичным для оспаривания сделки.

Возможные изменения документа могут относиться как к содержанию документа, так и к оформлению документа (стиль шрифта, цвет и размер текста), пространственному расположению элементов (расстояние между строками). Возможно добавление и удаление элементов содержимого, включая текст, рисунки, графики, таблицы, рукописное заполнение (пометки, подписи), печати.

Одним из подходов к сравнению двух изображений документов является использование распознавания. Самым простым методом сравнения является распознавание текста с использованием OCR с последующим использованием программных утилит (к примеру, утилиты `diff`) на основе алгоритма нахождения наибольшей общей последовательности (LCS) для сравнения результатов распознавания двух документов [155]. Недостатками такого метода являются большое количество ложных срабатываний из-за ошибок распознавания и потеря информации о шрифте, цвете и размере текста. Также распознавание текста не может быть использовано для сравнения печатей, изображений и других нетекстовых элементов, присутствующих на изображениях документов.

Другим подходом, не использующим распознавание символов, является использование дескрипторов с предварительной сегментацией текста на строки. В работах [154] использовались dense SIFT дескрипторы. Также может использоваться сегментация изображения документа не только на строки, но и на символы, как предложено в работе [156]. Результаты, представленные в этой работе, показывают, что предложенный метод может обрабатывать изображения многоязычных документов с различным разрешением и размером шрифта.

Еще одним методом сравнения изображений документов, является использование визуальной схожести документов. В работе [157] для сравнения документов предлагается использовать визуальную меру схожести, для подсчета которой используется макет документа и характеристики текста, полученные из текстовых примитивов: сложность текстового блока, основанная на энтро-

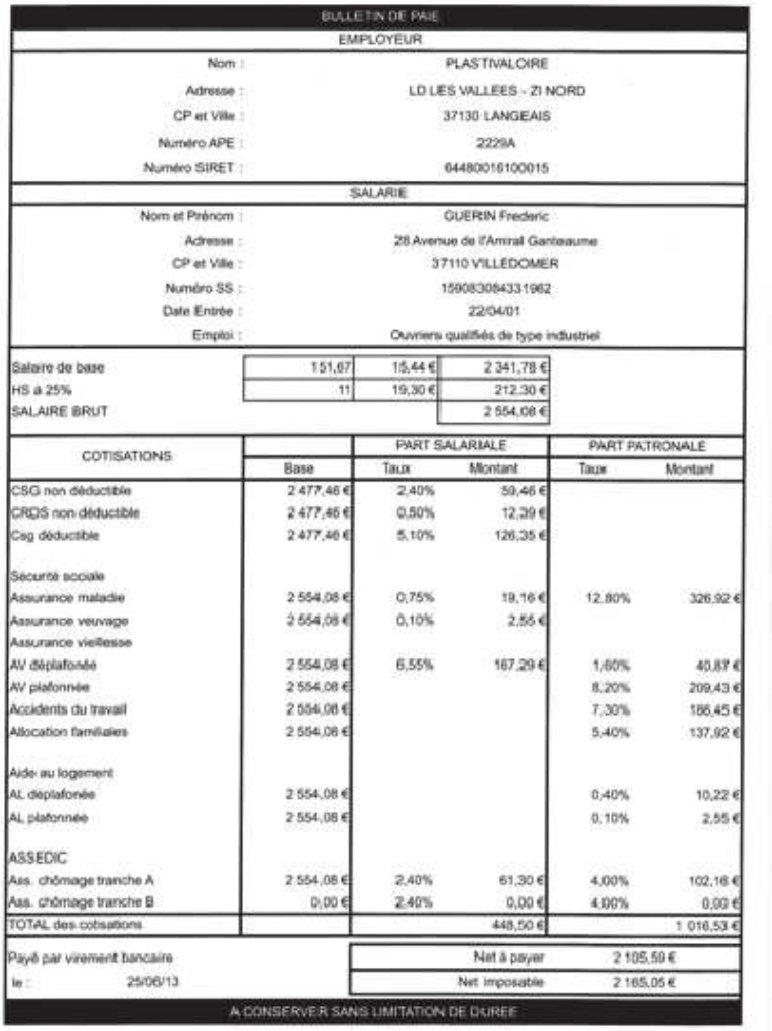
пии, и наглядность, имеющая дело с жирностью шрифта. Эта мера схожести изображений документов может использоваться для классификации документов и выдачи похожих документов при поиске.

Еще одной задачей сравнения документов является поиск дубликатов или почти дубликатов, например, при построении больших наборов данных, корпусов документов. В этом случае могут быть рассмотрены разные определения дубликатов. В некоторых случаях дубликатами могут называться версии документа, полученные в разных условиях [158]. Почти дубликатами могут считаться, например, изображения с одинаковым текстовым наполнением, но разными рукописными пометками [159].

Также к задаче сравнения изображений документов можно отнести определение того, является ли изображение подделкой по имеющему заранее набору данных [160; 161]. Рассматриваются изображения документов, поступающих из одного источника, например, счета из поликлиники, платежные ведомости, расчетные листы и т.д. Такие документы не содержат дополнительных средств защиты (таких как водяные знаки), но при этом имеют схожую структуру. По имеющемуся набору подлинных документов, который рассматривается как определенная обучающая выборка, выявляются признаки или особенности данных изображений документов. Далее на вход алгоритму поступает некоторое изображение и необходимо установить, является ли оно поддельным. Это осуществляется путем выявления определенных особенностей у входного изображения и сравнения их с теми, которые были выявлены у набора подлинных документов. Наиболее важным искажением, которое является в центре внимания работы [160], является неравномерное вертикальное масштабирование при повторной печати отсканированного документа. В работе [161] предложен подход для выравнивания документов, поступающих из одного источника.

Общей проблемой работ по обнаружению подделок является отсутствие публичного набора данных для оценки алгоритмов. Во-первых, естественно, что мошенники не хотят раскрывать свои методы и виды подделок, которые были ими сделаны. А во-вторых, большинство документов, которые подвергаются модификациям, содержат личную информацию и являются конфиденциальными. В [162] авторы решили эту проблему, синтезировав “реальные” документы, которые впоследствии были подделаны добровольцами. Представленный публичный набор данных, составлен из корпуса 477 подделанных (модифици-

рованных) платежных ведомостей, в которых было модифицировано около 6000 символов. Пример изображения из этого пакета данных представлен на рис. 1.8.



BULLETIN DE PAIE

EMPLOYEUR

Nom :	PLASTIVALOIRE
Adresse :	LD LES VALLEES - ZI NORD
CP et Ville :	37130 LANGEAIS
Numéro APE :	2229A
Numéro SIRET :	64480016100015

SALARIE

Nom et Prénom :	GUERIN Frederic
Adresse :	28 Avenue de l'Amiral Ganteaume
CP et Ville :	37110 VILLEDOMER
Numéro SS :	1590830843 3 1962
Date Entrée :	22/04/01
Emploi :	Ouvriers qualifiés de type industriel

Salaire de base	151,07	15,44 €	2 341,78 €
HS à 25%	11	19,30 €	212,30 €
SALAIRE BRUT			2 554,08 €

COTISATIONS		PART SALARIALE		PART PATRONALE	
	Base	Taux	Montant	Taux	Montant
CSG non déductible	2 477,46 €	2,40%	59,46 €		
CRDS non déductible	2 477,46 €	0,50%	12,29 €		
Csg déductible	2 477,46 €	5,10%	126,35 €		
Securité sociale					
Assurance maladie	2 554,08 €	0,75%	19,16 €	12,80%	326,92 €
Assurance veuvage	2 554,08 €	0,10%	2,55 €		
Assurance vieillesse					
AV déplafonnée	2 554,08 €	6,55%	167,29 €	1,60%	40,87 €
AV plafonnée	2 554,08 €			8,20%	209,43 €
Accidents du travail	2 554,08 €			7,30%	186,45 €
Allocation familiales	2 554,08 €			5,40%	137,92 €
Aide au logement					
AL deplafonée	2 554,08 €			0,40%	10,22 €
AL plafonnée	2 554,08 €			0,10%	2,55 €
ASSEDIC					
Ass. chômage tranche A	2 554,08 €	2,40%	61,30 €	4,00%	102,16 €
Ass. chômage tranche B	0,00 €	2,40%	0,00 €	4,00%	0,00 €
TOTAL des cotisations			448,50 €		1 016,53 €

Payé par virement bancaire	Net à payer	2 105,59 €
le : 25/06/13	Net imposable	2 165,05 €

A CONSERVER SANS LIMITATION DE DUREE

Рисунок 1.8 — Пример подделанного (модифицированного) изображения платежной ведомости [162].

1.5.3 Распознавание документов, удостоверяющих личность

Среди богатого разнообразия классов документов, удостоверение личности играет важнейшую роль, и для успешного применения систем распознавания документов общего назначения к такой задаче зачастую требуются значительные доработки и усовершенствования. Это связано с тем, что для успешного распознавания и анализа таких документов требуется учитывать

большое количество дополнительной информации о документе, при этом в целевых системах требования к качеству и скорости исполнения, как правило, выше, чем для документов других классов. Системы для распознавания удостоверений личности опираются как на классические методы компьютерного зрения и анализа изображений документов, так и на последние достижения в области машинного обучения и глубокого обучения.

Задача автоматического извлечения данных из изображений документов, удостоверяющих личность, стала предметом активного изучения начиная с 2000-х, в частности, чтобы обеспечить более эффективный ввод данных и проверку личной информации при регистрации в гостинице, посадке в самолет и т.д. [163]. Хотя изначально основными способами ввода были планшетные и специализированные сканеры, распознавание документов с помощью камер стало особенно актуально в течение последних 10 лет [164] благодаря широкому распространению портативных камер и мобильных устройств, например, смартфонов.

В течение последних лет был опубликован ряд работ, в которых описываются системы и подходы к распознаванию документов, удостоверяющих личность. В [163] рассматривается задача распознавания удостоверений личности на изображениях, полученных с помощью планшетного сканера. Документ обнаруживается и выравнивается с помощью преобразования Хафа, а определение типа документа выполняется с помощью классификации на основе цветовой гистограммы. Обнаружение текста осуществляется помощью анализа связанных компонентов, а распознавание текста – с помощью бинаризованных текстовых областей с последующей постобработкой используя геометрический и лингвистический контекст.

В работах [165], [166] и [167] описаны системы распознавания индонезийских удостоверений личности. Подход, описанный в [165], предназначен для документов, запечатленных с помощью камеры. Этапы обработки включают масштабирование, преобразование в оттенки серого и бинаризацию изображений документов, выделение областей текста с помощью анализа связанных компонентов, сегментацию текстовых строк на основе гистограмм и распознавание текста в шаблоне. В [166], символы индонезийских удостоверений личности распознавались с помощью CNN (сверточных нейронных сетей) и SVM (машины опорных векторов) с предварительной обработкой. Система, описанная в [167], включает сглаживание как один из этапов предварительной

обработки изображения, морфологические операции для обнаружения текстовых полей, а для распознавания текстовых строк используется Tesseract [119]. В [92] аналогичный подход описан для распознавания итальянских документов, удостоверяющих личность, с помощью камеры. Однако, в качестве этапа предварительной обработки для обнаружения и распознавания документов используется обнаружение и анализ вершин с классификацией типа документа с помощью CNN.

В работах [168] и [169] описаны системы обнаружения и распознавания текстовых полей вьетнамских документов, удостоверяющих личность. Этапы предварительной обработки в [169] включают преобразование в оттенки серого, коррекцию наклона, сглаживание и бинаризацию, а текстовые поля определяются отдельно: номер удостоверения личности детектируется на одной стороне документа, а анализ структуры таблицы проводится на другой стороне. Этап предварительной обработки изображения в [168] включает проективное преобразование на основе обнаруженных углов документа на изображении, полученном с помощью камеры, классификацию углов и геометрическую эвристику. В [168] используется SSD Mobilenet V2 [170] для обнаружения текста, а архитектура Attention OCR [171] – для распознавания текстовых строк.

В работах [172], [173] и [174] описаны системы распознавания документов, удостоверяющих личность, в частности, рассматривались китайские удостоверения личности, изображения которых получены с помощью камеры. Системы включают коррекцию наклона с помощью преобразования Хафа, предварительную обработку изображений документов – коррекцию яркости и преобразование в оттеки серого, обнаружение проекций и текста на основе морфологии и распознавание текста с помощью CNN [172; 173], либо на основе SVM и метода сравнения с эталоном [174]. Определение типа документа, описанное в [172], опирается на обнаружение национальных гербов с помощью признаков Хаара и детектора на основе алгоритма машинного обучения AdaBoost (Adaptive Boosting).

В работе [93] описана система анализа документов, удостоверяющих личность, эффективность которой была оценена в отношении колумбийских ID-карт. Целью описанного подхода является проверка подлинности удостоверения личности. Подход включает удаление фона с помощью глубокого обучения, определение углов и контуров для проективного преобразования, проверку яркости, гистограммы градаций серого, детекцию лица, связанные цветовые

компоненты и структурные маркеры для аутентификации (проверки подлинности) документа.

Важнейшей проблемой исследования новых методов и алгоритмов обработки документов, удостоверяющих личность, является наличие открытых датасетов. Поскольку документы, удостоверяющие личность, содержат персональную информацию, публичных датасетов, состоящих из реальных документов, не существует. Поэтому чтобы обеспечить воспроизводимость научных исследований в течение последних лет создаются и публикуются синтетические датасеты удостоверений личности. Так, существует Brazilian Identity Document Dataset (BID Dataset) [113], состоящий из изображений бразильских удостоверений личности с размытыми персональными данными, где заполнение полей синтезировано. Некоторые датасеты были созданы специально для задачи распознавания документов, удостоверяющих личность, например, база данных LRDE Identity Document Image Database (LRDE IDID) [175]. Более широкие датасеты для анализа документов в целом, например, SmartDoc family [176], также содержат примеры документов, удостоверяющих личность.

Хотя методы, используемые в отдельных этапах обработки, отличаются среди систем, в целом подходы и идеи довольно похожи. Типичная система распознавания документов, удостоверяющих личность, состоит из следующих этапов (см. рис. 1.9):

1. Предварительная обработка входного изображения. Этот этап включает общие шаги предварительной обработки изображения, такие как масштабирование или преобразование цветовой схемы, удаление фона, обнаружение контуров, краев и углов или семантическая сегментация.
2. Подготовка образа документа. Применяются геометрическая коррекция (ректификация – коррекция наклона, проекционные преобразования), коррекция яркости и т.д.
3. Извлечение текстовых полей и других важных элементов документов, удостоверяющих личность, таких как фотография лица.
4. Предварительная обработка текстовых полей (например, бинаризация и коррекция наклона), сегментация на символы, распознавание и постобработка с использованием языковых моделей.

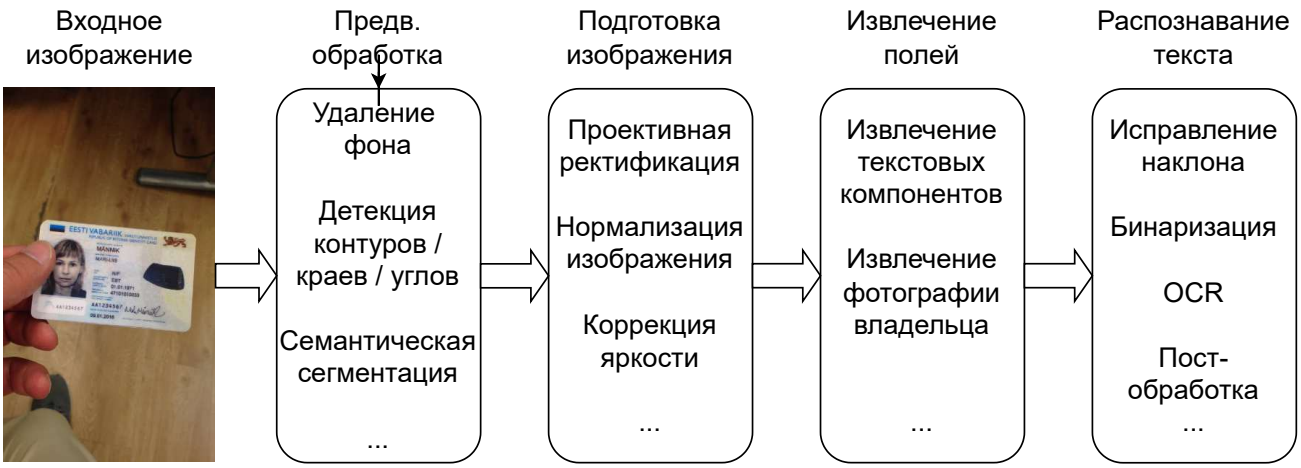


Рисунок 1.9 — Общая схема распознавания документов, удостоверяющих личность.

1.6 Дополнительные вопросы систем анализа и распознавания документов

1.6.1 Распознавание видеопоследовательностей

Помимо классических схем захвата входных изображений при помощи сканеров, в настоящее время являются крайне актуальными вопросы захвата изображений с помощью камеры смартфона или веб-камеры [177]. Несмотря на ряд недостатков, связанных с геометрическими или иными искажениями, преимущество мобильные камеры по сравнению со сканерами состоит в возможности захвата видеопотока, который может содержать кадры с разным освещением, под разными углами, с различными характеристиками фокусировки, что позволяет уменьшить спорадические ошибки OCR-систем.

При распознавании видеопотока возникает проблема выбора способов объединения информации, полученной из разных кадров видеопоследовательности. На сегодняшний день при распознавании видеопотока получили большую популярность методы выбора наиболее информативного кадра [178], методы «суперразрешения», создающие изображение более высокого качества на основе нескольких кадров низкого разрешения [179], методы отслеживания и комбинирования изображений распознанного объекта на последовательности кадров, методы компенсации размытия путем замены размытых областей в одном кадре их более четкими аналогами, взятыми из других кадров, или с использова-

нием методов глубокого обучения [180]. Также для лучшего восстановления распознанного изображения документа можно использовать данные, полученные от различных датчиков мобильного устройства, таких как, например, акселерометр или гироскоп. Однако для современных мобильных устройств погрешность в их измерениях может быть весьма значительной и препятствовать использованию этих данных для восстановления изображений. Недостатками методов первой группы, состоят в вычислительной сложности, чувствительности к геометрическим искажениям кадров и плохой масштабируемости в отношении видеопоследовательностей произвольной длины.

1.6.2 Оптимизация быстродействия алгоритмов распознавания

Распознавание документов основано на комбинации алгоритмов с различной математической сложностью. Быстродействие реализации распознавания на какой-либо вычислительной архитектуре должно быть оптимизировано по нескольким причинам. Во-первых, из-за экономии времени и других ресурсов, во-вторых, для улучшения эргономики пользовательских приложений. И, наконец, для платформ с ограниченными ресурсами, например, для устройств «Интернета вещей» (IoT) и мобильных устройств, оптимизация быстродействия непосредственно связана с оптимизацией энергопотребления и тепловыделения.

Одним из самых популярных механизмов классификации являются искусственные нейронные сети. ИНС могут использоваться практически на всех этапах распознавания документа. Модели ИНС предназначенные для классификации с высокой точностью, обычно требуют больших вычислительных ресурсов и энергии, так как часто ИНС основаны на большом числе операций умножения и большим числом параметров. Поэтому оптимизация быстродействия ИНС является актуальной задачей. Рассмотрим несколько популярных подходов оптимизации быстродействия ИНС.

Первая группа подходов основана на использовании альтернативных моделей нейрона. В работе [181] предложена аппаратная архитектура с использованием сверточной нейронной сети сумматора (AdderNet), в которой исходная свертка заменена ядром сумматора. Указывается, что AdderNet может достичь увеличения скорости на 16% и снижения энергопотребления на 47,85

– 77,9% по сравнению с аналогичной CNN традиционной архитектурой. Глубокая нейронная сеть (DNN) ShiftAddNet [182] выполняет умножение на основе суммирования и логическим сдвигом битов. В работе [183] описано семейство архитектур DeepShift, базирующееся на двух операциях: сверточные сдвиги и полносвязные сдвиги, которые заменяют все умножения вместе с побитовым сдвигом и заменой знака. Для представления весов требуется 6 битов. Показано сокращение времени задержки на 25% при выводе ResNet18 по сравнению с ядрами GPU на основе неоптимизированного умножения.

Вторая группа подходов использует целочисленную арифметику вместо вещественной [184]. При этом подходе повышается быстродействие и экономится занимаемая память при незначительном уменьшении точности. Впрочем, в работе [185] утверждается, что при использовании 4-битного детектора почти без потерь были достигнуты характеристики модели с полной точностью.

Третья группа подходов ориентирована на бинарные сети (BNN), в которых часть слоев содержит двоичные коэффициенты. Для достижения приемлемой точности при обучении BNN предлагается группа методов для избежания переполнения, такие как Normalization Layer Design, Small-pipeline Rule and Aggregated Convolutional Operation [186].

Применение малобитной вещественной арифметики (FP8) и малобитной гибридной арифметики (HFP8) позволяет успешно обучать DNN. В работе [187] демонстрируется возможность квантования предварительно обученной модели до 8-битного формата без потери точности.

Перспективным подходом является замена сверточного и полносвязного ядра тензорными приближениями низкого ранга. Такие методы как тензорная декомпозиция, тензорная последовательность, тензорное кольцо и модифицированная полиадическая тензорная декомпозиция показали эффективность сжатия при незначительном снижении точности [188].

Перечисленные подходы основаны на уменьшении квантования коэффициентов ИНС и на аппроксимации нелинейных функций. В дополнение к ним уместна низкоуровневая оптимизация для систем команд SIMD.

1.7 Выводы по главе

В первой главе показано, что задача распознавания документов, удостоверяющих личность, исследовалась в различных аспектах уже более сорока лет и продолжает оставаться актуальной по сей день. Большинство работ посвящено исследованию проблем распознавания документов, полученных с помощью сканирования, и только относительно небольшая часть новейших работ посвящена проблемам распознавания документов на фотографиях. Несмотря на наличие работ по отдельным аспектам и алгоритмам решающим узкие задачи, такие как: поиск и идентификация документа, поиск реквизитов документа, распознавание текста, выявление аномалий и изменений, в литературе не представлены работы по системному подходу к решению задачи распознавания документов в целом, с учетом всех особенностей, привнесенных повсеместным использованием мобильных телефонов для распознавания и неконтролируемости условий получения изображения, в отличии от множества работ по распознаванию документов со сканеров. Показано, что слабо изученными остаются возможности, которые предоставляет использование видеопотока, а также слабо рассмотренным остается вопрос ограниченности вычислительной и энергетической мощности мобильных устройств, используемых в качестве вычислительных платформ для распознавания, что ставит вопросы о применимости существующих алгоритмов без специальной оптимизации. Особую проблему в исследованиях распознавания документов, удостоверяющих личность, представляет практически полное отсутствие открытых и репрезентативных пакетов экспериментальных данных, которые бы позволили исследователям полноценно проводить исследования и верифицировать результаты.

Исходя из проведенного анализа современного состояния исследований в области распознавания документов, можно поставить следующие задачи:

1. Создать универсальный по источнику изображения подход к архитектуре систем распознавания документов, удостоверяющих личность, учитывающий особенности задачи, источников и способов захвата изображений, возможность анализа видеопотока и ограниченные ресурсы мобильных вычислительных платформ.

2. Изучить и предложить модель использования видеопотока с целью повышения качества решения задачи распознавания документов, удостоверяющих личность.
3. Предложить методы, модели, алгоритмы повышения качества распознавания документов, удостоверяющих личность, с использованием видеопотока.
4. Предложить и апробировать пути снижения вычислительной нагрузки в процессах распознавания документов, удостоверяющих личность.
5. Создать репрезентативные пакеты данных для исследования, замера и верификации результатов исследований в области распознавания документов, удостоверяющих личность, при этом учесть необходимость соблюдения авторских прав и ограничения, налагаемые на разглашение персональных данных.
6. Разработать методы анализа изображений (сканированных либо полученных путем фотографирования) и видеопоследовательностей документов, удостоверяющих личность, ставящие в качестве цели проверку подлинности документа и минимизации возможных фальсификаций процесса предъявления документа в автоматических системах ввода.

Глава 2. Распознавание и ввод идентификационных документов

2.1 Введение

В данной главе представлен анализ архитектуры систем распознавания документов, удостоверяющих личность. Будет показано, как трансформируются отдельные элементы и вся традиционная схема распознавания под влиянием особенностей природы документов, удостоверяющих личность, и способов получения их изображений на мобильных устройствах.

Классическая схема распознавания документов представлена на рисунке 2.1. Основные элементы и алгоритмы реализации этой схемы описаны в Главе 1. Реализация ее предполагает получение изображения в контролируемых условиях со сканера с известными оптическими и техническими характеристиками, что не свойственно для мобильных устройств. В этой главе рассмотрены особенности документов, удостоверяющих личность, особенности формирования изображений, оценка качества входных изображений, локализация и идентификация документов, поиск и распознавание текстовых полей. Кроме того, в настоящей главе сформулирована новая задача, связанная с процессом применения распознавания – проверка подлинности документа, удостоверяющего личность. В конце настоящей главы представлена универсальная схема распознавания документов, удостоверяющих личность, предназначенная как для обработки изображений со сканера, так и фотографий документов и видеопотока.

2.2 Документ, удостоверяющий личность: особенности и применение распознавания

Документом, удостоверяющим личность (удостоверяющим документом, ID-документом), считается любой документ, позволяющий идентифицировать личность его владельца. Набор информационных полей (как текстовых, так и графических) в таких документах соответствующий: различная персональная

Рисунок 2.1 — Классическая схема распознавания документов.

информация, относящаяся к владельцу документа. Наиболее часто на удостоверяющих документах указывают фамилию, имя и отчество, дату рождения, пол, государственный идентификационный номер, располагают фотографию владельца и оттиск подписи. Дополнительно документы могут содержать специальные кодифицированные зоны (машиночитаемые зоны, одномерные и двумерные штрихкоды). На рисунке 2.2 представлены примеры документов, удостоверяющих личность.

Рисунок 2.2 — Пример изображений документов, удостоверяющих личность: удостоверение личность гражданина Албании (образца 2004 года), удостоверение личности гражданина Франции (образца 2021 года), паспорт гражданина Сербии (образца 2008 года).

В целом, документы, удостоверяющие личность, могут иметь довольно простую структуру, особенно если они не содержат большого количества полей (например, служебные пропуска зачастую не обладают сложной структурой). Однако удостоверение личности государственного образца из соображений безопасности практически всегда оснащается элементами защиты (визуально контролируемыми, проявляющимися при облучении документа УФ, RFID-метками, микрочипами, содержащими биометрические данные, и т. п.) [189].

Хотя документы, удостоверяющие личность, могут быть выполнены в разном размере, чаще всего их изготавливают в виде идентификационной карты, которая по размерам совпадает с обычной банковской картой. Существует ряд международных стандартов, описывающих характеристики таких документов, например, ГОСТ Р ИСО/МЭК 7810-2015 [190] или ISO/IEC 7810:2003 [191], которые устанавливают требования к физическим характеристикам удостоверяющих документов. Такие стандарты стали необходимы в связи с повсеместным использованием удостоверений личности и направлены на унификацию характеристик удостоверяющих документов и упрощение процедур их обработки. Еще одним важным форматом документа, удостоверяющего личность, является паспорт или проездной документ. Паспорта обычно выдаются в виде буклета, часто снабженного встроенными микрочипами для машинного считывания.

Количество различных типов документов, удостоверяющих личность, выдаваемых в мире, естественно, очень велико. Собираются специальные базы данных с примерами документов, удостоверяющих личность, и некоторые из

них находятся в открытом доступе, например, Public Register of Authentic identity and travel Documents Online (PRADO) [192]. Однако там можно найти только часть всего разнообразия удостоверяющих документов. Более того, многие страны подразделяются на регионы или штаты, которым разрешено выпускать удостоверяющие документы по собственным стандартам. Например, каждый штат Соединенных Штатов Америки выдает специальное водительское удостоверение, и Американская ассоциация администраторов транспортных средств (American Association of Motor Vehicle Administrators, AAMVA) должна заботиться об их унификации и контроле таких водительских удостоверений [193]. Кроме того, удостоверяющие документы время от времени меняются в связи с обновлением дизайна и повышением требований к безопасности.

Системы автоматического распознавания удостоверяющих документов и ввода персональных данных применяются во многих сферах жизнедеятельности для решения конкретных прикладных задач. Первая группа таких задач связана с автоматизацией процесса идентификации личности клиента при обращении в различные офисы оказания услуг (например, в центры государственных услуг или финансовые организации). В данном случае оператор, обслуживающий клиента, лично контролирует принадлежность документа клиенту, валидность и подлинность удостоверяющего документа, а системы автоматического распознавания документов используются для ускорения обслуживания клиента (автоматического заполнения журналов посещения, подготовки договоров, заполнения анкет и т. п.).

Вторая группа актуальных задач, для которых необходимы автоматические системы ввода документов – это удаленная идентификация клиента. Возросший в последние время спрос на удаленные услуги, включающие в том числе необходимость качественной удаленной идентификации и верификации физического лица, определил требования к системам распознавания удостоверяющих документов. В данном случае, для успешного решения поставленной задачи оказывается недостаточным выделение текстовых реквизитов. Автоматизированные системы распознавания помимо традиционных функций должны обеспечивать возможность выделения и проверки признаков подлинности документа, контроля «живости» документа, а также возвращать все необходимые данные (фотографию держателя документа, подпись и т. п.) для дальнейшего использования биометрических систем.

Третья группа задач, в которой системы автоматического распознавания удостоверяющих документов играют критическую роль, связана с оказанием транспортных услуг. Так, для продажи билетов, регистрации пассажиров перед поездкой, обеспечения работы систем лояльности необходимо удаленно идентифицировать и верифицировать личность пассажира – задача, которая эффективно решается современными автоматическими системами распознавания удостоверяющих документов. С другой стороны, при пересечении границ производится детализированный досмотр и контроль пассажиров, при котором необходимо предъявление проездного документа (паспорта), что в современном мире также автоматизируется за счет применения автоматических систем сканирования, распознавания и проверки подлинности.

2.3 Особенности формирования изображений для цифровых фото и видео устройств

Рассмотрим использование фото- и видео камер для получения изображений документов, при этом будем использовать общий термин «съемка камерой» [194]. При такой съемке с помощью объектива на светочувствительной матрице лучами отраженного от объекта съемки света формируется изображение документа (см. рисунок 2.3).

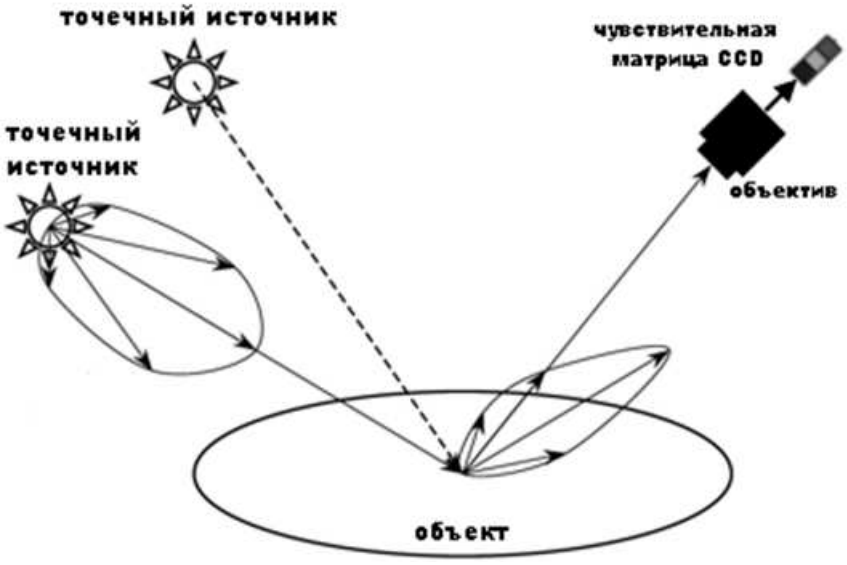


Рисунок 2.3 — Принципиальная схема формирования изображения в фото- и видео камерах.

По сравнению со сканерами, оптическая схема камеры является более сложной и сама по себе вносит больше искажений вследствие аберраций, бликов и отражений внутри оптической системы.

Использование фотосенсоров (матриц) и аналоговой электроники устройствами для регистрации изображений неизбежно приводит к появлению искажений изображений, которые называют цифровым шумом [195]. Источниками цифрового шума являются сам процесс оцифровки аналогового сигнала (ошибки квантования сигнала, тепловой шум и перенос заряда на матрице) и его дальнейшее усиление. Дополнительными источниками искажений являются загрязнения матрицы и дефектные элементы сенсора. Еще один источник искажений – алгоритмы сжатия изображений, что особенно заметно для кадров видеопотока [196; 197]. Цифровой шум заметен на изображении в виде наложенной маски из пикселей случайного цвета и яркости. Шум более заметен на однотонных участках изображения, в особенности – на темных. В сканерах гарантировано качественное освещение, в отличие от камер, для которых часто возникает ситуация недостаточной освещенности и влияние цифрового шума (см. рисунок 2.4) многократно усиливается.

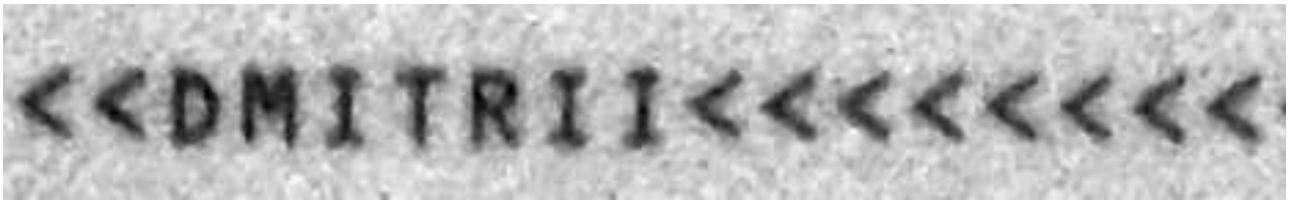


Рисунок 2.4 — Цифровой шум в условиях низкой освещенности.

Рассмотрим разрешающую способность различных устройств формирования изображений документа. Типичным разрешением для систем оптического ввода документов является 300 точек на дюйм (DPI). Такое разрешение обеспечивается всеми современными сканерами и позволяет качественно распознавать печатный текст с обычными для печатных документов размерами шрифтов. При таком разрешении изображение страницы формата A4 имеет размеры 3394×2400 пикселей. Характерное разрешение для камеры мобильного устройства в фоторежиме (5 мегапикселей) дает размер кадра 2592×1936 пикселей. Таким образом при фотосъемке максимальное рабочее разрешение не превышает 200 DPI (для формата A4 – 2262×1600 пикселей). Характерное разрешение web-камеры среднего уровня определяется режимом съемки FullHD с размером

кадра 1920×1080 пикселей, что дает еще более низкое рабочее разрешение, которое можно оценить как примерно 120 DPI (для формата A4 – 1358×960 пикселей).

В отличие от сканеров, при съемке камерой сам документ расположен в произвольной плоскости относительно плоскости сфокусированного изображения. Отклонение от перпендикулярной оптической оси плоскости приводит к проективному искажению (см. рисунок 2.5) изображения документа [198].

Рисунок 2.5 — Примеры слабого (слева) и сильного (справа) проективного искажения.

Предположения о выровненности, параллельности и перпендикулярности различных частей документа, например, изображений линий разграфки или строк текста часто используется в алгоритмах оптического распознавания текста [6; 199]. Такого рода свойства справедливы для исходных документов и выполняются для гомотетичных исходному документу изображений, получаемых при сканировании. Очевидно, что при съемке камерой углы и их отношения, а также пропорции объектов изменяются в зависимости от ракурса съемки. Это приводит к тому, что классические алгоритмы не могут применяться напрямую, а требуют предварительной проективной нормализации изображения.

Проективная нормализация основана на том, что съемка камерой хорошо моделируется преобразованием центральной проекции. Матрицу преобразования плоскости оригинального документа на плоскость изображения можно легко восстановить исходя из знания точного соответствия между документом

и его изображением всего по 4-м точкам. А из гомологии, при условии что найденные точки составляют прямоугольник, становится возможным определить отношение сторон оригинального документа. В большинстве случаев решается задача определения положения документа на изображении, а в качестве особых точек выбираются углы документа – таким образом одновременно находятся параметры проективной нормализации и границы документа, в которых ее нужно проводить.

В сканере фиксированы взаимное расположение документа и регистрирующей подсистемы, поэтому известен коэффициент подобия, а разрешение выходного изображения дает возможность вычислить точные размеры для различных элементов изображения. При съемке камерой коэффициент подобия с исходным документом неизвестен, поскольку определяется оптической схемой и взаимным расположением камеры и документа.

В зависимости от характеристик объектива и положения документа относительно плоскости наводки на резкость часть или все изображение документа может быть «размыто» (см. рисунок 2.6, слева). Подвижность самого документа или камеры, особенно в условиях недостаточной освещенности, приводит к ухудшению качества изображения – появляется «смазывание» (см. рисунок 2.6, справа) из-за смещения объекта во время экспозиции. С точки зрения «неподвижности» камеры и документа возможны следующие варианты:

- устройство и документ неподвижны (например, документ кладется на стол, над которым жестко зафиксировано регистрирующее устройство);
- устройство в руках, документ неподвижен (например, документ лежит на столе или закреплен на стене, съемка производится мобильный устройством «с рук»);
- устройство неподвижно закреплено, документ в руках;
- устройство и документ подвижны.

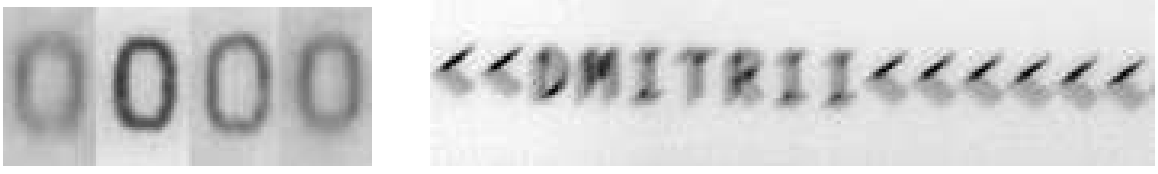


Рисунок 2.6 — Примеры дефокусированных символов (слева) и эффекта «смазывания» из-за смещения объекта (справа).

Как объект физического мира исходный документ подвержен механическим деформациям. Для обеспечения его «плоскости» при сканировании

используются прижимная крышка или ролики. При съемке камерой обеспечить отсутствие деформаций такого рода сложнее или просто невозможно. На изображении деформированного документа строки «изгибаются» (см. рисунок 2.7).

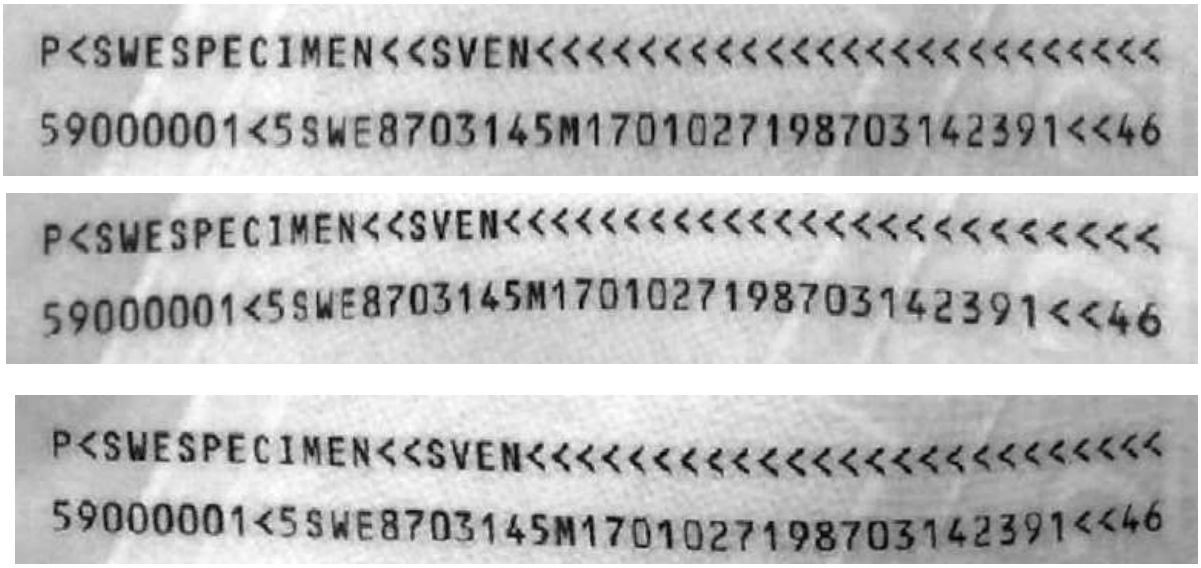


Рисунок 2.7 — Различные варианты изображения деформирования документов.

Механическая деформация исходного документа комбинируется с проективным искажением изображения. Выровненные в параллельные строки на исходном документе символы на изображении даже после проективной нормализации не имеют базовых линий. Более того, искажению подвергаются не только сами строки, но и изображения отдельных символов. То есть даже после правильной проективной нормализации изображение символа из деформированной на исходном документе области будет отличаться от изображения этого же символа из недеформированной области (см. рисунок 2.8).

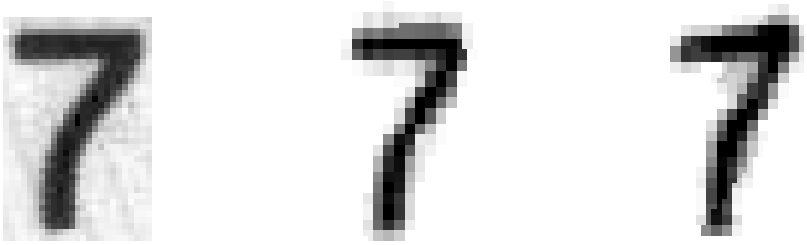


Рисунок 2.8 — Пример искажения изображения отдельного символа при механической деформации исходного документа (слева направо: отсканированный символ, символ при слабом изгибе, символ при сильном изгибе).

В силу указанных выше особенностей формирования изображения при сканировании, если зона документа занимает не все изображение, то либо края документа легко детектируются, либо заполнение изображения вне зоны документа не оказывает влияния на алгоритмы распознавания. При съемке на

камеру ориентированный неизвестным образом в пространстве документ расположен на фоне произвольных объектов и может быть освещен различным числом источников освещения с неизвестными геометрическими, цветовыми и яркостными характеристиками. При этом оригинальный документ в общем случае не обязан иметь строго прямоугольную форму – он может иметь скругленные, загнутые или обрезанные углы, дефекты прямых краев и деформации в плоскости документа. Сцена, в которой происходит съемка, может содержать в себе множество примитивов, не отличимых от границ документа, и быть похожей по цвету или текстуре на заполнение документа (см. рисунок 2.9). В таких условиях нахождение зоны документа и выделение текстовых строк само по себе становится довольно сложной задачей.

Рисунок 2.9 — Различные варианты организации сцены с точки зрения фона.

Схема освещения документа в сканерах минимизирует появление теней и бликов даже для «глянцевых» страниц документов. При съемке камерой в естественных сценах на изображениях часто возникают перепады яркости (тени, отражения и т.д.) и цветовые искажения, которые усложняют задачи анализа изображений и распознавания, например, за счет потери существующих или появления фальшивых границ объектов (см. рисунок 2.10).

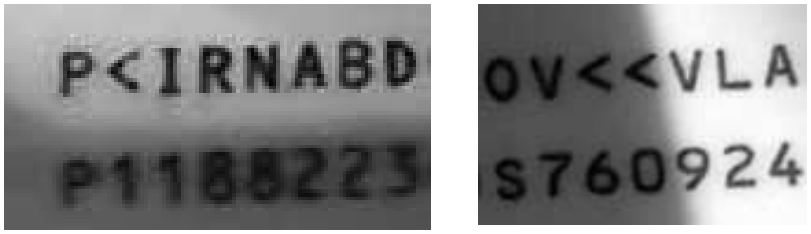


Рисунок 2.10 — Изображения документов при наличии затенения и перепада цветов.

Дополнительно элементы защиты [200] документа часто содержат области с «голографическими» элементами (см. рисунок 2.11 в центре и справа), которые также искажают изображение.

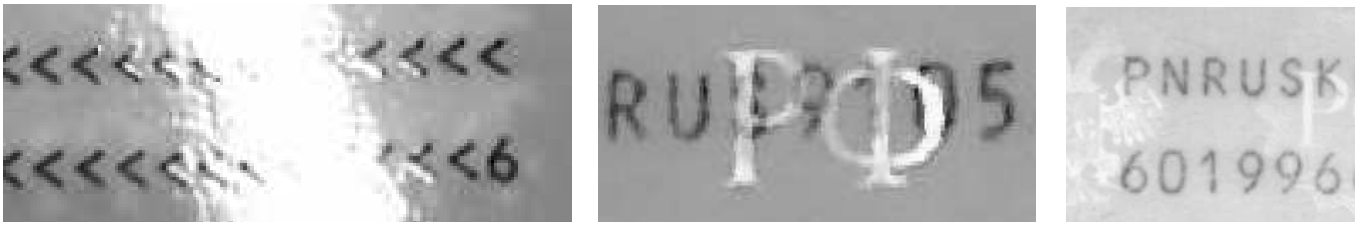


Рисунок 2.11 — Фрагменты зон документа с бликом от протяженного источника света (лампа дневного света, слева), и с голографической защитой (в центре и справа).

Стоит отметить, что физический процесс формирования изображения документа для малоразмерной цифровой видеокамеры аналогичен формированию фотоизображения. Необходимое для оптического распознавания разрешение изображения можно считать высоким с точки зрения технологий видеосъемки, что приводит к необходимости передачи большого потока информации. Последнее решается использованием аппаратных и программных средств сжатия потока кадров [196]. Таким образом, при использовании в задаче оптического распознавания изображения документа в качестве источника самого изображения видеокамеры необходимо учитывать еще более низкое, по сравнению с использованием сканера или фотоустройства, разрешение отдельного кадра. При этом каждый отдельный кадр изображения подвергается упаковке/распаковке и может содержать артефакты использования алгоритмов сжатия с потерями. Одновременно с этим использование видеопотока позволет использовать для распознавания последовательность изображений, на которых документ может располагаться под разными углами, в разных условиях освещения и при разных условиях фокусировки. Этим вопросам будет уделено отдельное внимание в Главе 3.

Рассмотрим основные задачи, которые должны быть решены для получения высокого качества результатов распознавания при создании систем оптического ввода документов, удостоверяющих личность, с использованием стационарных и мобильных малоразмерных цифровых видео камер в качестве источника изображения.

В первую очередь для целевых документов необходимо построить модель механических деформации исходных документов, провести анализ влияния та-

ких деформаций на распознавание и выбирать некоторые уровни «хороших», «допустимых» и «неприемлемых» деформаций. Если документы исполняются на плотном пластике, то страницы практически не деформируются, поэтому возможными малыми изгибами можно просто пренебречь. Выполненные на бумаге документы подвержены «изгибам» и «скручиванию» (чаще всего вдоль или поперек основного направления чтения), причем иногда возникают «волны», когда изгибы разнонаправленны в разных местах страницы. Отдельный важный и довольно сложный случай возникает при распознавании разворота документа, когда между двумя страницами есть естественный сгиб, а сами страницы могут быть по-разному ориентированы в пространстве.

Различные варианты организации сцены и процесса съемки очень существенно влияют на качество изображений, поэтому обязательным этапом является анализ процесса съемки. По результатам такого анализа определяются и формализуются границы применимости разрабатываемой технологии. Критическими пунктами являются:

- наличие и степень влияния на изображение особенностей оптической системы и регистрирующей подсистем (например, дисторсия, различные шумы, сжатие изображений и т.д.);
- расположение и крепление документа, камеры и источников освещения (в том числе допустимые диапазоны углов отклонений документа от плоскости фокусировки, диапазон расстояний от камеры до документа, скорость изменения различных характеристик);
- степень контролируемости фона и освещения.

Помимо общих вопросов желательно исследовать особенности конкретных моделей камер, например качество и скорость автофокусировки, цветопередачу и т.д.

2.4 Особенности систем распознавания изображений документов, полученных с фото и видео устройств

2.4.1 Сложности распознавания изображений

Хотя изначально основными способами ввода были планшетные и специализированные сканеры, распознавание документов с помощью камер стало особенно актуально в течение последних 10 лет благодаря широкому распространению портативных камер и мобильных устройств, например, смартфонов.

В предыдущих разделах мы обсудили особенности получения изображений с фотокамер и камер мобильных телефонов. Сложности при обработке фотографий по сравнению со сканами довольно многочислены. Во-первых, фон в случае сканированного изображения, как правило, однороден, в то время как фотография документа может быть сделана на произвольном фоне (см. рисунок 2.12). Разнородный и неконтролируемый фон может оказаться препятствием для точного детектирования и локализации документа, особенно если фон со структурной точки зрения слишком сложен, имеет много высококонтрастных линий или локальных областей или содержит текст, который используемый алгоритм распознавания может ложно «признать» за часть документа. Вторая существенная трудность распознавания фотографий – это неконтролируемые условия освещения. Изображения, полученные со сканеров, всегда равномерно освещены, в то время как фотография может быть получена при слабом и неравномерном освещении, быть пере- или недоэкспозированной (выполненной с недостаточной экспозицией). Неравномерное освещение создает проблемы для детектирования и локализации документа на фотографии, а также для анализа структуры документа, распознавания текста и других компонентов. Кроме того, серьезной проблемой является расфокусировка или наличие размытия.

Но самым важным различием между изображениями, полученными с помощью сканеров и камер, является геометрическое положение документа на изображении. На изображении, которое получено с помощью веб-камеры или камеры мобильного устройства, документ может быть повернут на все три угла Эйлера по отношению к оптической системе камеры. Если камера рассматривается в рамках модели камеры-обскуры, то семейство возможных

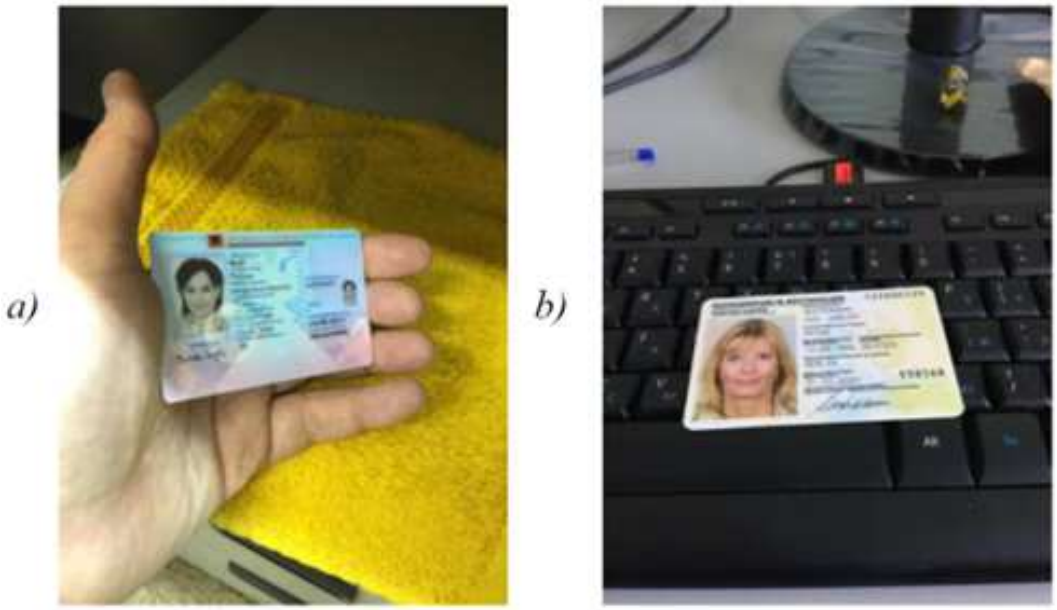


Рисунок 2.12 — Примеры фотографий удостоверений личности, сделанных с помощью смартфона.

геометрических преобразований документа теперь является подмножеством проективных, что, как было показано ранее, значительно усложняет задачу предварительной локализации документа. Более того, на одном изображении возможно несколько проективных преобразований для разных частей документа, как в случае захвата обеих страниц сброшюрованного документа, например, паспорта (см. рисунок 2.13a). Поскольку параметры объектива камеры могут быть неизвестны, изображения документов также могут быть подвержены радиальным искажениям. Наконец, если сам документ не является жестким, он может быть подвержен физической деформации, например, возможен изгиб бумажных страниц паспорта (см. рисунок 2.13b).

Помимо фотографий, использование веб-камер и мобильных устройств для получения изображений документов привело к появлению другого вида входных данных — последовательности видеокадров вместо одной фотографии. Использование видеопоследовательности вместо фотографии позволяет лучше контролировать процесс распознавания документов, а также делает входные данные менее подверженными подделке, поскольку видео сложнее подделать по сравнению с единичным загружаемым изображением.

С точки зрения распознавания и анализа документов, использование нескольких входных изображений одного и того же объекта дает дополнительные преимущества: оказывается возможным применять методы фильтрации и уточнения для повышения точности обнаружения и локализации объекта,

Рисунок 2.13 — Примеры фотографий внутреннего российского паспорта: a) книжный разворот, b) изогнутая страница.

использовать так называемые методы «сверхразрешения» для получения изображений более высокого качества, а также улучшить результаты распознавания текста путем накопления покадровых результатов распознавания и их интеграции в один наиболее достоверный. На рисунке 2.14 представлены примеры видеокадров, снятых для документа, удостоверяющего личность.

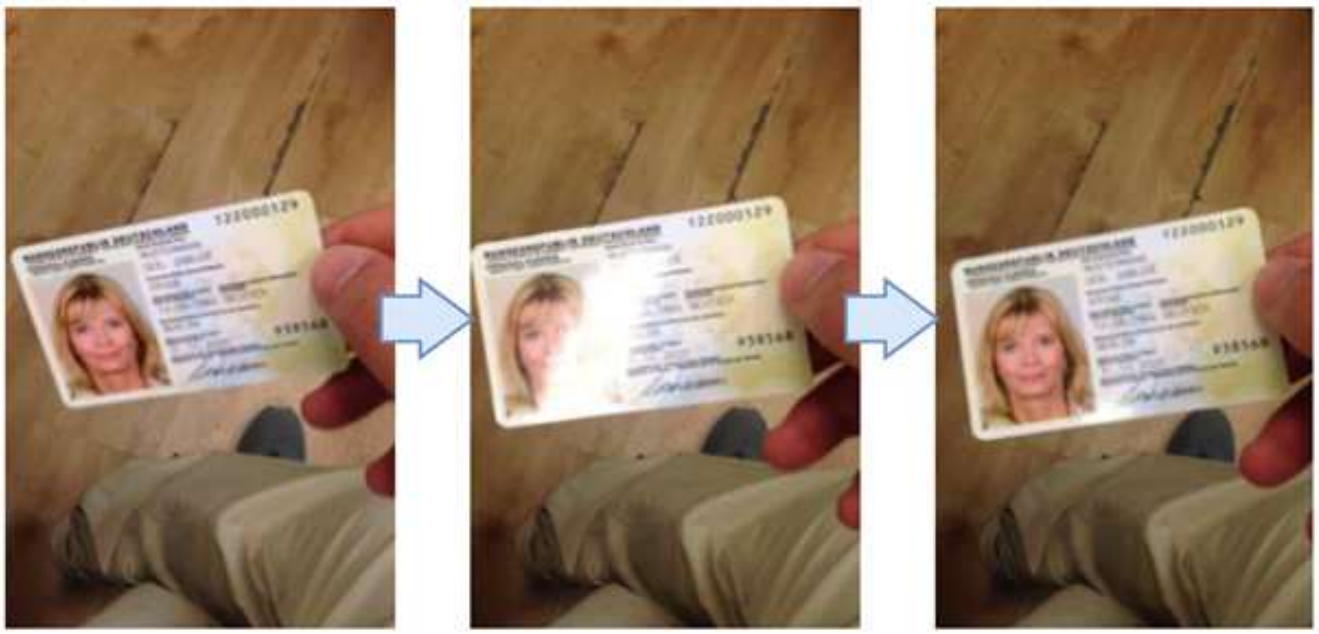


Рисунок 2.14 — Примеры видеокадров полученных для удостоверения личности, где сцена изменяется от кадра к кадру.

В современном мире практически интересны оказываются только такие системы распознавания удостоверяющих документов, которые получают на вход различные по своей природе данные.

2.4.2 Оценка качества изображения

Мы видели, что входные изображения систем оптического распознавания могут подвергаться большому количеству различных искажений, особенно в неконтролируемых или естественных условиях съемки: дефокусировка, смаз, цифровой шум, разладка регистратора, артефакты сжатия, блики, проективные искажения, повреждения объекта обработки и т.д. Проблема заключается в том, что при искаженных входных данных поведение систем распознавания не всегда предсказуемо, из-за чего невозможно корректное применение методов определения достоверности результатов распознавания. Следовательно, для построения систем распознавания заданной надежности необходима разработка методов контроля качества входных изображений.

Введем определения оценки качества изображений, необходимые для построения модели их внедрения в системы распознавания. Объективные методы оценки качества изображений могут быть классифицированы по доступности исходного изображения. Кроме этого, результатом метода может являться как скалярная оценка, так и изображение или другой объект, задающий положение и степень проявления различных искажений.

Обозначим за $\mathcal{I}$ множество входных изображений для рассматриваемого метода оценки Q, а за $\mathcal{D}$ – множество выходных оценок качества. Пусть имеется изображение $I \in \mathcal{I}$. Объективной функцией оценки качества в случае отсутствии информации об исходном изображении назовем функцию Q_{NR}, принимающую одиночное входное изображение:

$$Q_{NR} : \mathcal{I} \to \mathcal{D}. \tag{2.1}$$

При наличии исходного (не подвергаемого искажению) изображения $I^* \in \mathcal{I}$, соответствующий метод оценки Q_{FR} выглядит, как:

$$Q_{FR} : \mathcal{I} \times \mathcal{I} \to \mathcal{D}. \tag{2.2}$$

В отдельных случаях доступна лишь часть информации об исходном изображении, способствующая оценке степени его искажения. Обозначив за $\mathcal{P}$ множество указанной вспомогательной информации, получаем вид метода Q_{PR} с частичной информацией об исходном изображении:

$$Q_{PR} : \mathcal{I} \times \mathcal{P} \to \mathcal{D}. \tag{2.3}$$

Рассмотрим подробнее множество $\mathcal{D}$ выходов методов. В самом распространенном для общей оценки качества случае $\mathcal{D} = \mathcal{D}_{\mathbb{R}}$ представляет собой множество скалярных действительных чисел $\mathbb{R}$ с ограничением на[0,1], где 1 означает отсутствие повреждений, а 0 – наличие серьезных помех на изображении:

$$Q\left(I\right) = D \in \mathcal{D}_{\mathbb{R}}. \tag{2.4}$$

Однако в контексте применения в системах распознавания важно знать положение в пространстве и степень проявления соответствующих искажений на входном изображении. Данный эффект можно получить прямым расширением $\mathcal{D}$ до множества выходных изображений. Эти изображения могут представлять собой бинарные маски, задавая только положение искажения в пространстве:

$$Q\left(I\right) = D \in \mathcal{D}_B,\ D = \left\{d_{x,y} \in \{0,1\} | x,y : I_{x,y} \in I\ \right\}, \tag{2.5}$$

или карты вещественных чисел, оценивающие степень проявления искажений в каждой точке $I_{x,y}$, исходного изображения:

$$Q\left(I\right) = D \in \mathcal{D}_I,\ D = \left\{d_{x,y} \in \mathbb{R}_{[0,1]} | x,y : I_{x,y} \in I\ \right\}. \tag{2.6}$$

Для некоторых видов искажений информация о деградации в каждой точке исходного изображения может быть избыточной и неудобной для дальнейшей обработки, поэтому введем дополнительные типы возвращаемых значений методов оценки качества, такие, как компоненты связности для случая бинарных масок со степенью повреждения $q \in \mathcal{D}_R$ в данной компоненте:

$$Q\left(I\right) = D \in \mathcal{D}_C, D = \left\{(q_i, C_i)\right\}, C_i = \left\{(x,y)\,|x,y : I_{x,y} \in I\right\}, C_i \cap C_j = \varnothing, \tag{2.7}$$

или окаймляющие прямоугольники поврежденных областей, заданные координатами (x,y) верхней левой точки, шириной и высотой (w,h):

$$Q\left(I\right) = D \in \mathcal{D}_O, D = \left\{(q_i,\ \langle x_i, y_i, w_i, h_i\rangle)\right\}. \tag{2.8}$$

В дальнейшем под множеством будет подразумеваться одно из вышеперечисленных множеств:

$$\mathcal{D} \in \{\mathcal{D}_R,\mathcal{D}_B,\mathcal{D}_I,\mathcal{D}_C,\mathcal{D}_O\}\,, \tag{2.9}$$

в зависимости от конкретного приложения задающих информацию об искажениях.

Опишем процесс встраивания модулей оценки качества изображений в систему. Рассмотрим систему распознавания общего вида, состоящую из нескольких модулей. Ограничимся модулем S_i и теми выходами $r_{i,j}$, компоненты $v_{i,j}$ которых содержат данные из множества изображений, т.е. , $v_{i,j} \in \mathcal{I}$.

Пусть $Q_{i,j}$ – модуль оценки качества изображений для i,j-го выхода подсистемы S_i, принимающий на вход $r_{i,j}$. Данный модуль реализует одну из описанных в предыдущем разделе функций $Q\,(I)$ и возвращает оценки качества $q_{i,j} \in \mathcal{D}_{i,j}$, которые, как было обозначено ранее, являются пространственным распределением оценок качества для поступившего изображения $v_{i,j}$.

Введем модуль коррекции и принятия решений о дальнейшей обработке $\psi_{i,j}$, принимающий на вход как оценки качества $q_{i,j}$, так и результат $r_{i,j}$. Данный модуль возвращает модифицированные результаты $r^*_{i,j}$. Важной особенностью $\psi_{i,j}$ является возможность выдачи отказа в дальнейшей обработке, когда повреждение покрывает большую часть изображения $v_{i,j}$, и передачи сообщения об этом родительской подсистеме S_i посредством обратной связи.

Последним вводимым модулем в систему является опциональный модуль внимания или интереса A_i, соответствующий подсистеме S_i. Его задача – построить карту интереса $a_i\,(x,y)$ подсистемы S_i в участках обрабатываемого подсистемой объекта при условии его выделения на предыдущих этапах, за счет чего возможно контролировать приоритет обработки в видеопотоке.

На рисунке 2.15 проиллюстрирована модель графа подсистем и их связей после добавления модулей оценки качества изображений в процесс обработки.

2.5 Локализация и идентификация документов

Традиционно задача поиска и идентификации документа решалась на бинаризованном изображении с привлечением различных алгоритмов (описанных

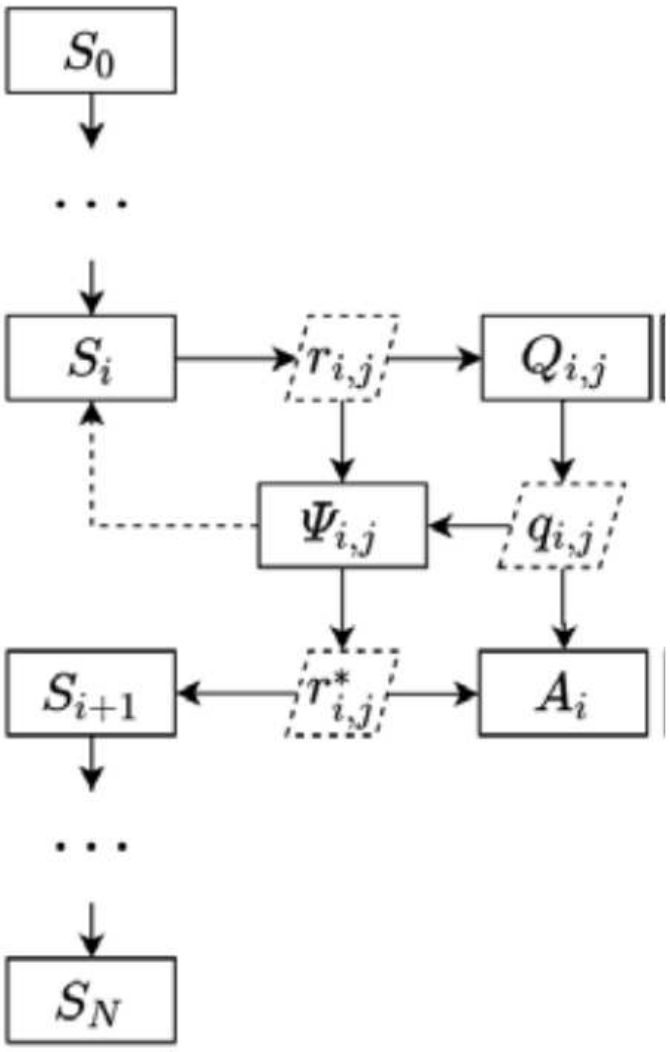


Рисунок 2.15 — Граф системы обработки с модулями оценки качества изображений.

в Главе 1), в основном базирующихся на наложении шаблонов или семантическом анализе результатов распознавания имеющихся текстов. Оба этих подхода слабо применимы к документам, удостоверяющим личность, в том числе за счет сложности получения анализируемого изображения. Неконтролируемые условия и сцена съемки делают качественную бинаризацию проблематичной. Кроме того, традиционные подходы локализации и идентификации документа обычно базируются на анализе хорошо различимых статических текстов. В задаче распознавания документов, удостоверяющих личность, как мы показали выше в настоящей главе, это далеко не так.

Таким образом необходимо применять другие подходы к решению этих задач, которые в современной научной школе принято называть «Lowe-based segmentation».

Обозначим множество известных типов документов как T. Поскольку некоторые информационные поля документа почти не различимы между собой иначе как по положению (например – даты рождения и выдачи), для успешной атрибуции полей мы должны не только решить задачу классификации изображения документа из множества T, но и привязать физическую систему координат этого типа документа к изображению. Для этого можно воспользоваться следующим приемом: для каждого типа из T заранее создать шаблон (образец изображения данного документа), а при распознавании последовательно решать задачу установления соответствия (корреспонденции) между входным изображением и каждым из образцов. При привязке изображений с мобильного телефона существенны три проблемы: геометрические искажения, неконтролируемое освещение и окклюзии. Даже если считать документ жестким и плоским, его образ подвержен проективному искажению. Распознаваемый документ освещен не так, как образец, причем, возможно, неравномерно и на нем могут быть блики. Наконец, документ может не попадать целиком в кадр или быть частично заслонен пальцами. К тому же, хотя рассматриваемые нами документы и изготавливаются с использованием бланка, содержащего графические элементы, общие для данного типа, они содержат и заполнение (текст, подписи, штампы, фотографии и др.), индивидуальное для каждого экземпляра, также «заслоняющее» бланк.

Для решения задачи установления соответствия в таких условиях хорошо себя зарекомендовал подход, базирующийся на сопоставлении локальных особенностей изображений [201]. Вместо того, чтобы рассматривать множество всех пикселей шаблона, на его изображении выделяются так называемые ключевые точки (особые точки), которые представляют собой достаточно уникальные устойчиво детектируемые области. В случае рассмотрения сдвиговой модели искажения относительно шаблона каждая ключевая точка фиксированного размера может быть задана двумя координатами. В аффинной модели точка помимо координат дополнительно кодируется направлением и масштабирующим коэффициентом. На данный момент известно множество алгоритмов детектирования ключевых точек (Harris, Shi-Tomasi, FAST, YAPE, SIFT и другие). Эти алгоритмы различаются как по быстродействию, так и по устойчивости детектора к различным классам искажений изображения. Главным критерием при выборе детектора в задаче распознавания идентификационных документов является соотношение скорости работы и вероятности выделения уникальных

особенностей при фиксированном числе детекций на изображении. Такому критерию отлично соответствует алгоритм YAPE.

Для каждого шаблона из T можно заранее вычислить его «созвездие» ключевых точек. При этом очевидно, что в такое созвездие должны входить только точки, принадлежащие бланку документа, а не его заполнению. По созвездию ключевых точек, выделенных уже на входном изображении документа, можно оценить соответствие каждому из шаблонов. Тогда задача классификации типа документа сводится к выбору наилучшего из полученных соответствий.

Для того, чтобы можно было при построении соответствия различать найденные ключевые точки, для каждой из них вычисляется так называемый дескриптор. По сути, он представляет собой вектор признаков локальной окрестности данной точки и позволяет сравнивать их между собой с помощью некоторой введенной метрики (обычно используется первая или вторая норма разности). Чем меньше вычисленное значение метрики, тем больше точки походят друг на друга.

На рисунке 2.16 приведена общая схема локализации и идентификации документа с применением дескрипторов особых точек.

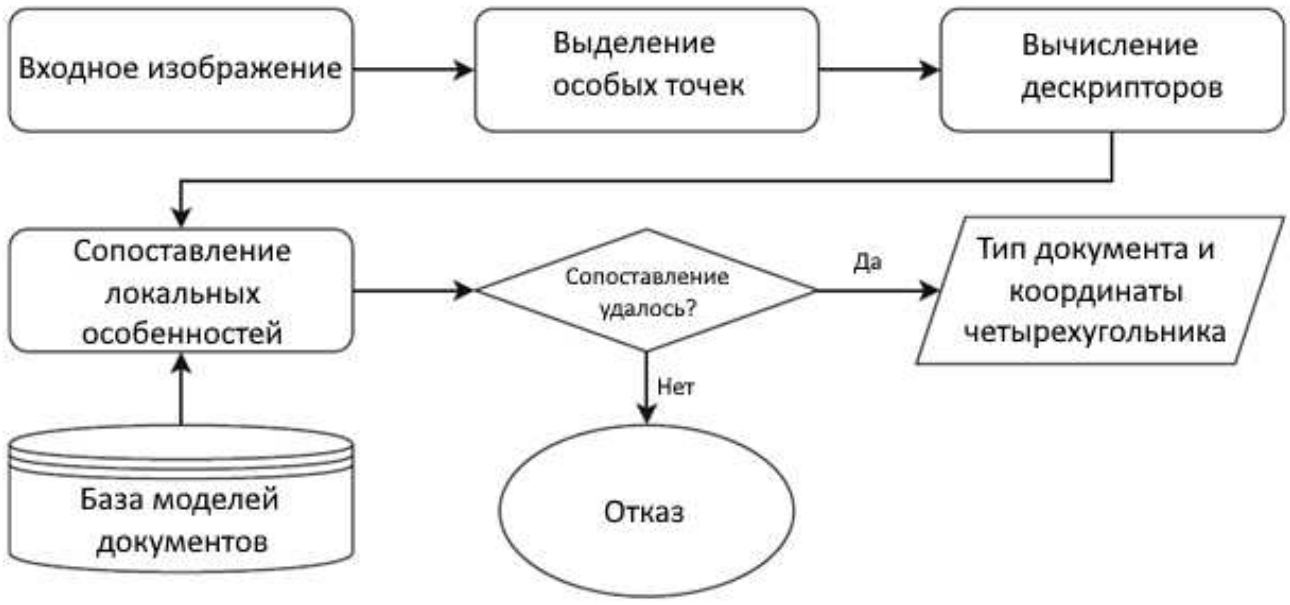


Рисунок 2.16 — Общая схема локализации и идентификации документа с применением дескрипторов особых точек.

Несмотря на то, что на сегодняшний день существует довольно большое количество различных способов построения дескрипторов, позволяющих

строить достаточно эффективные для целей описания регионов (патчей) изображений различной тематики, с точки зрения вычислений и потребления памяти эффективнее оказывается построение специального признакового описания патчей, характерных для изображений документов. В следующем разделе описано семейство дескрипторов, разработанное как раз для эффективного решения задачи сопоставления документов.

2.5.1 Дескрипторы для задачи локализации и классификации ID-документов

Как сказано выше, выбор правильного алгоритма построения дескриптора в задаче локализации и классификации документов, удостоверяющих личность, является крайне важным. Промышленные распознающие системы поддерживают распознавание нескольких тысяч различных удостоверяющих документов (в настоящее время во всем мире действует более 2000 различных удостоверяющих документов [202]). Если предположить, что каждый документ может быть представлен для целей локализации и идентификации порядка 300 точками с соответствующими дескрипторами, а каждый дескриптор требует 2048 бит = 256 байт (соответствует дескриптору SURF [78]), то получается, что только лишь для хранения такой описательной базы документов потребуется порядка 220 Мб. Кроме того, размер дескриптора кратно влияет на скорость вычислений при решении задачи локализации и идентификации.

Помимо известных и зарекомендовавших себя инженерных дескрипторов SIFT [203] и SURF [78], в настоящее время активно использую обучаемые дескрипторы: RFD [204], BEBLID [205], BinBoost [206]. По качественным и вычислительным характеристикам такие дескрипторы обходят инженерные дескрипторы. Однако важно отметить, что при обучении таких дескрипторов очень важно использовать правильный набор данных для обучения.

Дескриптор RFDoc – обучаемый дескриптор, построенный по принципу RFD [204], предназначенный специально для решения задач локализации и идентификации документов.

В качестве набора данных для обучения дескриптора были использованы следующие наборы данных документов, удостоверяющих личность:

MIDV-500 [207] и MIDV-2019 [208] (см. рисунок 2.17). В настоящей работе указанным наборам данных будет уделен отдельный раздел. Однако здесь важно отметить, что по своей структуре данные наборы данных содержат изображения муляжей удостоверяющих документов, полученные путем съемки на камеры мобильных устройств. Все изображения аннотированы (указан четырехугольник документа на каждом изображении, координаты статических объектов и данных заполнения и т. п.).

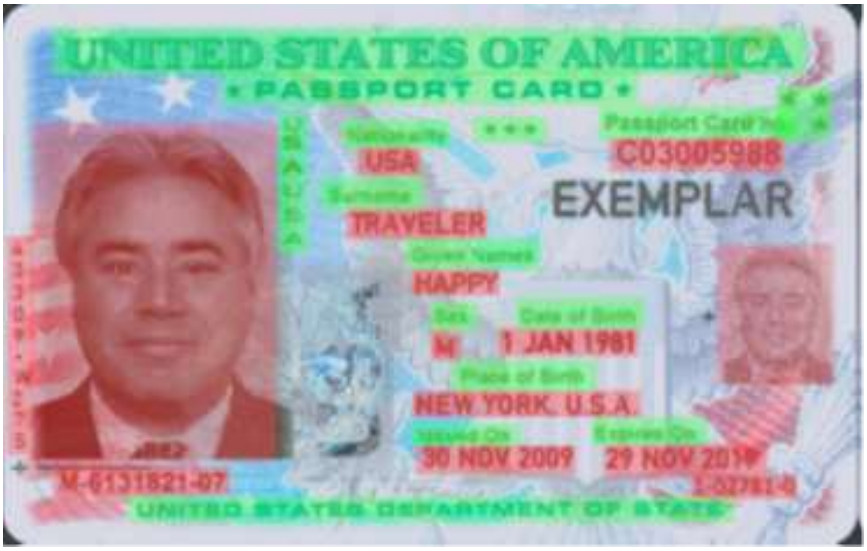


Рисунок 2.17 — Иллюстрация регионов статических элементов документа (выделены зеленым цветом) и переменных персональных данных (выделены красным цветом).

На рисунке 2.17 приведен пример изображения документа, на котором дополнительно отмечены зеленым и красным цветами области статических элементов документа, а также области, относящиеся к заполнению. Использование таких аннотированных изображений позволяет подготовить обучаемый набор – коллекцию положительных и отрицательных патчей, необходимую для обучения дескриптора (см. рисунок 2.18).

Рисунок 2.18 — Примеры положительных и отрицательных пар патчей, используемых при обучении дескриптора RFDoc.

Таблица 2 — Качество локализации и идентификации документов с использованием различных типов дескрипторов на наборе MIDV-500

№	Дескриптор	Качество классификации	Качество локализации	Размер дескриптора
1	RFDoc	93.458%	85.128%	192 бита
2	BEBLID-512 [205]	93.508%	85.226%	512 бит
3	BEBLID-256 [205]	92.783%	84.072%	256 бит
4	BinBoost-256 [206]	91.116%	81.132%	256 бит
5	BinBoost-128 [206]	85.958%	73.588%	128 бит
6	SURF [78]	91.241%	82.783%	2048 бит

Таблица 3 — Качество локализации и идентификации документов с использованием различных типов дескрипторов на наборе MIDV-2019

№	Дескриптор	Качество классификации	Качество локализации	Размер дескриптора
1	RFDoc	88.875%	75.535%	192 бита
2	BEBLID-512 [205]	85.854%	75.001%	512 бит
3	BEBLID-256 [205]	83.833%	72.368%	256 бит
4	BinBoost-256 [206]	79.916%	63.074%	256 бит
5	BinBoost-128 [206]	68.791%	50.262%	128 бит
6	SURF [78]	75.666%	61.542%	2048 бит

Дескриптор RFDoc обучался по принципу обучения дескриптора RFD, с дополнительным ограничением на размер дескриптора. В результате обучен бинарный дескриптор размером 192 бита.

Качество работы обученного детектора проверялось в рамках решения задачи локализации и идентификации удостоверяющих документов на наборах данных MIDV-500 и MIDV-2019. Таблицы 2 и 3 содержат количественные показатели работы системы распознавания при использовании различных дескрипторов.

Для оценки эффективности дескриптора RFDoc с точки зрения потребления памяти был выполнен следующий эксперимент. Для всех шаблонов документов из набора данных MIDV-500 были посчитаны особые точки и со-

Таблица 4 — Необходимое количество памяти для описания шаблонов всех документов набора данных MIDV-500

№	Дескриптор	Размер дескриптора	Необходимый объем памяти, Мб
1	RFDoc	192 бита	8.6
2	BEBLID-512 [205]	512 бит	22.7
3	BEBLID-256 [205]	256 бит	11.5
4	BinBoost-256 [206]	256 бит	11.5
5	BinBoost-128 [206]	128 бит	5.8
6	SURF [78]	2048 бит	82.0

ответствующие дескрипторы. Общий объем памяти, необходимый для такой операции, представлен в таблице 4.

2.5.2 Алгоритм выбора лучшего шаблона

Так как количество заготовленных шаблонов документов может быть велико (на сегодняшний день известно более 2000 уникальных шаблонов различных удостоверяющих документов [202]), то имеет смысл перед геометрическим сопоставлением шаблонов отбросить заведомо неподходящие. Это можно делать в модели «мешка дескрипторов», для чего требуется эффективный алгоритм поиска соответствия между найденным на изображении документа дескриптором и всеми локальными особенностями на всех шаблонах, которые после однократного предварительного вычисления можно запомнить в некоторой структуре. Для этого логично использовать поисковые деревья, позволяющие осуществлять быстрый приближенный поиск k наиболее близких точек по выбранной метрике. Это может быть, например, метод иерархической кластеризации. С его использованием все шаблоны сортируются по количеству успешно сопоставленных особенностей, после чего отбирается M лучших шаблонов-кандидатов и далее рассматриваются только они.

Известно, что искомый шаблон и входное изображение документа связаны некоторым проективным преобразованием. Будем оценивать его параметры для каждого из M отобранных шаблонов. Это позволяет выделить среди шаблонов

лучший (по некоторой метрике). При этом, среди найденных сопоставлений локальных особенностей нередко встречаются ложные, поэтому метод оценки параметров проективного преобразования должен быть робастным. Таким свойством обладает широко известная схема RANSAC [209]. В этом подходе устойчивость обеспечивается тем, что параметры преобразования оцениваются не по всем данным, а по максимальной самосогласованной подвыборке.

2.5.3 Идентификационные документы с «бедным» шаблоном

Существует класс документов, установление шаблона для которых методом ключевых точек весьма проблематично. Связано это либо с малым количеством потенциально детектируемых ключевых точек на шаблоне документа, либо с неуникальностью их дескрипторов. Последнее подразумевает ситуацию, когда внутриклассовая вариация для дескриптора ключевой точки на наборе изображений сравнима (а не пренебрежимо мала) с межклассовым расстоянием до ближайшего дескриптора среди прочих ключевых точек. Одной из причин такой ситуации является совпадение окрестностей ключевых точек. Типичным примером документа с «бедным» шаблоном является внутренний паспорт гражданина РФ, шаблон которого состоит из текстовых блоков, многократно повторяющихся горизонтальных линий и орнамента в виде густой сети волнистых фигурных линий (см. рисунок 2.19). Ясно, что такой шаблон не обеспечит достаточной для решения задачи установления соответствия с шаблоном уникальности дескрипторов ключевых точек.

На сегодняшний день традиционный способ установления соответствия шаблону для такого рода документов заключается в выделении на изображении графических примитивов (статических элементов), после чего выполняется их сопоставление с примитивами, заданными в шаблонах [13]. Однако такой подход обладает существенным недостатком: поиск статических элементов – сложная задача, требующая для решения либо использования методов машинного обучения [12], либо нетривиальной предварительной обработки изображения документа, которая сама по себе должна опираться на знание о типе обрабатываемого документа [98].

Рисунок 2.19 — Пример паспорта гражданина РФ.

На практике часто оказывается, что документы с «бедным» шаблоном обладают жесткой геометрией (различные образцы одного типа документа практически совмещаются «на просвет»). В частности, паспорт РФ удовлетворяет этому свойству. Для такого рода документов задача установления шаблона может быть переформулирована в терминах детектирования на изображении объекта заданной геометрии, которая может быть решена, например, методом Виолы и Джонса, сводящим задачу детекции к задаче бинарной классификации всевозможных подокон на изображении [97]. Показано, что этот алгоритм обеспечивает высокое качество детектирования документов (98% в реальной распознающей системе), удостоверяющих личность, как на сканах, так и на фотографиях [98]. Благодаря использованию признаков Хаара метод Виолы и Джонса показывает высокое быстродействие и робастность к незначительным вариациям образа документа.

Однако на практике обучение классификатора Виолы и Джонса достаточной точности часто оказывается нетривиальной задачей. В большинстве работ, рассматривающих вопрос обучения подобных классификаторов, приводятся алгоритмы обучения очередного уровня каскада при некоторой локальной целевой функции, но не обсуждаются проблемы обучения каскада в целом. При этом общеизвестные эвристические подходы (как, например, доведения до нуля числа

ложноотрицательных результатов на очередной подвыборке) не обеспечивают высокой точности для каскада в целом.

Нужную точность классификации с большим запасом обеспечивают нейросетевые модели (задача бинарной классификации окна по признаку содержания документа не сложнее задачи распознавания символа на произвольном фоне), но вызов нейросети для каждого подокна неприемлем с точки зрения производительности. С учетом того, что проективные искажения, как изложено выше, можно оценивать независимо по взаимному расположению пучков линий и строк, и детектор Виолы и Джонса работает в геометрической модели подобия, число подокон на изображении составляет 10^6–10^7. Приемлемое же число запусков нейросети можно оценить по числу символов на документе (10^2–10^3), из соображения, что времена работы привязки (подсистемы локализации документа) и распознавания должны быть сопоставимы. Поэтому следует использовать комбинированный классификатор, состоящий из последовательности «недоученного» каскада Виолы и Джонса (который имеет достаточную с точки зрения конечной целевой функции полноту, а точность – всего лишь порядка 1%) и двухклассовой нейронной сети [210]. Достижение такой точности означает сокращение числа подокон в 10000 раз. После этого среднее число вызовов нейросети, отнесенное к числу искомых объектов, составит 100, что разумно, с учетом того, что документ содержит порядка 100 знакомест.

Таким образом, метод Виолы и Джонса в паре с пост-классифицирующей нейронной сетью представляет собой мощный инструмент для поиска на изображениях документов жесткой геометрии с «бедным» шаблоном.

2.6 Поиск текстовых полей

Рассмотрим постановку задачи поиска границ текстовых полей документа. Исходный документ состоит из одной или нескольких зон, каждая их которых содержит текстовые поля, составляющие содержимое документа. Зона состоит из строк, каждая из которых состоит из одного или несколько текстовых полей. Поля в строке отделены друг от друга расстоянием, характерным для данного документа, более близкие текстовые поля считаются одним общим полем.

Количество строк и полей на документе может быть переменной величиной, сами текстовые поля выделены с помощью яркости и размеров. Требуется определить границы таких полей и сопоставить их соответствующим атрибутам документа. Рассмотрим решение задачи на примере национального паспорта РФ в качестве исходных данных. Для зоны данных страницы паспорта РФ (см. рисунок 2.20) необходимо определить положение следующих полей: ФИО, пол, дата и место рождения. Поле отчество может отсутствовать, количество строк полей места рождения является переменной величиной, от одной до трех (такие поля выделены на изображении более светлыми рамками).

Рисунок 2.20 — Поля зоны данных национального паспорта РФ.

Алгоритм решения поставленной задачи состоит из трех этапов:

1. Предобработка изображения
2. Выделение строк и текстовых полей
3. Сопоставление найденных полей шаблону зоны документа

2.6.1 Предобработка изображения

Основная идея предобработки заключается в преобразовании исходного изображения таким образом, чтобы представить искомые текстовые поля в ви-

де легко находимых и визуально определяемых объектов. Для решения этой задачи будем использовать морфологические операции. Основными морфологическими операциями являются дилатация и эрозия, которые выполняются для каждой точки изображения с вычислением максимальных или минимальных значений в её окрестности, заданной некоторым примитивом. В обычных условиях может использоваться алгоритм Ван Херка, который позволяет вычислять морфологические операции с прямоугольным примитивом за время, не зависящее от его размеров.

Алгоритм предобработки проиллюстрирован на рисуноке 2.21 и состоит из следующих шагов:

Рисунок 2.21 — Алгоритм предобработки изображения.

1. Текстовые поля документов в большинстве случаев представляют собой последовательные наборы букв, яркость которых отличается от яркости фона. Таким образом, первым шагом осуществляем переход от цветного изображения к полутоновому. Если фон имеет характерный цвет, то вместо усреднения каналов предпочтительно взять данные конкретного цветового канала для лучшего контрастирования фона от текста.
2. Производится масштабирование исходной зоны к заданному размеру шаблона зоны. Это делается по следующим соображениям: (а) исходные изображения имеют различные размеры, в зависимости от источников данных, поэтому требуется их унификация; (б) размер шаблона выбирается небольшим, но одновременно текстовые поля должны быть достаточно отличимыми от фона; (в) масштабирование позволяет сгладить особенности гильоша документа и преобразовать его в близкий к однородному фон.
3. Производится отделение фона от текста путём замыкания изображения шаблона зоны с примитивом небольшого размера $[4, 4]$, в результате

получаем почти однородное изображение, состоящее из усредненного цвета фона и сохраняющее особенности неравномерности освещения различных частей документа.

4. Инвертируем изображение фона.
5. Складываем полученное изображение фона с изображением шага 2 и получаем результат, где фон стал близок к белому, а текстовые поля остались контрастными.
6. Производим размыкание полученного изображения с примитивом размера $[s, 0]$, где s – приблизительная ширина символа. В результате такого преобразования получаем изображение, представляющее собой склеенные в компоненты текстовые поля на однородном фоне.
7. Произведем замыкание изображения с примитивом размера $[0, h/2]$, где h – приблизительная высота символа. Это преобразование уберёт выступы компонент, а также небольшие статические тексты и связи с ними. В результате получается выходное изображение, на котором текстовые поля представлены в виде визуально отличимых компонент на почти однородном фоне.
8. В ряде случаев, связанных с ошибками нахождения границ документа или проблемами при печати данных, текстовые поля документа могут иметь значительные углы наклона, что затрудняет их поиск и последующее распознавание. Для определения угла наклона и коррекции полей используется быстрое преобразование Хафа, после чего осуществляем поворот изображения.

На рисунке 2.22 проиллюстрированы результаты предобработки изображения по этапам.

2.6.2 Выделение строк и текстовых полей

Рассмотрим алгоритм выделения текстовых строк и полей, в качестве исходных данных получающий результат предобработки изображения с предыдущей стадии. Схема алгоритма представлена на рисунке 2.23. Ниже описаны шаги алгоритма.

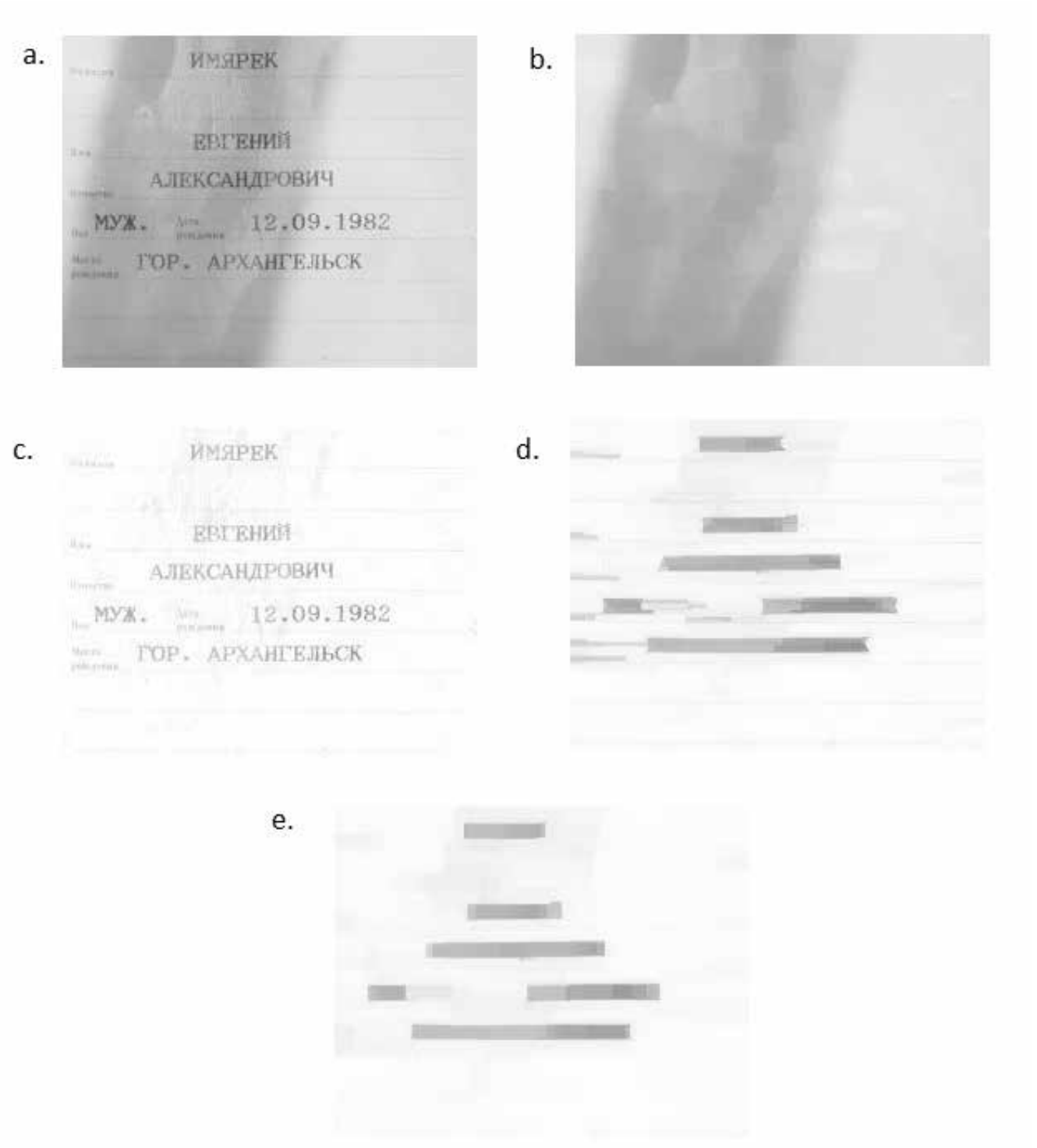


Рисунок 2.22 — Результаты предобработки изображения по этапам.

1. Исходное изображение содержит характерные компоненты текстовых полей. Для отделения их от фона вычисляем пороговое значение одним из методов бинаризации, например, методом Отсу. Данное значение используется для подсчета гистограмм на следующих шагах алгоритма.

Рисунок 2.23 — Алгоритм поиска текстовых полей.

2. Вычисляем вертикальную гистограмму, суммируя лишь те значения, которые меньше порога, так как компоненты текста темнее фона (см. рисунок 2.24).
3. Вычисляем верхние и нижние границы строк, содержащих текст, находя выбросы на гистограмме. Для отсечения случайных выбросов используются такие характеристики как минимальные и максимальные размеры высоты символов. Выбросы, размеры которых значительно меньше минимальной высоты, интерпретируем как случайные. Если же они больше максимального размера, то это случай слипания строк и потребуется дополнительный анализ для нахождения точек разрезания путём определения перепадов уже внутри самого выброса.
4. Для каждой найденной строки вычисляем горизонтальную гистограмму с использованием порога отсечения, аналогично шагу 2.
5. Находим границы полей внутри строк с помощью анализа горизонтальной гистограммы для каждой строки (рисунок 2.25). Для отсечения выбросов используется минимальный допустимый размер текстового поля (как минимум ширина одного символа). Для отделения одного текстового поля от другого внутри одной строки установим такую характеристику как минимальное расстояние между полями. Это возможно, ввиду того что в большинстве существующих документов для визуального отличия текстовых полей, содержащих различные данные, используется отделение их друг от друга на расстояние, значительно превышающее размер символа. Таким образом, близкие компоненты сливаются в одно поле, далекие друг от друга поля останутся отдельными полями и в результате получаем прямоугольники найденных полей документа.

6. Морфологические операции могут “съедать” часть элементов первых и последних символов полей, а также терять над и под строчные символы (например, умляуты) ввиду их малости. Чтобы избежать этого, границы найденных прямоугольников немного расширяются по горизонтали и вертикали (см. рисунок 2.26).

Рисунок 2.24 — Вычисление вертикальной гистограммы.

Рисунок 2.25 — Вычисление горизонтальной гистограммы.

2.6.3 Сопоставление найденных полей шаблону зоны документа

Приведенный выше метод поиска текстовых полей выдает как результат некоторый набор координат строк и найденных в них полей, но также необходимо сопоставить найденные текстовые поля атрибутам документа, с учетом возможного частичного их отсутствия. Например, для рассматриваемой зоны паспорта РФ “отчество” может отсутствовать, а количество строк, содержащих

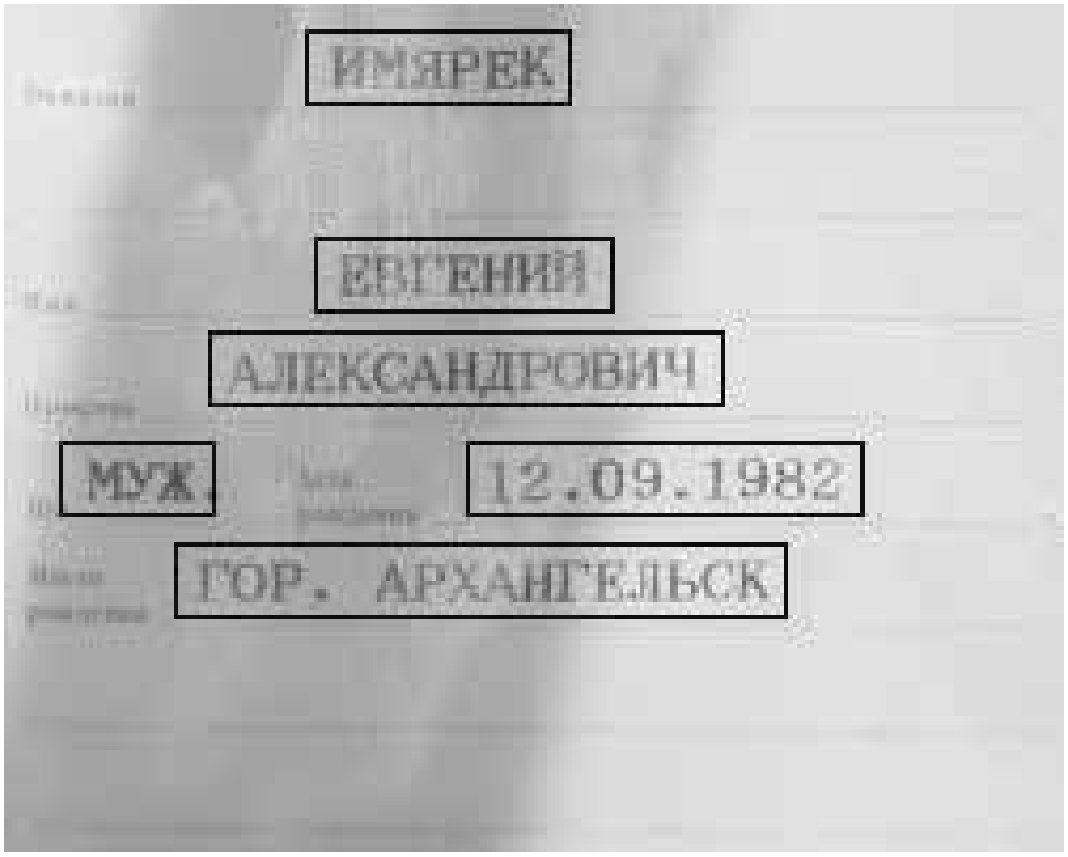


Рисунок 2.26 — Результат работы алгоритма поиска текстовых полей.

данные о “месте рождения”, варьируется от одной до трёх. Для решения этой задачи предлагается производить оценку полученного набора строк и полей на соответствие шаблону, описывающему зону документа. Для создания описания шаблона используется геометрическая модель соответствия строк и полей друг другу. Рассматриваем зону документа как набор строк, а каждую строку как набор полей. Для строк вводится отношение выше/ниже, для полей – левее/правее. Строки и поля могут быть необязательными, также для них могут вводиться дополнительные характеристики, такие как характерные размеры, положение внутри зоны (например, поле прижато к левому краю) и другие, все они используются для оценки соответствия полученных данных шаблону зоны. После этого производится рекурсивный перебор всех возможных вариантов сопоставления строк и полей, найденных на изображении, со строками и полями на шаблоне, с оценкой каждого такого сопоставления (см. рисунок 2.27).

При отсутствии такого сопоставления или очень низкой оценки выдаётся отказ, при наличии несколько вариантов – выбирается вариант с наибольшей оценкой или рассматривается дополнительная задача выбора. Отказ в большинстве случаев означает, что границы документа и его зоны были найдены неверно или исходное изображение слишком низкого качества, а данные – не читаемые.

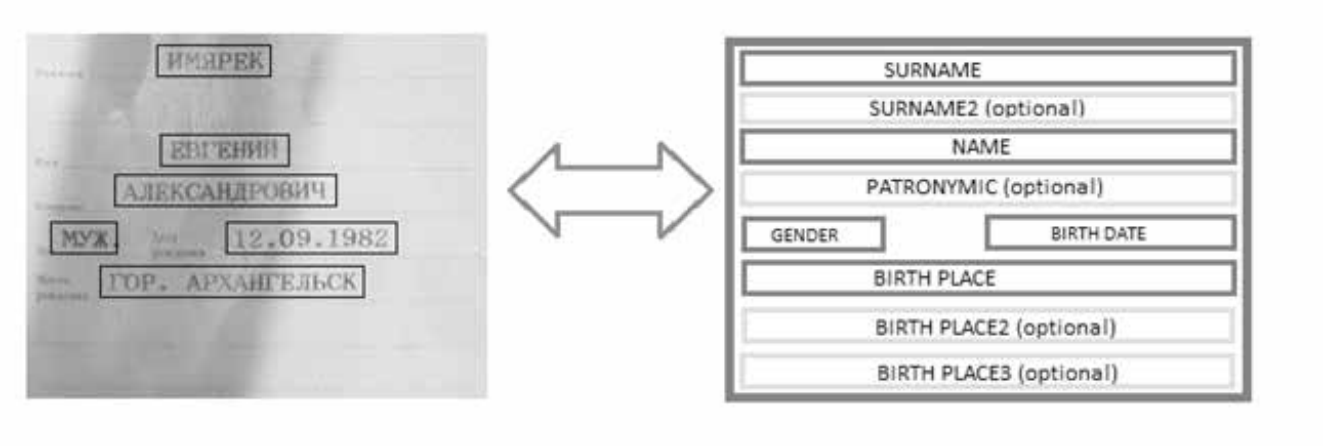


Рисунок 2.27 — Найденные поля (слева) и шаблон зоны документа (справа).

2.7 Особенности распознавания текстовой строки

Распознавание текстовых строк на изображениях документов, удостоверяющих личность, принципиально отличается от задач, в которых решается общая проблема оптического распознавания текста. Как было уже ранее сказано, важными особенностями текстовых строк, присутствующих на удостоверяющих документах и существенно усложняющих процесс распознавания текста, являются: а) сложный текстурный фон и б) мульти-языковость.

Отметим, что сама по себе задача распознавания текстовой строки, может быть разбита на следующие подзадачи:

– сегментация текстовой строки на символы;
– распознавание каждого символа;
– обобщение полученного результата распознавания.

При этом, решая задачу распознавания текстовой строки описанным способом, следует использовать различные нейронные сети для сегментации строки на символы и последующего распознавании каждого символа. Ниже представлен универсальный алгоритм построения подсистемы распознавания текстовой строки, который может использоваться в рамках системы распознавания удостоверяющих документов (см. рисунок 2.28).

Пусть у нас есть полносверточная нейронная сеть (NN_{segm}), которая возвращает оценку наличия в данной точки изображения разреза между буквами. Пусть также есть классифицирующая (распознающая) сверточная нейронная сеть (NN_{class}), позволяющая выполнить непосредственно распознавание каждого символа. Алгоритм распознавания состоит из следующих шагов:

1. Подготовим изображение текстовой строки к распознаванию, обрезав его по заранее определенным базовым линиям и приведя его к целевому масштабу (по высоте).
2. Применим сегментирующую полносверточную нейронную сеть NN_{segm} для построения карты потенциальных разрезов P.
3. Применим механизм подавления немаксимумов (non-maximum suppression) к карте потенциальных разрезов P для исключения ложных срабатываний, построив в результате P_1.
4. Построим пары ненулевых точек из P_1, расстояние между которых fdлежит в заранее заданном интервале $fd \in [fd_{min}; fd_{max}]$.
5. Предварительно проведем классификацию найденных участков изображений с помощью классифицирующей сети NN_{class}.
6. Применим технику динамического программирования для построения пути сегментации строки на символы, максимизируя сумму степеней уверенностей точек разреза и степеней уверенности распознавания каждого символа.
7. Проведем финальное распознавание символов в найденных после предыдущего шага позициях.

На рисунке 2.28 приведена схема описанного алгоритма распознавания текстовой строки, а также проиллюстрирован каждый шаг алгоритма.

Предложенная схема распознавания текстовый строки была испытана с точки зрения качества и скорости распознавания на наборе данных MIDV-500. В следующих таблицах представлены сравнительные характеристики предложенной схемы в сравнении с популярным системами распознавания.

Для оценки качества распознавания использовался метод оценки посимвольного распознавания (per-character recognition, PCR), определяемый следующим образом:

$$PCR = 1 - \frac{\sum_{i=1}^{L_{total}} \min\left(lev\left(l_{i_{ideal}}, l_{i_{recog}}\right), len\left(l_{i_{ideal}}\right)\right)}{\sum_{i=1}^{(L_{total})} len\left(l_{i_{ideal}}\right)}. \tag{2.10}$$

В таблицах 5 и 6 представлено сравнение качества распознавания и времени распознавания полей различного типа на разных системах распознавания строки.

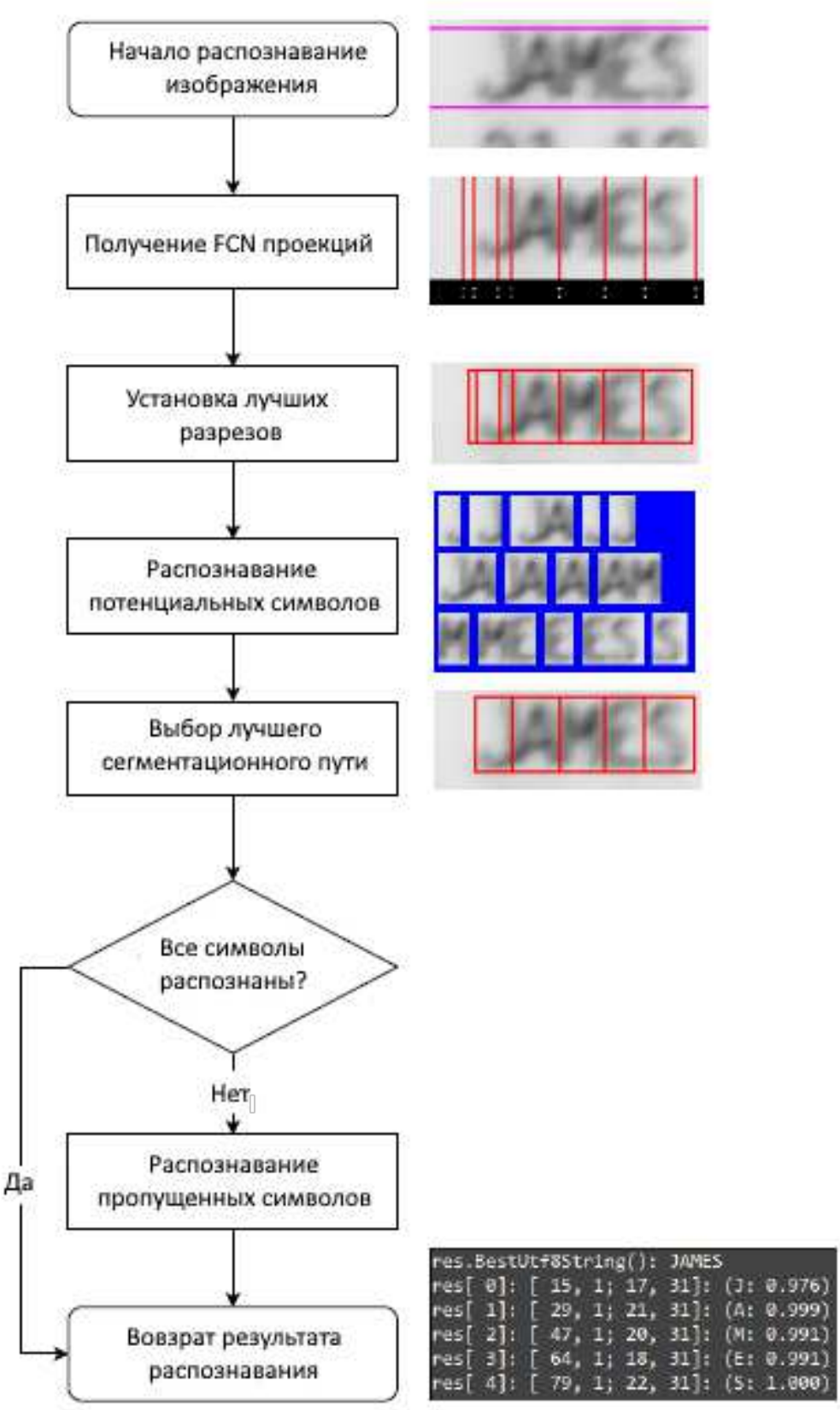


Рисунок 2.28 — Схема распознавания текстовой строки.

Таблица 5 — Сравнение качества распознавания отдельных полей на стенде MIDV-500 (в единицах $PCR \times 100\%$)

Вид строки	Предложенный метод	Tesseract 4.00	Tesseract 3.05	FineReader 15
Латинские имена	79.04	75.44	37.29	55.76
Даты	84.59	57.80	41.85	56.67
MRZ	92.98	47.94	58.52	74.11
Номер документа	80.06	41.83	27.27	57.11

Таблица 6 — Сравнение времени распознавания всех изображений для каждого поля в наборе данных MIDV-500 (в секундах)

Вид строки	Предложенный метод	Tesseract 4.00	Tesseract 3.05
Латинские имена	112.697	304.269	714.159
Даты	121.755	371.914	696.674
MRZ	233.179	586.808	731.757
Номер документа	110.067	227.961	446.367

2.8 Проверка подлинности

Как уже было сказано выше, документы, удостоверяющие личность, образуют специальный класс документов, который отличается от остальных особенностями их изготовления и применения. Кроме того, существуют международные стандарты изготовления паспортно-визовых документов, которые определяют дополнительные черты таких документов, например, на всех паспортах предназначенных для путешествий имеется специальная зона – машиночитаемая зона (МЧЗ, MRZ), которая специально наносится для автоматического считывания и в которой дублируется часть данных. Для чего используется избыточное кодирование информации за счет использования контрольных сумм. Также большинство документов такого типа изготовляются с применением различных технологий, усложняющих их фальсификацию. Используется павловская печать, специальные шрифты (иногда они являются секретными), печать методом тиснения, лазерная гравировка и много другое видимое без использования специальных средств. Все эти специальные

особенности можно и нужно использовать для доказательства подлинности документа. Таким образом, можно сформулировать следующую постановку задачи: в процессе распознавания документа необходимо проверить все видимые специальные особенности изготовления документа. Важной особенностью этой задачи является низкая толерантность к ошибкам типа ложного срабатывания. Это связанно с тем, что в большинстве стран попытка предъявления заведомо поддельного документа уголовно преследуется, поэтому ложные срабатывания крайне “токсичны”, хотя внешне это и не очевидно.

Таким образом возникает внутренние противоречие, с одной стороны, необходимо максимально тщательная проверка всех особенностей и данных, с другой стороны, ложные срабатывания обходятся дорого. В принципе подбор правильного порога зависит от способа и места применения. В этой работе не ставиться цель детального исследования этого вопроса. В настоящем разделе ограничимся постановкой задачи и исследуем основные способы контроля: сверку избыточных данных, способы нанесения текстов и анализ печатей.

2.8.1 Сверка избыточных данных

Одним из признаков подлинности являются распознанные символы полей. Из-за ошибок оцифровки образа возможно появление ошибок распознавания. Одним из способов повышения надежности распознавания является объединение результатов распознавания двух или более полей, содержащих одну и ту же информацию. Примером является заграничный паспорт гражданина РФ, в котором информация из поля «Фамилия» дублируется в части строки MRZ. Очевидно, что совпадение результатов распознавания двух образов, содержащих одну и ту же информацию, повышает уверенность в результате.

В данном разделе рассмотрены вопросы оценки надежности распознавания полей удостоверяющих личность документов с дублированием информации.

Вероятностная модель

Введем модель распознавания текста документа. Пусть имеется конечный алфавит A множества допустимых символов. Будем считать, что изображение символа на документе – это элемент некоторого множества I_A возможных изображений символов. Введем два оператора.

1. $T : I_A \to A$ – отображение, возвращающее изображенный символ от его изображения.
2. $G : I_A \to A$ – оператор распознавания, отображение, возвращающее распознанный символ от изображения символа.

Будем считать, что операторы G и T определены и детерминированы, а на пространстве I_A задана сигма-алгебра, относительно которой T и G измеримы, и вероятностная мера на ней.

Пусть S – множество слов фиксированной длины, составленных из символов алфавита A, I_S – множество изображений слов на документе. Будем считать, что элементы S – это упорядоченные наборы фиксированной длины элементов A, а элементы I_S – упорядоченные наборы фиксированной длины элементов I_A. Сигма-алгебра на I_S индуцируется цилиндрическими множествами, образованными по элементам сигма-алгебры I_A и выбранным позициям в наборе. Расширим область определения операторов G и T на слова.

1. $T : I_S \to S$ – возвращает изображенное слово от его изображения.
2. $G : I_S \to S$ – возвращает распознанное слово от изображения слова.

Запишем условие согласованности для изображений слов и изображений символов.

Пусть

$$is = (ia_1, ia_2, \ldots, ia_n),\ is \in I_S,\ ia_k \in I_A,\ k = 1, \ldots, n, \tag{2.11}$$

тогда

$$\begin{aligned} &T(is) = s = (a_1, a_2, \ldots, a_n) \Leftrightarrow \\ &\Leftrightarrow T(ia_1) = a_1,\ T(ia_2) = a_2, \ldots, T(ia_n) = a_n, \\ &s \in S,\ a_k \in A,\ k = 1, \ldots, n. \end{aligned} \tag{2.12}$$

Будем считать, что распознавание слов происходит посимвольно и независимо:

$$\begin{aligned} &G(is) = (G(ia_1), G(ia_2), \ldots, G(ia_n)), \\ &is \in I_S,\, ia_k \in I_A,\, k = 1, \ldots, n. \end{aligned} \tag{2.13}$$

Распознавание слова

Рассмотрим упрощенную вероятностную меру на множестве I_S. Предположим, что изображения символов в словах, длина изображенного слова – независимы в совокупности. Так же предположим, что изображения символов в словах имеют одинаковое распределение. Пусть нам известны вероятности

$$P(G(ia) = b | T(ia) = a),\, i_a \in I_A,\, \forall\, b, a \in A. \tag{2.14}$$

Для упрощения, будем считать, что

$$P(G(ia) = a | T(ia) = a) = p,\ p \in [0{,}1], \forall a \in A. \tag{2.15}$$

Утверждение 1. *Пусть дано изображение is слова длины n, тогда*

$$P(G(is) = s | T(is) = s) = p^n. \tag{2.16}$$

Доказательство. Запишем

$$is = (ia_1, ia_2, \ldots, ia_n),\, s = (a_1, a_2, \ldots, a_n), \tag{2.17}$$

тогда

$$P(G(is) = s | T(is) = s) = P(G(ia_1) = a_1, \ldots, G(ia_n) = a_n). \tag{2.18}$$

В предположении независимости изображений символов справедливо

$$\begin{aligned} &P(G(ia_1) = a_1, \ldots, G(ia_n) = a_n) = \\ &= P(G(ia_1) = a_1) P(G(ia_2) = a_2) \ldots (G(ia_n) = a_n) = p^n. \end{aligned} \tag{2.19}$$

□

Отсюда следует, что вероятность верного распознавания слова экспоненциально убывает при увеличении длины слова в такой модели. Заметим также, что в данной вероятностной модели вероятность верного распознавания не зависит от символов в слове, но зависит от его длины.

Распознавание нескольких полей

Теперь рассмотрим вероятность согласованности результатов распознавания в нескольких полях.

Утверждение 2. *Пусть имеется k изображений слова $is_1, is_2, \ldots, is_k$, на которых изображено слово s длины n. Обозначим за H событие $\{T(is_1) = T(is_2) = \ldots = T(is_k) = s\}$. Тогда для вероятности совпадения результатов распознавания верно*

$$\begin{aligned} &1)\ P(G(is_1) = G(is_2) = \ldots = G(is_k)|H) \leqslant (p^k + (1-p)^k)^n, \\ &2)\ P(G(is_1) = G(is_2) = \ldots = G(is_k)|H) \geqslant \left(p^k + \frac{(1-p)^k}{(|A|-1)^{k-1}}\right)^n. \end{aligned} \tag{2.20}$$

Доказательство. Запишем

$$\begin{aligned} &is_j = (ia_{j,1}, ia_{j,2}, \ldots, ia_{j,n}),\ j \in \{1,2,\ldots,k\}, \\ &s = (a_1, a_2, \ldots, a_n). \end{aligned} \tag{2.21}$$

Тогда

$$\begin{aligned} &P(G(is_1) = G(is_2) = \ldots = G(is_k) = s|H) = \\ &= \prod_{j=1}^{n} P\left(G(ia_{1,j}) = G(ia_{2,j}) = \ldots = G(ia_{k,j})|H\right). \end{aligned} \tag{2.22}$$

Каждую вероятность в произведении можно записать как

$$\begin{aligned} &P\left(G(ia_{1,l}) = G(ia_{2,l}) = \ldots = G(ia_{k,l})|H\right) = \\ &= \sum_{j \in A} P(G(ia_{1,l}) = j, \ldots, G(ia_{k,l}) = j|H) = \\ &= p^k + \sum_{j \in A \setminus \{a_l\}} P(G(ia_{1,l}) = j|H) \ldots P(G(ia_{k,l}) = j|H). \end{aligned} \tag{2.23}$$

Отметим, что вероятности $P(G(ia_{i,l}) = j|H)$ равны для всех $i \in \{1, \ldots, k\}$, в силу одинаковой распределенности изображений символов. Решим задачу оптимизации

$$f(x_1, x_2, ..., x_n) = \sum_{j=1}^{n} x_j^k \to \min,$$
$$x_j \geqslant 0,\ j = 1, \ldots, n, \tag{2.24}$$
$$\sum_{j=1}^{n} x_j = a.$$

Функция Лагранжа примет следующий вид

$$L(x, \lambda, \mu) = \sum_{j=1}^{n} x_j^k + \sum_{j=1}^{n} \mu_j x_j + \lambda \left(\sum_{j=1}^{n} x_j - a \right). \tag{2.25}$$

Условия из теоремы Куна-Таккера [211]:

$$\begin{aligned} &1)\ \mu_j x_j = 0,\ j = 1, \ldots, n; \\ &2)\ L'_\lambda = \sum_{j=1}^{n} x_j - a = 0; \\ &3)\ L'_{x_j} = k x_j^{k-1} + \mu_j + \lambda = 0,\ j = 1, \ldots, n;\ x_j \neq 0. \end{aligned} \tag{2.26}$$

После получения решений системы и нахождения минимума среди них, получаем решение задачи

$$x_1 = x_2 = \ldots = x_n = \frac{a}{n}, \tag{2.27}$$

из этого

$$\begin{aligned} &\sum_{j \in A \setminus \{a_l\}} P(G(ia_{1,l}) = j|H) \ldots P(G(ia_{k,l}) = j|H) \geqslant \\ &\geqslant (|A| - 1) \left(\frac{1-p}{|A|-1} \right)^k, \end{aligned} \tag{2.28}$$

тем самым получаем второе неравенство.

Кроме того, из задачи

$$f(x_1, x_2, ..., x_n) = \sum_{j=1}^{n} x_j^k \to \max \tag{2.29}$$

имеем

$$\sum_{j \in A \setminus \{a_l\}} P(G(ia_{1,l}) = j|H) \ldots P(G(ia_{k,l}) = j|H) \leqslant (1-p)^k. \tag{2.30}$$

Итак, окончательно получаем

$$\begin{aligned} &P\left(G(ia_{1,l}) = G(ia_{2,l}) = \ldots = G(ia_{k,l})|H\right) \leqslant \\ &\leqslant \prod_{j=1}^{n}(p^k + (1-p)^k) = (p^k + (1-p)^k)^n. \end{aligned} \tag{2.31}$$

Первое неравенство также доказано.

□

Вероятность ошибки второго рода

Произведем оценку вероятности ложного совпадения результатов распознавания полей.

Утверждение 3. *Пусть имеется* 2 *изображения слов* is_1, is_2 *длины* n. *Пусть* $p > 0.5$. *Обозначим за* H *событие* $\{T(is_1) \neq T(is_2)\}$. *Тогда для вероятности совпадения результатов распознавания верно*

$$P(G(is_1) = G(is_2)|H) \leqslant 2p(1-p)(p^2 + (1-p)^2)^{n-1}. \tag{2.32}$$

Доказательство. Запишем

$$\begin{aligned} &is_j = (ia_{j,1}, ia_{j,2}, \ldots, ia_{j,n}), \\ &s_j = (a_{j,1}, a_{j,2}, \ldots, a_{j,n}),\ j \in \{1,2\}. \end{aligned} \tag{2.33}$$

Оценим вероятности совпадения результатов распознавания изображений разных букв, находящихся в одинаковых позициях, для двух разных изображений слов.

$$\begin{aligned} &P\left(G(ia_{1,l}) = G(ia_{2,l}))|H\right) = \\ =&\sum_{j\in A} P(G(ia_{1,l}) = j, G(ia_{2,l}) = j|H) = \\ =&\sum_{j\in A} P(G(ia_{1,l}) = j|H)P(G(ia_{2,l}) = j|H) = \\ =&p\left(P(G(ia_{1,l}) = a_{2,l}|H) + P(G(ia_{2,l}) = a_{1,l}|H)\right) + \\ +&\sum_{j\in A\setminus\{a_{1,l},a_{2,l}\}} P(G(ia_{1,l}) = j|H)P(G(ia_{2,l}) = j|H). \end{aligned} \tag{2.34}$$

Сделаем оценку сверху в предположении $p > 0.5$. Аналогично доказательству второго неравенства из утверждения 2

$$\begin{aligned}&\sum_{j\in A\setminus\{a_{1,l},a_{2,l}\}} P(G(ia_{1,l}) = j|H)P(G(ia_{2,l}) = j|H) \leqslant \\ &\leqslant (1-p-P(G(ia_{1,l}) = a_{2,l}|H))(1-p-P(G(ia_{2,l}) = a_{1,l}|H)).\end{aligned} \tag{2.35}$$

Тогда

$$\begin{aligned}&P\left(G(ia_{1,l}) = G(ia_{2,l}))|H\right) \leqslant \\ &\leqslant p\left(P(G(ia_{1,l}) = a_{2,l}|H) + P(G(ia_{2,l}) = a_{1,l}|H)\right) + \\ &+ (1-p-P(G(ia_{1,l}) = a_{2,l}|H))(1-p-P(G(ia_{2,l}) = a_{1,l}|H)).\end{aligned} \tag{2.36}$$

Введем обозначения

$$\begin{aligned}&P(G(ia_{1,l}) = a_{2,l}|H) = p_1 \in [0, 1-p], \\ &P(G(ia_{2,l}) = a_{1,l}|H) = p_2 \in [0, 1-p], \\ &F(p_1, p_2) = p\left(p_1 + p_2\right) + (1-p-p_1)(1-p-p_2).\end{aligned} \tag{2.37}$$

Для функции $F(p_1, p_2)$ справедливо

$$\begin{aligned}&F(p_1, p_2)'_{p_1} = p - (1-p-p_2) = 2p - 1 + p_2 \geqslant 0, \\ &\text{т.к.} \quad 2p-1 > 0;\ p_2 \geqslant 0. \\ &\text{Аналогично} \ \ F(p_1, p_2)'_{p_2} \geqslant 0.\end{aligned} \tag{2.38}$$

Следовательно, функция F принимает свой максимум на правых границах значений p_1 и p_2. Таким образом,

$$\begin{aligned}&P\left(G(ia_{1,l}) = G(ia_{2,l}))|H\right) = F(p_1, p_2) \leqslant \\ &\leqslant F(1-p, 1-p) = 2p(1-p).\end{aligned} \tag{2.39}$$

Так как вероятность совпадения результатов распознавания $ia_{m,j}$ и $ia_{m,j}$ выше при $T(ia_{m,j}) = T(ia_{l,j}),\ p > 0.5$, то для оценки сверху можно считать, что слова совпадают во всех позициях кроме одной. Оценка сверху при $T(ia_{m,j}) = T(ia_{l,j})$ была получена в утверждении 2 и равна $p^2 + (1-p)^2$. Тогда

$$P\left(G(is_1) = G(is_2))|H\right) \leqslant 2p(1-p)(p^2 + (1-p)^2)^{n-1}. \tag{2.40}$$

□

Применение оценок к вероятности распознавания даты

Применим полученные оценки к задаче распознавания даты в нескольких полях. Будем считать, что дата это слово длины 6 (2 символа дня, 2 – месяца, 2 – года) с символами из алфавита мощности 10. Тогда из утверждения 1, вероятность верного распознавания даты в одном поле соответствует функции p^6, график которой изображён на рисунке 2.29, где по горизонтальной оси отмечено значение параметра p, а по вертикальной искомая вероятность.
График иллюстрирует, что даже при общей для цифр вероятности верного распознавания равной 0.98, вероятность верного распознавания целого поля равна 0.8858. Для целевого качества распознавания поля даты в 0.98 необходимо распознавать цифры с качеством не меньше 0.9966. Обратный график изображен на рисунке 2.30.

Пусть система для проверки документа на подлинность проверяет совпадение некоторых результатов распознаваний его полей, причем в случае хотя бы одного несовпадения не подтверждает подлинность. Назовем ситуацию ложным срабатыванием, если слова в полях в действительности совпадают, но их результаты распознаваний в совокупности различны. Это эквивалентно тому, что на вход системе дан подлинный документ, но система не подтвердила его подлинность. Предположим, что для проверки системой документа на подлинность необходимо проверить на совпадение четыре поля даты, воспользуемся оценкой из утверждения 2. График нижней оценки вероятности совпадения распознаваний при условии действительного совпадения дат в полях изображён на рисунке 2.31. «Экспоненциальный» эффект усиливается в такой постановке, при распознавании цифры с качеством 0.98 мы имеем вероятность ложного срабатывания $1 - 0.6158$ в худшем случае.

Пусть есть два поля с несовпадающей датой, график верхней оценки вероятности подтвердить их равенство по результату распознавания изображен на рисунке 2.32 зеленым цветом. Отметим, что в реальности вероятность получить одинаковый результат распознавания для двух несовпадающих дат будет близка к графику только в тех случаях, где даты различаются только по одному символу, так была построена оценка в утверждении 3. Для сравнения

на рисунке 2.32 синим цветом изображен график верхней оценки той же вероятности, но при различии символов обоих дат во всех позициях. Этот график задается уравнением $y = (2x(1-x))^n$, которое легко получить, используя конструкции из вывода утверждения 3. Наличие нетривиального экстремума у зеленого графика можно объяснить тем, что при росте p вероятность подтвердить равенство в $n-1$ реально совпадающей позиции увеличивается, но при этом вероятность подтвердить равенство в несовпадающей позиции остается существенной, тем самым общая вероятность подтвердить равенство дат растет; но когда p становится слишком большим, мы почти наверное не подтверждаем равенство в несовпадающей позиции и общая вероятность подтвердить равенство дат стремиться к нулю.

Приведем пример расчета необходимой точности распознавания символа для целевого качества 0.01 по ложным срабатываниям в случае трех полей. Пусть в полях изображены три даты длины 6. По пункту 1 утверждения 2 имеем

$$\begin{aligned} &1 - 0.01 \leqslant (p^3 + (1-p)^3)^6, \\ &0.99833 \leqslant 3p^2 - 3p + 1, \\ &p \geqslant 0.99944. \end{aligned} \tag{2.41}$$

Точность распознавания символа 0.99944 является необходимой для заданной целевой вероятности, но достаточная точность будет близка к данной, так как правые части пунктов 1 и 2 из утверждения 2 асимптотически равны при $p \to 1$.

Проведем численный эксперимент для ID-документа, состоящего из пяти полей (фамилия, имя, отчество, пол, дата), каждое из которых присутствует в двух экземплярах: непосредственно и как часть MRZ-строки). Рассчитаем достаточную точность распознавания одного символа для обеспечения целевого качества по ложным срабатываниям, равного 0.01. Обозначим изображения слов в парах полей фамилии, имени, отчества, пола и даты за

$$(is_1, is_2), (in_1, in_2), (ip_1, ip_2), (ie_1, ie_2), (id_1, id_2) \tag{2.42}$$

соответственно.

Алфавит для фамилии, имени, отчества состоит из 33 букв, для пола из 2, для даты из 10.

Имеем

$$\begin{aligned} &P\left(G(is_1) = G(is_2), G(in_1) = G(in_2), \ldots, G(id_1) = G(id_2)\right)) = \\ &= P\left(G(is_1) = G(is_2)\right) \ldots P\left(G(id_1) = G(id_2)\right)), \end{aligned} \tag{2.43}$$

в силу независимости распознаваний. По утверждению 2 получаем

$$\begin{aligned} &P\left(G(is_1) = G(is_2)\right) \ldots P\left(G(id_1) = G(id_2)\right)) \geqslant \\ &\left(p^2 + \frac{(1-p)^2}{32}\right)^l \left(p^2 + (1-p)^2\right) \left(p^2 + \frac{(1-p)^2}{9}\right)^6 \geqslant 0.99, \end{aligned} \tag{2.44}$$

где l суммарная длина фамилии, имени и отчества.

Найдя наименьшую точность распознавания символа p из таких неравенств при фиксированных l и взяв эмпирическое распределение l, мы получили математическое ожидание p по распределению l, равное 0.99983.

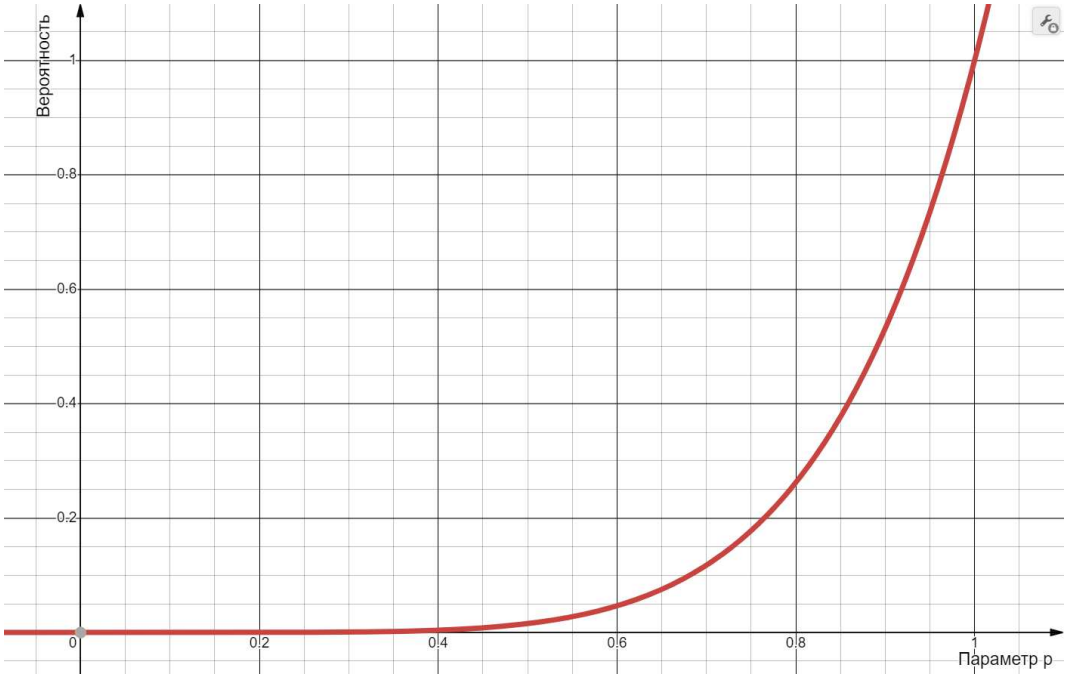


Рисунок 2.29 — Вероятность верного распознавания даты в одном поле.

2.8.2 Распознавание текста на изображении оттиска печати

Печати и штампы играют важную роль при рассмотрении задачи распознавания документов, удостоверяющих личность. Помимо придания необходимой юридической силы документу, печати и штампы зачастую используются для контроля валидности и подлинности.

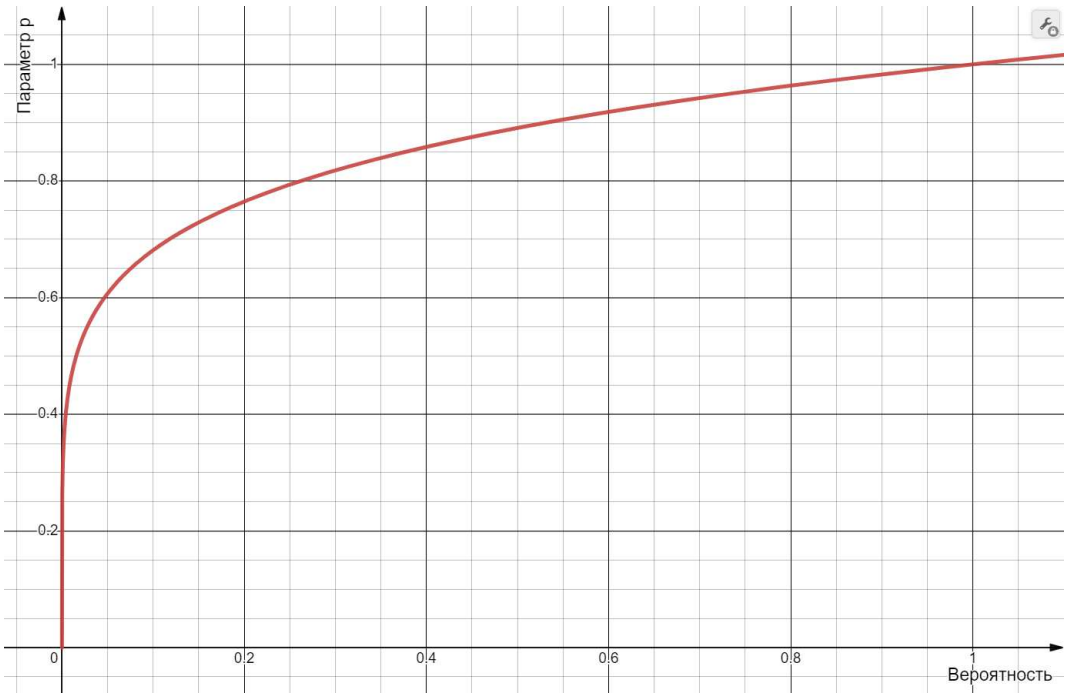


Рисунок 2.30 — Зависимость необходимого качества распознавания цифры для распознавания поля с заданной вероятностью.

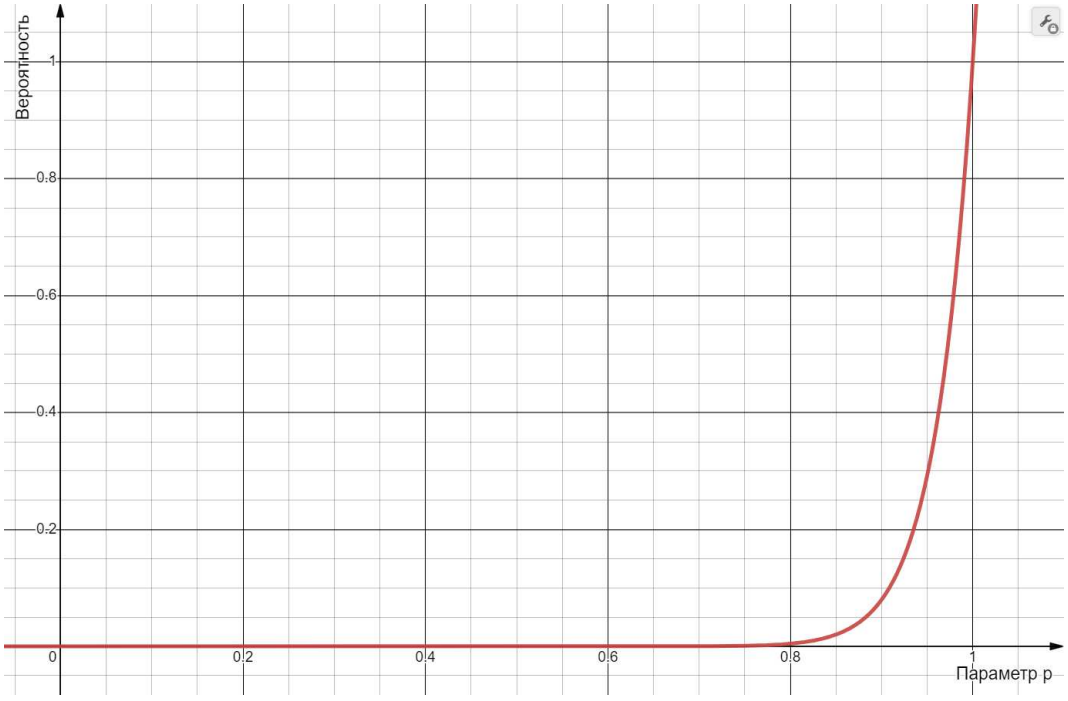


Рисунок 2.31 — Вероятность совпадения результатов распознавания четырех полей дат, при их совпадении в реальности.

Наибольшую популярность в Российской Федерации получили печати круглой формы. При этом, вдоль обода такой печати принято наносить юридически значимую информацию (индивидуальные номера отделений организаций, ИНН юридических лиц и т. п.). Распознавание такого текста является отдельной

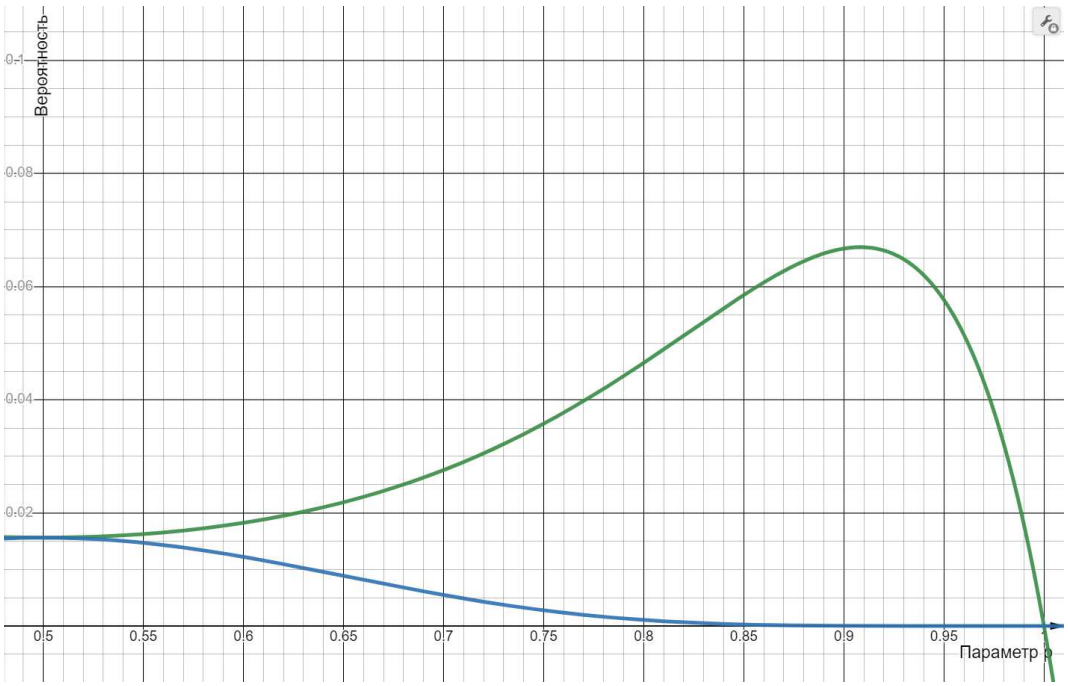


Рисунок 2.32 — Вероятность совпадения результатов распознавания двух полей дат, при их различии в реальности, зеленым показан график при различии в одной позиции, синим – во всех позициях.

важной подзадачей общей задачи распознавания документа, удостоверяющего личность.

В настоящем разделе будет представлен способ распознавания оттисков круглых печатей на примере общегражданского паспорта гражданина РФ, который в соответствии с требованиями содержит оттиск печати органа выдачи документа.

Локализация оттиска печати

Задача локализации оттиска печати на нормализованном изображении документа относится к классу задач детекции объектов и заключается в определении точного положения изображения оттиска печати.

В качестве метода детекции точного положения печати могут использоваться следующие подходы:

1. Обобщенное преобразование Хафа. Для каждой точки на изображении существует конечное множество окружностей, которым она может принадлежать. Тогда, т.к. окружность задается тремя параметрами (координатами

центра и радиусом), введя трехмерный массив (аккумулятор), можно провести процедуру голосования, где каждая точка голосует за все окружности, которым она может принадлежать, победителей которой можно объявить истинными окружностями, соответствующими положениям печатей [212].

2. Метод Виолы и Джонса. Данный метод позволяет с помощью техники машинного обучения построить бинарный классификатор, который с помощью метода скользящего окна используется для решения задачи локализации искомого объекта [213].

3. Аппроксимация компонент связности краев изображения фигурами искомой формы. На бинарном изображении, являющемся результатом работы детектора Канни, можно выделить группы пикселей, транзитивно соседствующих друг с другом по одному из 8-ми направлений. Такие множества будем называть треками. Если на исходном изображении были печати, то соответствующие треки будут похожи на дугу окружности. Поиск и анализ таких треков позволяет локализовать печати на исходном изображении.

Нормализация изображения текстовой строки

Для применения методов оптического распознавания текста к изображению печати зона текста геометрически нормализуется: происходит «разворот» круглой (круговой) полосы печати на изображении B в прямоугольник.

Пусть заданы предполагаемый радиус печати R_{pred}, желаемый отступ от края печати $indent$, ширина вырезаемой полосы $strip$ и координаты центра C_{stamp} (C_x, C_y).

Пусть необходимая высота итогового «развернутого» прямоугольника H_{unwrap} (см. рисунок 2.34). Ширина «развернутого» прямоугольника рассчитывается так (в соответствии с длиной окружности радиуса $(R_{pred} - indent)$:

$$W_{unwrap} = 2 \cdot \pi \cdot (R_{pred} - indent). \tag{2.45}$$

Для каждой точки «развернутого» изображения рассчитывается соответствующая ей точка исходного. Пусть идет расчет координат точки P_{src} соответствующей точке «развернутого» изображения P_{dst} (row, col). Порядок вычислений следующий:

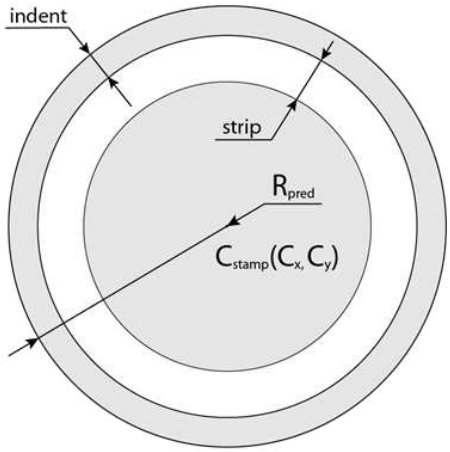


Рисунок 2.33 — Структурная схема оттиска круглой печати.

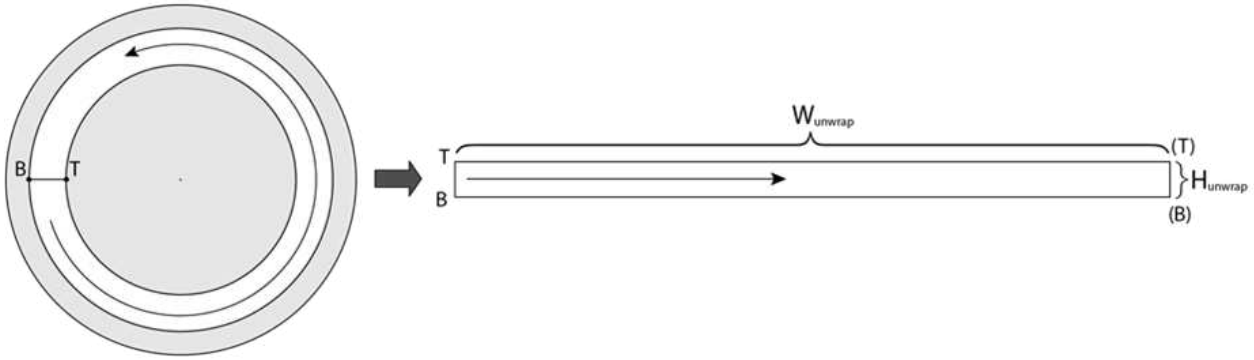


Рисунок 2.34 — Схема «разворота» оттиска печати.

1. Вычисляют расстояние от P_{src} до внешней окружности вырезаемой полосы:

$$distance2strip\_border \ = strip * (1.0 \ - \ row/H_{unwrap}). \tag{2.46}$$

2. Вычисляют расстояние от P_{src} до внешней окружности печати:

$$distance2stamp\_border = indent \ + \ distance2strip\_border. \tag{2.47}$$

3. Вычисляют расстояние от P_{src} до центра печати C_{stamp}

$$distance2center = R_{pred} - distance2stamp\_border. \tag{2.48}$$

4. Вычисляют угол между горизонталью и отрезком CP_{src} по часовой стрелке:

$$\alpha = \ 2 * \pi * (1 - col/W_{unwrap}). \tag{2.49}$$

5. Вычисляют координаты P_{src} посчитав сдвиги вдоль осей координат P_{src} от центра печати:

$$x = C_x \ - \ distance2center * cos(\alpha) y = C_y \ - \ distance2center * sin(\alpha) \tag{2.50}$$

Таким образом, точке P_{dst} (row,col) соответствует точка P_{src} (x,y). Результат представлен на рисунке 2.35.

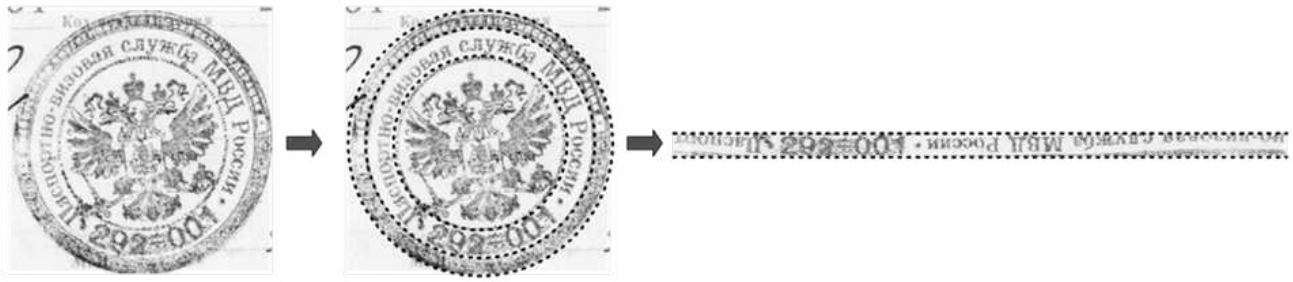


Рисунок 2.35 — Пример «разворота» оттиска печати.

Поиск и распознавание области кода подразделения

После «разворота» текстовой строки необходимо выполнить локализацию целевого текста. Код подразделения наносится по шаблону XXX-XXX. Благодаря наличию шаблона, непосредственно поиск области кода подразделения можно выполнить с помощью алгоритма Виолы и Джонса на изображении.

Распознавание кода подразделения на изображении производят с помощью оптического распознавания символов. Результатом является текстовая строка с альтернативами и оценками распознавания, также называемая матрицей альтернатив или AP-цепь. Процесс происходит следующим образом: распознаваемая зона (строка) обладает массивом точек разрезания x_0, x_1, ..., x_N. Для каждой из пар точек x_i и x_j образ символа распознается методом, дающим штрафную оценку $r(x_i, x_j)$. Для путей τ, являющихся подмножеством исходного набора отрезков разрезания, подсчитывается мера $m(\tau)$ как наименьшая из оценок пар соседних точек разрезания $r(x_i, x_{i+1})$. Целью является нахождение пути с максимальной оценкой $m(\tau)$. Оптимальный путь определяется с помощью динамического программирования, опирающегося в каждом отрезке на уже построенные оптимальные пути в промежуточные точки, этим достигается

построение оптимального пути, ведущего из начальной точки зоны сегментации в ее конечную точку за один проход.

На рисунке 2.36 приведен пример матрицы альтернатив (AP-цепи), соответствующей изображению номера.

$2_{0,988}$	$9_{0,999}$	$2_{0,980}$	$[\ ]_{0,770}$	$0_{0,988}$	$0_{0,999}$	$1_{0,980}$
$7_{0,012}$	$7_{0,01}$	$9_{0,020}$	$-_{0,230}$	$6_{0,012}$	$8_{0,001}$	$4_{0,020}$

Рисунок 2.36 — Пример матрицы альтернатив (AP-цепи), соответствующей изображению номера.

2.8.3 Контроль способа нанесения текстовой информации

С целью обеспечения должного уровня долговечности документа, а также для придания дополнительной защиты от фальсификации, в документах, удостоверяющих личность, используются различные техники нанесения текстовой информации. Так, например, в текстовых поля помимо традиционных способов печати (лазерной, струйной, трафаретной или офсетной) могут применяться лазерная гравировка, лазерная перфорация, эмбоссирование символов и другие (см. рисунок 2.37).

Кроме того, важной задачей является контроль семейства и начертаний шрифтов, которые использовались при нанесении текстовых данных. В большинстве стран при описании стандартов оформления документов, удостоверяющих личность, жестко регламентированы шрифты, которыми должны наноситься текстовые поля (см. рисунок 2.38). Отступление от данных стандартов является, в частности, значимыми признаком предъявления фальсификации вместо оригинального документа.

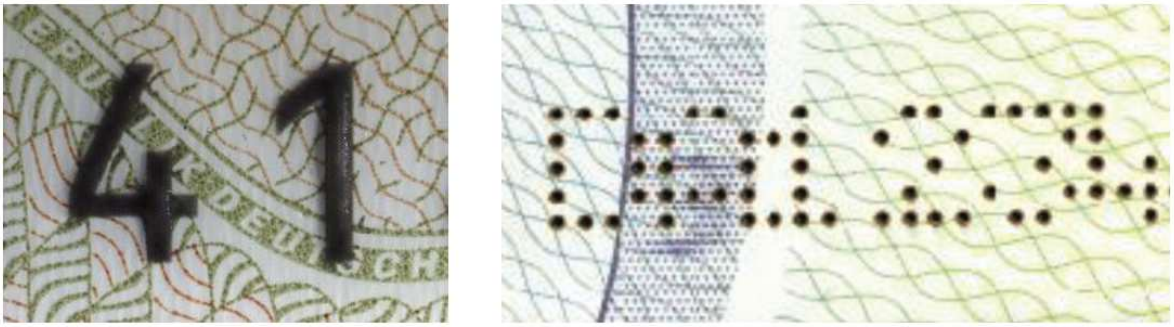


Рисунок 2.37 — Примеры текста, нанесенного методом лазерной гравировки с тактильным эффектом (слева) и лазерной перфорации (справа).

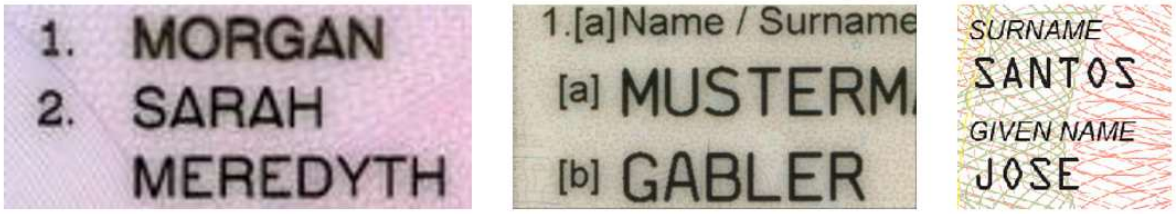


Рисунок 2.38 — Примеры используемых шрифтов при изготовлении (слева направо) британских водительских удостоверений, немецких идентификационных карт и филиппинских идентификационных карт.

Таким образом, проверка способа нанесения текстовых полей по изображению документа, удостоверяющего личность, является важной подзадачей в рамках задачи распознавания ID документов.

В настоящем разделе описан способ построения подсистемы контроля способа нанесения текстовых полей и выявления аномалий с использованием инструмента полносвязных нейронных сетей.

Рассмотрим следующую задачу. Пусть для определенного класса документов, удостоверяющих личность, заданные текстовые поля с точки зрения особенности нанесения обладают свойством A. Пусть есть исследуемое изображение текстового поля f. Необходимо проверить наличие свойства A у исследуемого изображения текстового поля f.

Алгоритм решения сформулированной задачи будем строить на базе бинарного нейросетевого детектора, представляющего из себя полносверточную нейронную сеть, которая в результате обработки изображения текстового поля f строит карту оценок присутствия и отсутствия свойства A в каждой точке изображения f.

Алгоритм предполагает анализ горизонтальных текстовых полей. Поэтому от оценок присутствия исследуемого свойства в каждом пикселе можем

перейти к интегральной оценке вдоль каждой вертикальной линии. Пусть ω_A^i и $\omega_{\overline{A}}^i$ – интегральные оценки присутствия и отсутствия свойства A по вертикальной линии $i \in 0, 1, \ldots, W_f - 1$, где W_f – ширина изображения текстовой строки f в пикселях. Определим суммы значений оценок S_A и $S_{\overline{A}}$ по всем вертикальным линиям следующим образом:

$$S_A = \sum_{i=0}^{W_f-1} \omega_A^i, \qquad S_{\overline{A}} = \sum_{i=0}^{W_f-1} \omega_{\overline{A}}^i. \tag{2.51}$$

Важно отметить, что рассматриваемое текстовое поле в целом обладает свойством A, если $S_A > S_{\overline{A}}$.

Назовем аномалией ситуацию, при которой в текстовом поле f была частично нарушена однородность свойства A текстового поля. Пример такой ситуации представлен на рисунке 2.39.

Рисунок 2.39 — Пример текстового поля без аномалий (слева) и с аномалиями в результате подмены символов (справа).

Из особенностей задачи нас интересуют аномалии, физический размер которых по ширине сопоставим с шириной символа. Пусть W_S – средняя ширина символов, присутствующих в текстовом поле f. Пусть T – минимальный порог для детекции класса при поиске аномалии. Тогда, анализируя интегральные оценки вдоль вертикальных линий можно найти наибольший участок длины L, для которого нарушена однородность свойства A текстового поля, то есть такую позицию j, для которой справедливо $\omega_{\overline{A}}^i > T \ \forall i \in j, \ldots, j + L$. Тогда, если $L > W_S$, то аномалия в исследуемом текстовом поле задетектирована. Блок-схема описанного в данном разделе алгоритма приведена на рисунке 2.40.

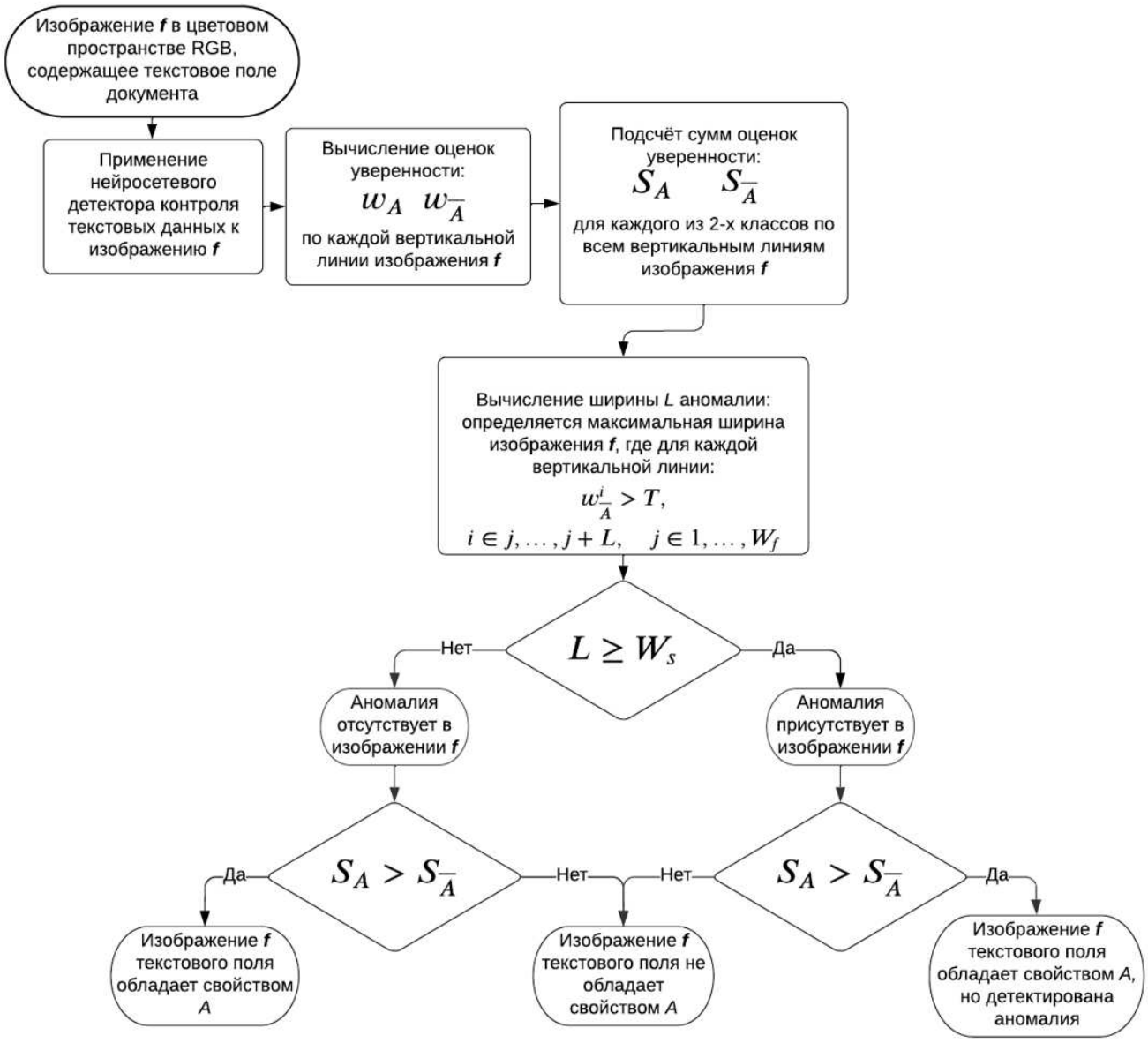


Рисунок 2.40 — Блок-схема алгоритма контроля способа нанесения текстовой информации.

2.9 Модель универсальной системы распознавания документов, удостоверяющих личность

2.9.1 Определения

Перед тем как перейти к рассмотрению ключевых проблем и особенностей, формально введем понятия 2D, 3D и 4D распознаваний.

Определение 1. *Будем называть распознавание изображения документа «распознаванием 2D», когда документ на исследуемом изображении расположен в плоскости изображения.*

Определение 2. *Будем называть распознавание изображения документа «распознаванием 3D», когда документ на изображении может быть повернут на все три угла Эйлера по отношению к оптической системе камеры, которая рассматривается в рамках модели камеры-обскуры.*

Определение 3. *Будем называть распознавание серии изображений документов «распознаванием 4D», когда исследуемые изображения являются отдельными кадрами видеопотока и могут быть упорядочены с учетом временной шкалы.*

2D системы распознавания удостоверяющих документов традиционно имеют дело с изображениями, полученными с помощью сканеров. В данном случае, помимо обычных для такого рода систем подзадач (включающих идентификацию визуальных элементов документа, анализ структуры, распознавание текстовых полей и т. д.), возникают дополнительные проблемы: произвольное смещение и поворот документа на изображении, а также, в некоторых случаях, неизвестное разрешение изображения (например, когда процесс получения изображения со сканера не контролируем). Тем не менее, все указанные подзадачи, связанные с определением геометрии документа, решаются фактически на плоскости (2D распознавание).

Если изображение документа получено с помощью веб-камеры или камеры мобильного устройства (случай 3D распознавания), то вместо анализа фактически плоского объекта необходимо анализировать трехмерную сцену, где документ должен быть локализирован с учетом проективных преобразований, возможных нелинейных искажений и произвольного фона. Помимо геометрических особенностей расположения документа, дальнейшие методы анализа документов должны учитывать также возможную расфокусировку, размытие, неравномерное или недостаточное освещение, а также блики на отражающей поверхности документа.

Наконец, если входные данные – это не единичное изображение, а последовательность видеокадров (случай 4D изображения), документ необходимо обрабатывать во времени с учетом меняющихся условий съемки. При этом избыточность визуальной информации может быть использована для повышения

надежности результатов анализа и распознавания документов, а изменяющиеся условия могут быть использованы для обнаружения и анализа оптических переменных эффектов (голограмм, оптически переменных красок и т. д.), которые широко используются для защиты документов, удостоверяющих личность.

Все перечисленные проблемы и особенности, возникающие при 2D, 3D и 4D распознавании удостоверяющих документов, схематично представлены на рисунке 2.41.

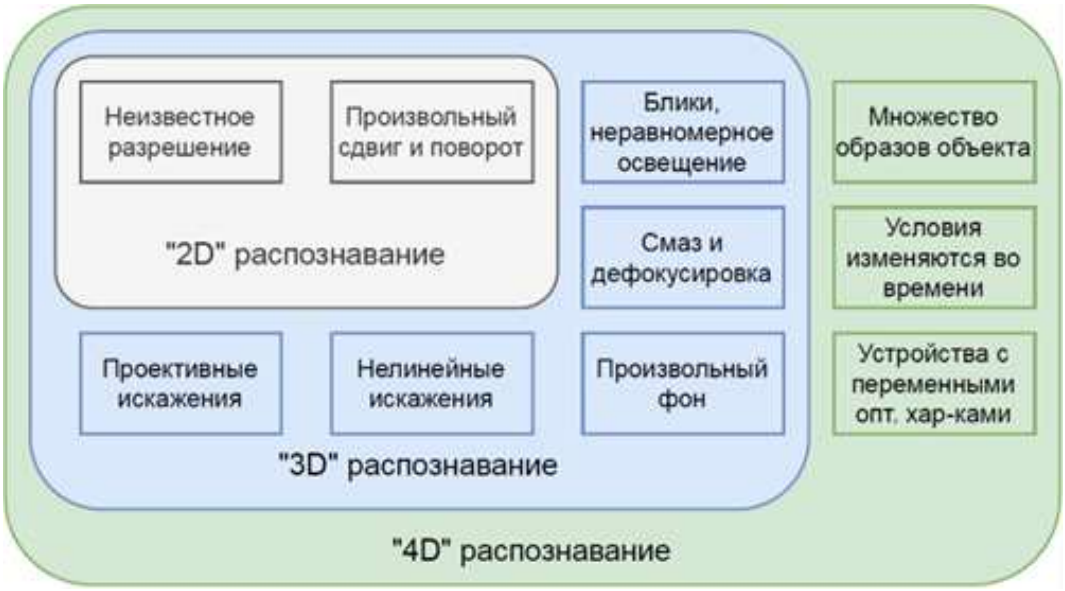


Рисунок 2.41 — Особенности 2D, 3D и 4D систем распознавания документов, удостоверяющих личность.

2.9.2 Подход к построению

Несмотря на различные условия получения входных данных (будь то изображения или последовательность видеокадров), различные модели геометрического положения документа и различные дополнительные искажения (такие как наличие размытия, бликов, неравномерного освещения и т.д.), непосредственно объект распознавания – документ, удостоверяющий личность – остается неизменным.

Поэтому разумным является подход построения современной системы распознаваний идентификационных документов, учитывающей специфику различных режимов получения изображения. При этом отдельные компоненты

при таком подходе и их взаимосвязь должны обеспечивать поддержку множества различных типов идентификационных документов, а архитектура должна поддерживать возможность включения дополнительных методов анализа и обработки документа.

В настоящем разделе представлен список предложений, которых следует придерживаться при построении универсальной распознающей системы (универсальность в данном случае подразумевает поддержку входных данных, соответствующих понятиям 2D, 3D, 4D).

1. **Использование методов локализации шаблона, устойчивых к проективным искажениям.** Применение таких методов позволяет автоматически обработать проективные преобразования, нелинейные искажения или произвольный фон, которые усложняют проблему распознавания 3D документов.
2. **Определение разрешения документа по результатам локализации шаблона.** Информацию о фактическом разрешении следует вычислять после этапа локализации и идентификации документа. Документы, удостоверяющие личность, относятся к классу документов с фиксированной («жесткой») структурой. Поэтому, зная физические размеры идентифицированного документа и определив границы документа на изображении, можно с высокой точностью вычислить фактическое разрешение.
3. **Независимый анализ качества изображения.** Анализ проблем расфокусировки, размытости, бликов или неравномерного освещения и прочие проблемы, свойственные случаям распознавания документов в 3D и 4D, следует вынести в отдельную подсистему. Выделение данного модуля облегчит анализ в сложных случаях и позволит на ранних этапах отфильтровать входные данные плохого качества.
4. **Анализ элементов документа на скорректированных и отмасштабированных изображениях.** Использование уже скорректированного (от проективных искажений) изображения шаблона документа позволяет применять одинаковые методы анализа отдельных элементов как в случае распознавания документов в 2D, так и в 3D и 4D.
5. **Обеспечения доступа к результатам распознавания предыдущих кадров.** В случае 4D распознавания документов условия съемки, изменяющиеся во времени, могут быть компенсированы на отдельных

этапах процесса благодаря доступу к временному хранилищу цикла распознавания и возможности использовать результаты распознавания из предыдущих видеокадров.

6. **Использование модулей комбинирования результатов распознавания.** Добавление модулей комбинирования позволяет повысить ожидаемую точность распознавания полей документа за счет использования информации из нескольких кадров, а также проводить анализ оптически изменяемых элементов.
7. **Фильтрация успешно распознанных полей.** Идея настоящего предложения заключается в реализации модуля, останавливающего распознавания отдельных полей на очередном кадре видеопоследовательности, если такое поле уже успешно распознано. Из-за особенностей удостоверяющих документов зачастую возникает ситуация, когда различные информационные поля требуют разного количества входных кадров для успешного распознавания. Ранний останов по отдельным полям для случая 4D распознавания позволяет сократить общее количество распознаваемых элементов, существенно сокращая итоговое время распознавания документа.

Следование указанным предложениям позволяет построить систем распознавания удостоверяющих документов, одинаково хорошо работающую с различными типами входных данных.

2.10 Выводы по главе

Главным результатом этой главы является переход от классических методов распознавания документов на основе бинаризации, анализа текстов, линий, углов, меток и т.п., к методам компьютерного зрения, применяющимся в анализе трехмерных сцен.

1. Проведен системный анализ архитектуры комплексов распознавания документов, удостоверяющих личность. Выявлены особенности документов, удостоверяющих личность, оказывающие влияние на процесс распознавания и особенности формирования изображений этих документов.

2. Представлен метод выделения текстовых строк, построенный с применением инструментов морфологического анализа и динамического программирования.
3. Предложен подход к проверке подлинности документов, удостоверяющих личность, на основе сверки избыточных данных, контроля способа нанесения текстовой информации и анализа круглых печатей.
4. Предложен подход к построению универсальной системы распознавания идентификационных документов с проверкой подлинности как композиции алгоритмов компьютерного зрения, применяемых в анализе трехмерных сцен, учитывающий выявленные особенности как документов, так и процессов формирования изображений.

Глава 3. Распознавание объектов в видеопотоке

3.1 Введение

В предыдущей главе были рассмотрены особенности системы распознавания, привносимые заменой сканирования на съемку камерой в неконтролируемых условиях. Также были проанализированы особенности, связанные с природой документа, удостоверяющего личность.

В этой главе проведен анализ изменений процесса распознавания при переходе к рассмотрению видеопотока. В рамках настоящей главы исследованы вопросы повышения качества распознавания за счет выбора наилучшего кадра или комбинирования результатов распознавания отдельных кадров, построено и исследовано распределение оценок распознавания, изучен важнейший вопрос о месте автоматического останова, а также детально исследованы вопросы производительности.

3.2 Модель системы распознавания в видеопотоке

3.2.1 Особенности процесса распознавания в видеопотоке

Внедрение технологических, социальных и коммерческих процессов, основанных на использовании мобильных устройств и технологий, в условиях современного мира уже является обыденностью. Системы технического зрения с использованием мобильных технологий, к примеру, системы автоматического ввода и анализа документов на мобильных устройствах, продолжают вытеснять традиционные стационарные системы. Развитие технологий технического зрения с применением мобильных устройств в условиях аппаратных ограничений, связанных с ними, становится все более актуальной задачей.

Классические системы распознавания и автоматического ввода предполагают использование сканированного изображения или фотографии объекта

в качестве его оцифрованного представления. При использовании мобильных устройств для оцифровки образов распознаваемых объектов возникает дополнительная возможность использовать видеопоток цифровой камеры помимо отдельных фотографий или кадров. Процесс фотографии объекта при помощи современных мобильных устройств предполагает этап “наведения” оператором объектива камеры на объект с отображением кадров видеопотока на экране устройства в реальном времени для контроля оператора. В случае, если обработка изображения производится с одного изображения, информация, которая содержится в захваченных предварительных кадрах используется лишь косвенно (оператором). При рассмотрении цельного видеопотока в качестве цифрового образа объекта появляется возможность использовать гораздо больше визуальной информации [214].

Рисунок 3.1 — Процесс съемки идентификационного документа при помощи мобильного устройства (в качестве документа используется макет идентификационной карты Германии).

Использование видеопотока позволяет решать задачи, недоступные для решения при анализе одиночной фотографии. Внешние условия съемки могут привести к тому, что распознаваемый объект сильно искажен на одиночном изображении [215]. Примером является блик от протяженного источника света, проявляющийся на глянцевой поверхности плоского объекта (см. рис. 3.1) Поскольку в видеопотоке геометрическое положение снимаемого объекта, как правило, меняется между кадрами, блик также “сдвигается”, что позволяет получить информацию о скрываемом объекте на другом кадре видеопотока. Существуют также важный класс объектов, детектирование и распознавание которых невозможно на одиночных снимках – к примеру, голографические эле-

менты защиты, которые на единичных изображениях могут быть неотличимы от бликов или рисунков [215].

Главным отличием видеопотока как цифрового образа распознаваемого объекта является тот факт, что для одного и того же объекта рассматривается последовательность наблюдений, которые отличаются между собой. Рассмотрим причины, по которым результат распознавания объекта может быть ошибочным, исходя из предположения, что система действует всегда детерминировано, т.е. в любой момент времени и при любых внешних условиях результаты распознавания одного и того же набора входных данных всегда совпадают. Таким образом любая ошибка является следствием неспособности системы различить объект того или иного класса. Ошибки распознавания можно условно разделить на три группы:

1. Ошибки, обусловленные несовершенством алгоритма распознавания, т.е. ошибки, являющиеся “внутренними” с точки зрения системы распознавания объектов и которые могут проявляться даже при идеальном функционировании других подсистем. Данный класс ошибок является безусловным атрибутом любой системы распознавания, вне зависимости от модели входа.
2. Ошибки, обусловленные дефектами надсистемы. Система распознавания одиночного изображения, как правило, является одной из подсистем некоторого комплекса, и изображения, подаваемые на вход системе распознавания формируются в результате действия других подсистем (см. рис. 3.2). Как следствие, могут возникнуть ошибки, связанные с несовершенством предшествующих подсистем. К примеру, пусть в результате разбиения изображения текстовой строки на изображения отдельных символов была допущена ошибка, в следствии которой положение правой границы изображения латинской буквы «P» было найдено некорректно, в результате чего на изображении буквы была утеряна перемычка между двумя горизонтальными штрихами. Изображение, полученное в результате, с точки зрения системы распознавания одиночного символа, может быть неотличимо от латинской буквы «F».
3. Ошибки, обусловленные шумом среды. Возникают такие ошибки в случае, если в условиях внешней среды, в которой находится распознаваемый объект, его изображение становится неотличимым от

изображения объекта другого класса. К примеру, предположим, что производится съемка фотографии документа, удостоверяющего личность, содержащего поле «Имя» с истинным значением «HANNA». Данное поле начертано на белом фоне и документ покрыт защитной глянцевой поверхностью. В момент съемки на документе проявился блик от внешнего источника света, полностью закрывший букву «H» и оставивший изображения остальных букв неизменными. Таким образом, изображения данного поля будет неотличимо от изображение поля «ANNA» на аналогичном документе.

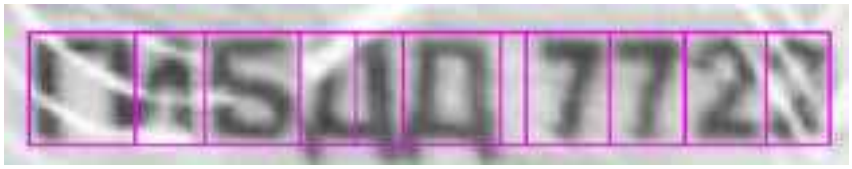


Рисунок 3.2 — Пример ошибочной сегментации текстовой строки на отдельные символы в условиях размытости изображения и дефектов, связанных с защитным голографическим слоем документа.

По отношению к системе распознавания одиночного изображения ошибки, связанные с шумом среды либо с дефектами надсистемы, являются следствием искажения входного изображения. Обладая возможностью использовать несколько наблюдений объекта можно ожидать, что влияние шума среды и дефектов надсистемы на эти наблюдения будут различны. Однако даже при фиксировании системы распознавания одиночного объекта, вне зависимости от дефектов надсистемы, остаются ошибки, обусловленные несовершенством модели классификации. Даже наиболее эффективные методы распознавания изображений, которые в ряде отдельных задач демонстрируют результаты, способные конкурировать с человеком [216—218], тем не менее, могут показывать неустойчивый результат при минимальных изменениях входного изображения [219; 220], даже если эти изменения касались всего лишь одного пикселя [221]. Так, даже используя наиболее точный метод распознавания, но обладая единственным входным изображением объекта, может быть невозможно отделить полезный сигнал от шума, влияние которого может кардинальным образом поменять результат. Рассматривая в качестве цифрового образа объекта не одиночное изображение, а видеопоток, появляется возможность уменьшить влияние ошибок за счет вариативности шума применительно к отдельным кадрам

видеопотока, которой не обладают классические системы распознавания объектов.

Одним из методов, позволяющих производить анализ множества изображений одной и той же сцены с целью уменьшить влияние шума оптической системы и дефектов, связанных с неконтролируемыми условиями съемками, является техника “супер-разрешения” – процесс получения изображения высокого разрешения из нескольких изображений того же объекта с более низким разрешением. Данной задаче уделялось большое внимание в литературе и предложено большое количество подходов, принимающих во внимание специфику финальной задачи обработки изображения и распознавания объекта или сцены [222]. Однако как было отмечено ранее, дальнейшая обработка полученного единого изображения объекта остается подверженной ошибкам алгоритма распознавания, в частности, неустойчивости сверточных нейронных сетей.

Рассматривая распознавания объекта на одиночном изображении, в рамках традиционных систем распознавания можно выделить две основные задачи – непосредственно задачу классификации образа объекта (т.е. определение принадлежности к одному классу из заранее заданного набора) и определение надежности распознавания (т.е. определение степени уверенности системы в собственном ответе с принятием решения об отказе, если в алфавите классификации нет специального “пустого” класса).

Применительно же к системе распознавания объекта в видеопотоке возникает ряд новых задач, в числе которых можно выделить следующие:

1. Задача предварительной оценки и выбора кадров для распознавания – поскольку в качестве графического представления изображения рассматривается не единственное изображения, а последовательность, возникает необходимость определять пригодность отдельных кадров для распознавания перед его непосредственным запуском;
2. Задача межкадрового комбинирования или *интеграции* – имея решение задачи классификации объекта, входом к которой является одно изображение целевого объекта, возникает задача аккумулировать единый результат классификации для множества изображений одного и того же объекта, результаты одиночной классификации которого, вообще говоря, могут быть противоречивы;
3. Задача принятия решения об остановке процесса распознавания – поскольку процесс распознавания в видеопотоке, вообще говоря, может

быть не ограничен во времени, возникает необходимость принимать на уровне системы распознавания решение о том, что захват новых изображений объекта следуюет прекратить и принять текущий аккумулированный результат распознавания за окончательный.

В данном разделе будет описана модель системы распознавания в видеопотоке и будут представлены формальные постановки новых задач, которые возникают в подобного рода системах.

3.2.2 Описание системы распознавания объектов в видеопотоке

Рассмотрим процесс распознавания объекта x на последовательности кадров видеопотока. Пусть задано множество классов $C = \{c_1, c_2, \ldots, c_M\}$. В случае распознавания текстового символа множеством классов может выступать какой-либо фиксированный алфавит. При рассмотрени задачи типизации страницы документа на изображении после локализации ее границ и проективного исправления, множеством классов может выступать коллекция типов страниц документов, доступных для дальнейшей обработки. Отдельно следует упомянуть, что иногда в задачах распознавания объектов и явлений допускается наличие "пустого класса", который должен быть ответом системы распознавания на входное изображение объекта, о котором системе не известно, либо на изображение, которое не содержит объекта.

Пусть задано изображение объекта $I(x)$ из некоторого множества всевозможных изображений $\mathbb{I}$ и в рамках модели взаимодействия системы распознавания с надсистемой или с пользователем (оператором) существует класс $c^* \in C$, к которому принадлежит объект x (истинный класс). Задачу распознавания изображения одиночного объекта можно рассматривать как задачу классификации – т.е. определения истинного класса. Результатом работы классификатора в общем виде можно считать всюду определенное отображение из множества классов C в множество оценок принадлежности: $r(I(x)) : C \to \mathbb{R}$. Учитывая, что множество классов C содержит ровно M элементов:

$$r(I(x)) = \{(c_1, q_1), (c_2, q_2), \ldots, (c_M, q_M)\}, \tag{3.1}$$

где $q_i \in \mathbb{R}, \quad i \in \{1, \ldots, M\}$ – вещественные оценки принадлежности объекта x к классам $c_i \in C$ при условии, что наблюдается изображение объекта $I(x)$.

В качестве окончательного решения классификации принимается класс, соответствущий максимальной оценке принадлежности. Таким образом, с точки зрения системы распознавания постановкой задачи классификации является поиск классификатора $r : \mathbb{I} \to \mathbb{R}^C$ из множества всевозможных классификаторов (или поиск набора параметров из множества параметров классификаторов какого-то определенного вида), который бы минимизировал ошибку классификации (несоответствие класса $\arg\max_{i=1}^{M} r(I(x))$ истинному классу c^*) и который не зависел бы от самого истинного класса c^*.

Рассмотрим теперь задачу распознавания объекта x в видеопотоке. Источником видеопотока является некоторое захватывающее устройство, предоставляющее последовательность различных кадров $I_1(x), I_2(x), I_3(x), \ldots, I_n(x)$, где $\forall k \in \{1, 2, \ldots, n\} : I_k(x) \in \mathbb{I}$, каждый из которых является изображением объекта x. Рассмотрим классификатор (семейство классификаторов) $R^{(n)} : \mathbb{I}^n \to \mathbb{R}^C$, ставящий в соответствие последовательности изображений длины n отображение из множества классов в оценки принадлежности. Таким образом можно формализовать задачу классификации объекта x на последовательности изображений как поиск классификатора (если последовательность имеет фиксированную длину) или семейства классификаторов (если последовательность изображений может иметь различную длину), который бы минимизировал ошибку классификации в том же смысле, как и в исходной задаче классификации одиночного изображения.

Следует обратить внимание, что при поиске семейства классификаторов возникает также и необходимость определить классификатор $R^{(1)} : \mathbb{I}^1 \to \mathbb{R}^C$, принимающий на вход последовательность изображений длины 1, постановка задачи для которого совпадает с постановкой задачи классификации на одиночном изображении. В связи с этим возникает естественный вопрос – поскольку решение задачи классификации одиночного изображения (т.е. поиска классификатора r) потребуется в любом случае, можно ли выразить все семейство классификаторов R, решающее задачу классификации объекта в видеопоследовательности, через одиночный классификатор r? Выразить семейство классификаторов видеопоследовательности R имея классификатор изображения r можно, поставив две дополнительные задачи: выбор изображений для

классификации и комбинирование результатов распознавания одиночных изображений.

Зададим функцию выбора $S : 2^{\mathbb{I}} \to 2^{\mathbb{I}}, \quad \forall X : S(X) \subseteq X$, ставящую в соответствие последовательности изображений ее подмножество, каждый элемент которого будет классифицироваться одиночным классификатором изображений. Также зададим функцию (семейство функций) $F^{(n)} : \left(\mathbb{R}^C\right)^n \to \mathbb{R}^C$, которая ставит множеству результатов классификации объекта единый аккумулированный результат классификации. Теперь семейство классификаторов последовательности изображений можно выразить следующим образом:

$$R^{(n)}(I_1(x), I_2(x), \ldots, I_n(x)) \stackrel{\text{def}}{=} F^{(m)}\left(\{r(I) \mid I \in S(\{I_1(x), \ldots, I_n(x)\})\}\right), \quad (3.2)$$

где $m = \text{Card}(S(\{I_1(x), \ldots, I_n(x)\}))$, тем самым сводя задачу классификации последовательности изображений к классификации одиночных изображений (из выбранных при помощи функции выбора S) и комбинирования результатов классификации при помощи функций комбинирования $F^{(n)}$.

Однако представление видеопотока как заранее заданной последовательности изображений объекта не в полной мере отражает сценарий распознавания объекта в реальном времени (к примеру, распознавание объекта при помощи мобильного устройства), поскольку такая модель предполагает в качестве входа лишь множество кадров и не предполагает изменения состояния системы в процессе съемки. Для того, чтобы более точно соответствовать процессу распознавания объекта в видеопотоке мобильного устройства рассмотрим процесс съемки изображений объекта x с дискретным временем. Представим видеопоток как генерирующуюся во времени последовательность изображений объекта x: зададим дискретное время $t = 0, 1, 2, \ldots$ и видеопоток, содержащий изображения наблюдаемого объекта $I_t(x) \in \mathbb{I}$. Подобная дискретная модель видеопотока, хоть и является упрощенной, соответствует принципам представления кодированного видеопотока в программных системах обработки видео.

Для определения системы распознавания объекта в видеопотоке, который генерируется независимо, необходимо определить модель обслуживания, которая бы являлась промежуточным слоем между видеопотоком и непосредственным потоком обрабатываемых системой распознавания изображений. Наиболее тривиальной является схема обслуживания, при которой изображения, генерируемые во время обработки системой распознавания предыдущих

изображения, сбрасываются. В случае, если возможно хранение коллекции изображений альтернативной моделью является схема обслуживания с буфером, позволяющим накапливать входящие изображения и выдавать их по запросу системы в произвольный момент времени, без ограничений, связанных с дискретизацией генерации изображений источником. С точки зрения непосредственно системы распознавания последовательности изображений набор методов и алгоритмов распознавания и интеграции результатов не зависят от схемы обслуживания, поэтому в дальнейшем будет предполагаться тривиальная схема со сбрасыванием изображений в периоды загрузки системы.

Пусть система распознавания поддерживает некоторое внутреннее состояние $\omega_t \in \Omega$, изменяющееся во времени. Время Δ_t, необходимое для получения обновленного результата после ввода очередного образа $I_t(x)$, в общем случае, является функцией от изображения и от внутреннего состояния системы: $\Delta_t = \Delta(I_t(x), \omega_t)$. Обратим внимание, что значение Δ_t может быть невычислимо в момент времени t. Результат распознавания объекта x, учитывающий информацию, содержащуюся в изображении, которое было захвачено в момент времени t, может быть доступен только в момент времени $T(t) = t + \Delta_t$.

В начальный момент времени $t = 0$ инициализировано внутреннее состояние системы ω_0. В каждый момент времени t происходит захват изображения $I_t(x)$. В случае, если система предполагает синхронную обработку изображений, то в момент захвата изображения $I_t(x)$ может происходит обработка другого изображения. Аналитически такое условие можно записать как $t < T(t_{\text{prev}})$, где t_{prev} – время захвата последнего изображения, поступившего в обработку. Однако, поскольку это условие может быть невычислимо в момент времени t, для целей описания модели можно считать, что внутреннее состояние системы ω_t хранит информацию о том, находится ли какое-либо изображение в обработке в момент времени t. Если в момент t система уже обрабатывает какое-либо изображение, то вновь полученно изображение $I_t(x)$ сбрасывается (в рамках тривиальной схемы обслуживания). В противном случае, изображение $I_t(x)$ поступает в обработку (в случае схемы распознавания (3.2) – поступает на вход классификатору одиночных изображений r). Результаты обработки изображения становятся доступными для вывода в момент времени $T(t)$.

Таким образом, в моменты времени $t \in \{0, 1, \ldots, T(0) - 1\}$ результат распознавания объекта не определен, а результат распознавания, который бы учитывал информацию, которая содержится в n различных (последовательно

захваченных) изображениях, может быть получен только в момент времени $T^n(0)$. При этом индексы изображений, поступающих в обработку, равны, соответственно, $0, T^1(0), T^2(0), \ldots, T^{n-1}(0)$, где под надстрочным знаком функции $T(t)$ подразумевается не возведение в степень, а множественная композиция функции. Пример описанной схемы представлен на рис. 3.3.

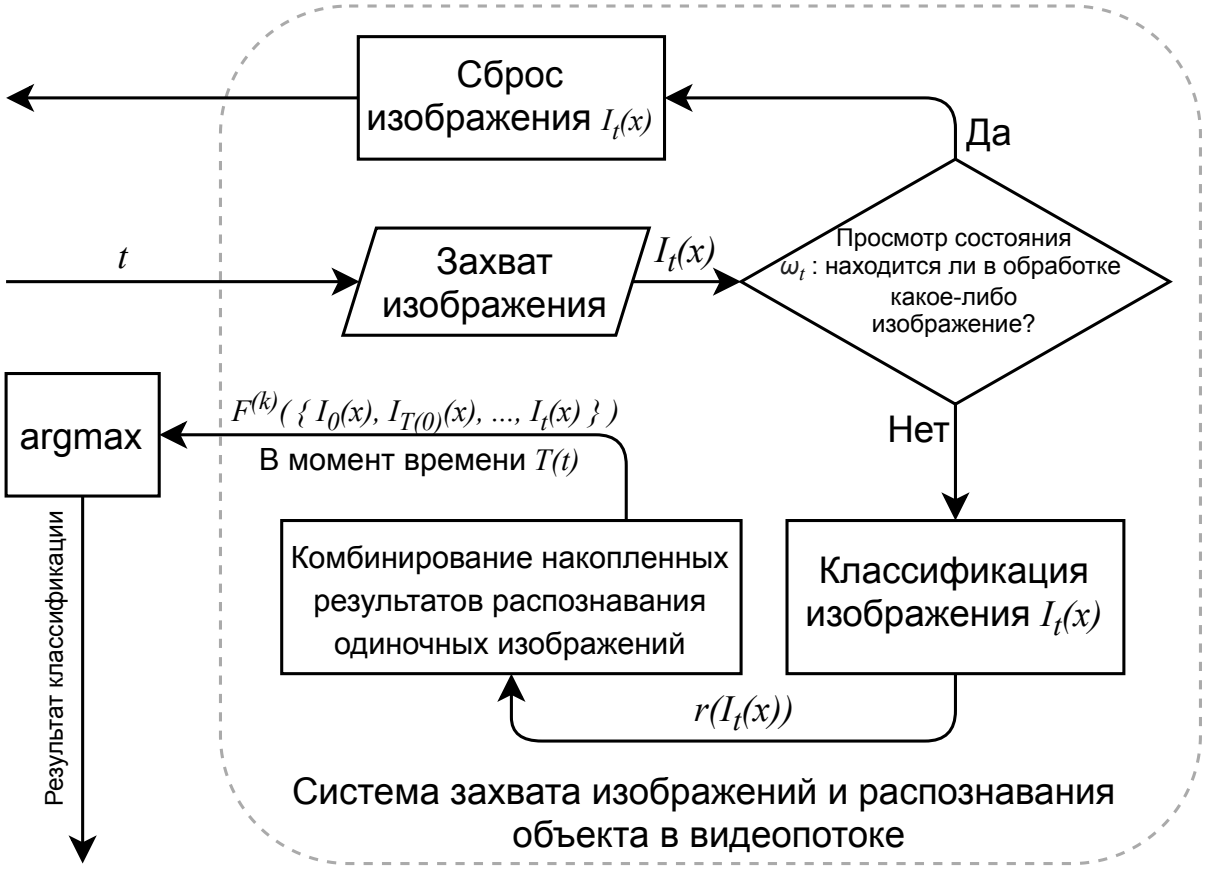


Рисунок 3.3 — Пример схемы системы распознавания объекта в видеопотоке с тривиальной моделью обслуживания, преобразующей видеопоток в последовательность обрабатываемых изображений и со схемой классификатора последовательности (3.2).

По сравнению с классическими системами распознавания, описанная система обладает рядом специфических свойств. В первую очередь необходимо отметить усиленное влияние производительности алгоритмов распознавания одиночного изображения на выход системы. Действительно, уменьшение времени Δ_t, необходимого для распознавания одного изображения $I_t(x)$, позволяет обработать большее количество информации об объекте x за одно и то же абсолютное время (т.е. за одно и то же время с точки зрения пользователя или оператора). Помимо этого, применительно к такой системе естественным образом возникает нетипичная для традиционных систем распознавания объектов

задача – задача остановки процесса распознавания. Такая задача заключается в принятии решения в момент времени $T(t)$ о том, что вновь полученный результат распознавания $R^{(n)}(I_0(x), I_{T^1(0)}, I_{T^2(0)}, \ldots, I_t(x)))$ можно считать окончательным и цикл захвата изображений можно прекратить. При распознавании сложных объектов, которые состоят из множества независимо распознаваемых объектов, решение об остановке процесса распознавания отдельных объектов также влияет на время Δ_t, необходимое для распознавания составного объекта, а значит и на количество информации, обрабатываемой в рамках общей системы. Таким образом, задача остановки (тесно связанная с задачей классификации последовательности изображений) является важным аспектом системы распознавания в видеопотоке, в особенности в рамках систем распознавания составных объектов, таких как текстовое поле или документ в целом. Правило остановки в общем виде можно формально представить в виде предиката, действующего на видеопоследовательности: $P : \mathrm{I}^* \to \{0, 1\}$. Истинность предиката влечет остановку процесса захвата и распознавания изображений:

$$P(\{I_1(x), I_2(x), \ldots, I_n(x)\}) = \begin{cases} 1: & \text{решение об остановке,} \\ 0: & \text{продолжение работы.} \end{cases} \tag{3.3}$$

Когда в рамках задачи распознавания объекта в видеопоследовательности, даже не принимая во внимание динамическую модель системы распознавания во времени, совместно с задачей максимизации точности классификации встает задача минимизации времени распознавания, то решение задачи остановки также может прямым образом использоваться для уменьшения абсолютного времени работы. К примеру, в рамках схемы классификации множества изображений (3.2), в случае, если определен предикат остановки P, моделируя последовательный процесс захвата изображений $I_1(x), I_2(x), \ldots, I_n(x)$, приходим к следующей схеме классификации:

$$\begin{aligned} R^{(n)}(I_1(x), I_2(x), \ldots, I_n(x)) &\stackrel{\text{def}}{=} \\ &\stackrel{\text{def}}{=} F^{(m)}\Big(\Big\{r(I)\Big| I \in S(\{I_1(x), \ldots, I_k(x)\}), \\ &\qquad k = \min\{i \leqslant n : P(I_1(x), \ldots, I_i(x)) = 1\}\Big\}\Big), \end{aligned} \tag{3.4}$$

где k – минимальная длина префикса последовательности изображений, на которой срабатывает предикат остановки, а $m = \mathrm{Card}(S(\{I_1(x), \ldots, I_k(x)\}))$.

3.2.3 Постановки задач

Формализуем приведенные выше задачи исходя из того, что основной целью системы распознавания объектов (на одиночных изображениях или в видеопотоке) является максимизация точности распознавания (т.е. максимизации доли корректных классификаций). При решении задачи распознавания одиночного изображения объекта положим, что задан набор объектов $X = \{x_1, x_2, \ldots, x_K\}$ мощности K и набор изображений $B_s = \{I_1(x_{b_1}), I_2(x_{b_2}), \ldots, I_H(x_{b_H})\}$ мощности H (тестовый набор), где b_h – индекс объекта из множества X для каждого $h \in \{1, 2, \ldots, H\}$ и каждое изображение $I_h(x_{b_h}) \in \mathbb{I}$ является образом объекта $x_{b_h} \in X$. Задано множество классов $C = \{c_1, c_2, \ldots, c_M\}$ и информация об идеальной принадлежности каждого объекта к соответствующему классу $c^* : X \to C$. Задачу распознавания объекта на одиночном изображении можно сформулировать как поиск классификатора $r : \mathbb{I} \to \mathbb{R}^C$, максимизирующего точность распознавания:

$$V_s(B_s) = \frac{1}{H}\left(\sum_{h=1}^{H}\left[\arg\max_{i=1}^{M} r(I_h(x_{b_h})) = c^*(x_{b_h})\right]\right) \to \max_r. \tag{3.5}$$

Для формализации задачи распознавания объекта в видеопоследовательности необходимо также ввести тестовый набор последовательностей изображений: пусть задано множество последовательностей изображений $B_m = \{J_1(x_{b_1}), J_2(x_{b_2}), \ldots, J_H(x_{b_H})\}$ мощности H, где b_h – индекс объекта из множества X для каждого $h \in \{1, 2, \ldots, H\}$ (так же как и в тестовом наборе одиночных изображений B_s), а $J_h(x_{b_h})$ является последовательностью изображений объекта $x_{b_h} \in X$ длины n_h:

$$J_h(x_{b_h}) = \{I_{h1}(x_{b_h}), I_{h2}(x_{b_h}), \ldots, I_{hn_h}(x_{b_h})\} \subset \mathbb{I}^{n_h}. \tag{3.6}$$

В общем виде задачу распознавания объекта на последовательности изображений теперь можно сформулировать как поиск семейства классификаторов $R^{(n)} : \mathbb{I}^n \to \mathbb{R}^C$, которые бы максимизировали точность распознавания:

$$V_m(B_m) = \frac{1}{H}\left(\sum_{h=1}^{H}\left[\arg\max_{i=1}^{M} R^{(n_h)}(J_h(x_{b_h})) = c^*(x_{b_h})\right]\right) \to \max_{R^{(n)}}. \tag{3.7}$$

В более частном случае, где схема распознавания последовательности изображений предполагает выбор изображений при помощи функции выбора

$S : 2^{\mathbb{I}} \to 2^{\mathbb{I}}$, классификации выборанных изображений при помощи классификатора $r : \mathbb{I} \to \mathbb{R}^C$ и комбинирования результатов при помощи семейства функций комбинирования $F^{(n)} : (\mathbb{R}^C)^n \to \mathbb{R}^C$ (3.2), задача формулируется схожим образом как поиск соответствующих функций S и $F^{(n)}$:

$$V_m(B_m) = \frac{1}{H}\left(\sum_{h=1}^{H}\left[\arg\max_{i=1}^{M} F^{(m_h)}(\{r(I)|I \in S(J_h(x_{b_h}))\}) = c^*(x_{b_h})\right]\right) \to \max_{F^{(n)},S}, \tag{3.8}$$

где $m_h = \mathrm{Card}(S(J_h(x_{b_h})))$.

Для формализации задачи распознавания объекта в видеопоследовательности в схеме с предикатором остановки (3.3) введем следующие обозначения. Пусть $J^{(k)}(x)$ – префикс последовательности изображений $J(x)$ длины $k \leqslant \mathrm{Card}(J(x))$. Обозначим через $n_P(J(x))$ количество изображений, которые будут обработаны системой распознавания до срабатывания правила остановки, определенного предикатом P:

$$n_P(J(x)) \stackrel{\text{def}}{=} \min\{k \leqslant \mathrm{Card}(J(x)) : P(J^{(k)}(x)) = 1\}. \tag{3.9}$$

Обозначим также как $J(x|P)$ префикс последовательности изображений $J(x)$, имеющий длину $n_P(J(x))$. С учетом правила остановки, основанного на предикате P, при обработке видеопоследовательности $J(x)$ на распознавание подаются только изображения из подпоследовательности $J(x|P)$. Для формализации задачи остановки воспользуемся моделью взаимодействия системы распознавания с пользователем, которая используется в задачах определения достоверности результата распознавания объекта [223; 224] и для оценки эффективности работы системы использует функционал, описанный в экономических терминах. Пусть w_c – стоимость ввода корректного результата распознавания объекта, w_e – стоимость ввода ошибочного результата, и w_f – стоимость распознавания одного изображения объекта. Тогда общая функция эффективности системы распознавания с правилом остановки, порождаемым предикатом P, может быть записана в виде средней стоимости работы системы:

$$W(B_m) = w_e + (w_c - w_e) \cdot V_m\left(\left\{J(x|P) \;\middle|\; J(x) \in B_m\right\}\right) + \\ + w_f \cdot \frac{1}{H}\left(\sum_{h=1}^{H} n_P(J_h(x_{b_h}))\right), \tag{3.10}$$

где V_m – точность распознавания видеопоследовательностей (3.7), вычисляемая по префиксам последовательностей, полученных после применения правила остановки, определяемого предикатом P.

Задача распознавания объекта в видеопослеовательности в схеме с остановкой (к примеру, в схеме (3.4)) может быть рассмотрена как задача минимизации общего функционала стоимости (3.10).

В последующих разделах будут рассмотрены конкретные примеры задач выбора изображений для распознавания, комбинирования результатов распознавания одиночных изображений и принятия решения об остановке, и будут приведены способы их решения применительно к конкретным системам распознавания.

3.3 Выбор кадров и комбинирование результатов распознавания

3.3.1 Возможные подходы к комбинированию

Как было описано в предыдущем разделе, при решении задачи классификации объектов в видеопотоке возникает задача выбора метода комбинирования информации, полученной с различных кадров. Подходы к комбинации межкадровой информации можно условно разделить на две группы:

1. методы, основанные на комбинировании изображений, ставящие в качестве цели получение единого представления объекта с более высоким “качеством”, что позволило бы использовать классификатор одиночного изображения и достигнуть более высокой ожидаемой точности;
2. методы, основанные на комбинировании результатов классификации одиночных изображений.

К методам первой группы можно отнести методы выбора наиболее информативного кадра [178; 225], методы “супер-разрешения”, получающие изображение с более высоким эффективным разрешением из нескольких кадров [226; 227], методы слежения за конкретными объектами на кадрах видеопоследовательности и комбинирования локальных областей конкретных объектов [228], а также методы компенсации конкретных локальных искажений, таких как сма-

зывание, путем замены локальных областей на соответствующие им области из других кадров, и методы, основанные на глубоком машинном обучении, принимающие на вход сразу множество изображений [180]. Здесь стоит отметить, что рассматривая системы мобильного распознавания, производящие съемку и вычисления на мобильном устройстве, в качестве входных данных можно использовать не только непосредственно кадры, полученные с камеры мобильного устройства, но и измерения с других сенсоров, таких, как акселерометр и гироскоп, однако даже для современных мобильных устройств ошибки измерения таких сенсоров могут быть настолько значительными, что использование таких данных для реконструкции изображений высокого качества может быть затруднено или невозможно, в особенности если конкретное устройство неизвестно заранее [229]. Методы первой группы, предполагающие комбинирование изображений, могут обладать высокой трудоемкостью, чувствительностью к геометрическим искажениям и быть трудно расширяемыми на случай неизвестной заранее длины последовательностей изображений.

Методы второй группы, описанные в литературе, используют правила комбинирования распределений, аналогично правилам, используемым при ансамблировании классификаторов (т.е. по правилу суммы, произведения, максимума, медианы и т.п. [230; 231]). В отличие от методов первой группы, на методы второй группы могут сильно влиять свойства оценок принадлежности, порождаемые классификаторами одиночных изображений, и, соответственно, выбор конкретного метода комбинирования может сильно зависеть от структуры и свойств классификатора.

Более того, выбор конкретного метода комбинирования результатов распознавания в видеопоследовательности может так же быть обусловлен требованиями, накладываемые в целом на систему распознавания. Рассмотрим в качестве примера задачу распознавания текстовой строки как последовательности символов. Сравним три подхода к комбинированию покадровых результатов распознавания:

1. Естественно предполагать, что на изображениях, на которых искажения вида «смаз» и «размытие» менее выражены, ожидаемая точность классификации будет выше [232]. Пусть задана функция $f : \mathbb{I} \to [0, 1]$, реализующая оценку сфокусированности изображения (к примеру, алгоритмом, описанным в работе [233]), где 0 означает наименее сфокусированное изображение, 1 – наиболее сфокусирован-

ное, с некоторым набором промежуточных градаций. Выберем из множества изображений объекта $\{I_1(x), I_2(x), \ldots, I_n(x)\}$ единственное изображение $I_f(x)$, обладающее максимальной оценкой фокусировки: $I_f(x) = \arg\max_{k=1}^{n} f(I_k(x))$. В качестве результата распознавания видеопоследовательности примем результат распознавания выбранного изображения:

$$R_1^{(n)}(I_1(x), I_2(x), \ldots, I_n(x)) = r(\arg\max_{k=1}^{n} f(I_k(x))), \tag{3.11}$$

где r – классификатор одиночного изображения.

2. Значения оценок принадлежности к классам каждого отдельного символа также можно использовать как апостериорный критерий качества – чем выше максимальное значение оценки принадлежности к классу, тем выше ожидаемая точность классификации. Пусть, в случае распознавания текстовой строки x, составленной из нескольких символов, на изображении $I(x)$, результатом является последовательность результатов классификации одиночных символов:

$$r(I(x)) = \left\{ \begin{array}{c} \{(c_{11}, q_{11}), (c_{12}, q_{12}), \ldots, (c_{1M}, q_{1M})\}, \\ \{(c_{21}, q_{21}), (c_{22}, q_{22}), \ldots, (c_{2M}, q_{2M})\}, \\ \ldots \\ \{(c_{K1}, q_{K1}), (c_{K2}, q_{K2}), \ldots, (c_{KM}, q_{KM})\} \end{array} \right\}, \tag{3.12}$$

где K – количество результатов классификации одиночных символов в ответе, M – количество классов, $c_{ij} \in C$ – класс (символ из некоторого алфавита), $q_{ij} \in [0, 1]$ – оценка принадлежности i-го символа к классу c_{ij}. Положим в качестве оценки качества результата распознавания строки x минимальное значение максимальной оценки принадлежности символа:

$$q(r(I(x))) \stackrel{\text{def}}{=} \min_{i=1}^{K} \max_{j=1}^{M} q_{ij}. \tag{3.13}$$

Выберем из множества результатов распознавания строки на одиночных изображениях $\{I_1(x), I_2(x), \ldots, I_n(x)\}$ единственный результат, обладающий максимальной оценкой качества, и в качестве результата распознавания видеопоследовательности примем этот результат:

$$R_2^{(n)}(I_1(x), I_2(x), \ldots, I_n(x)) = \arg\max_{k=1}^{n} q(r(I_k(x))). \tag{3.14}$$

3. В качестве третьего подхода рассмотрим прямое комбинирование одиночных результатов распознавания текстовой строки методом ROVER [234] с правилом усреднения в качестве правила комбинирования результатов распознавания одиночных символов:

$$R_3^{(n)}(I_1(x), I_2(x), \ldots, I_n(x)) = \mathrm{ROVER}(r(I_1(x)), \ldots, r(I_n(x))). \quad (3.15)$$

В первую очередь эмпирически проанализируем характеристики достигаемой точности распознавания строк в видеопоследовательностях с использованием трех описанных подходов. Для этого используем два открытых пакета данных: MIDV-500 [207] и MIDV-2019 [208]. Пакет данных MIDV-500 состоит из 500 видеопоследовательностей идентификационных документов, снятых при помощи мобильных устройств, с незначительными геометрическими искажениями. Пакет MIDV-2019 содержит 200 видеопоследовательностей, снятых в условиях низкого освещения (подмножество MIDV-2019-L) и с сильными проективными искажениями (подмножество MIDV-2019-D). Проанализируем четыре группы текстовых полей – номер документа, даты, имя держателя документа, написанное латинским алфавитом, и машиночитаемые зоны. При сравнении распознанных значений игнорировался регистр, а также игнорировались различия между цифрой «0» и латинской буквой «O». В качестве метрики качества использовалось нормализованное расстояние Левенштейна [235] до истинного значения распознаваемой строки. Для распознавания текстовых строк использовался алгоритм, описанный в работе [236].

Поскольку два из трех описанных подходов сводятся к выбору одного лучшего результата, для контроля также рассмотрим точность распознавания строки, которая достигалась бы при идеальном выборе (т.е. при выборе строки, заведомо наиближайшей к истинному значению).

На рис. 3.4, 3.5 и 3.6 представлены графики зависимости достигнутой точности распознавания строки с применением описанных выше подходов к комбинированию и с контрольным методом идеального выбора на пакетах данных MIDV-500, MIDV-2019-L и MIDV-2019-D, соответственно. Средние достигнутые значения нормализованного расстояния Левенштейна до истинного ответа после комбинирования на полях этих пакетов данных приведены в таблицах 7, 8 и 9.

Как можно увидеть из рис. 3.4 и таблицы 7, на пакете данных MIDV-500 наиболее высокое качество распознавания достигается при комбинировании

Таблица 7 — Среднее значение нормализованного расстояния Левенштейна до истинного ответа после комбинирования описанными методами для текстовых полей пакета данных MIDV-500

Метод комбинирования	Количество комбинированных кадров					
	5	10	15	20	25	30
Подход № 1 (выбор изображения)	**0,0833**	**0,0629**	**0,0573**	**0,0565**	**0,0554**	**0,0549**
Подход № 2 (выбор результата)	0,0999	0,0838	0,0788	0,0792	0,0786	0,0781
Подход № 3 (ROVER)	0,0995	0,0756	0,0689	0,0677	0,0680	0,0652
Выбор заведомо наиближайшего	0,0639	0,0408	0,0339	0,0321	0,0315	0,0309

Таблица 8 — Среднее значение нормализованного расстояния Левенштейна до истинного ответа после комбинирования описанными методами для текстовых полей пакета данных MIDV-2019-L

Метод комбинирования	Количество комбинированных кадров					
	5	10	15	20	25	30
Подход № 1 (выбор изображения)	**0,2331**	**0,1888**	**0,1751**	**0,1653**	**0,1588**	**0,1587**
Подход № 2 (выбор результата)	0,2685	0,2386	0,2300	0,2179	0,2281	0,2209
Подход № 3 (ROVER)	0,2392	0,1950	0,1833	0,1732	0,1675	0,1643
Выбор заведомо наиближайшего	0,1733	0,1231	0,1017	0,0868	0,0789	0,0750

Таблица 9 — Среднее значение нормализованного расстояния Левенштейна до истинного ответа после комбинирования описанными методами для текстовых полей пакета данных MIDV-2019-D

Метод комбинирования	Количество комбинированных кадров					
	5	10	15	20	25	30
Подход № 1 (выбор изображения)	**0,0519**	0,0459	0,0431	0,0447	0,0452	0,0463
Подход № 2 (выбор результата)	0,0568	0,0514	0,0552	0,0530	0,0538	0,0540
Подход № 3 (ROVER)	0,0546	**0,0412**	**0,0388**	**0,0365**	**0,0370**	**0,0371**
Выбор заведомо наиближайшего	0,0277	0,0171	0,0149	0,0131	0,0113	0,0103

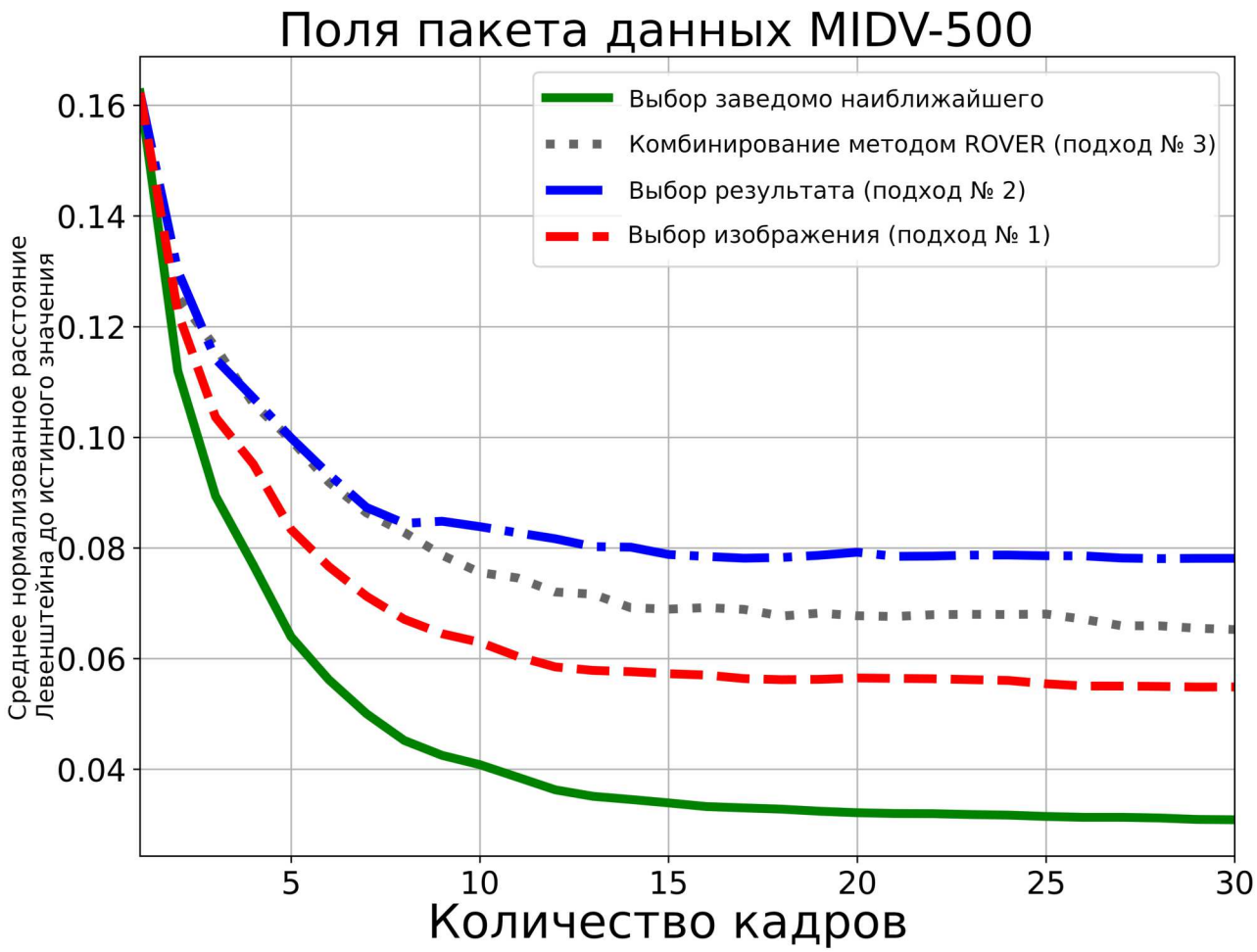


Рисунок 3.4 — Зависимость достигнутой точности распознавания строки от метода комбинирования, пакет данных MIDV-500.

методом выбора единственного изображения с максимальной оценкой сфокусированности (подход № 1). Тот же эффект наблюдается на подмножестве видеопоследовательностей в условиях низкого освещения из пакета данных MIDV-2019 (рис. 3.5, таблица 8), хотя различия между этим подходом и подходом полного комбинирования методом ROVER (подход № 3) уже становятся несущественными. При этом, на подмножестве видеопоследовательностей с сильными проективными искажениями из пакета данных MIDV-2019 (рис. 3.6, таблица 9) картина меняется: наиболее высокая точность распознавания достигается при полном комбинировании результатов распознавания методом ROVER. Тем самым, специфические свойства потока входных данных (или их распределение) могут влиять на выбор оптимальной стратегии комбинирования результатов распознавания в видеопоследовательности.

Стоит также заметить, что во всех трех случаях наилучшего качества достигала стратегия выбора одного результата (заведомо наиближайшего),

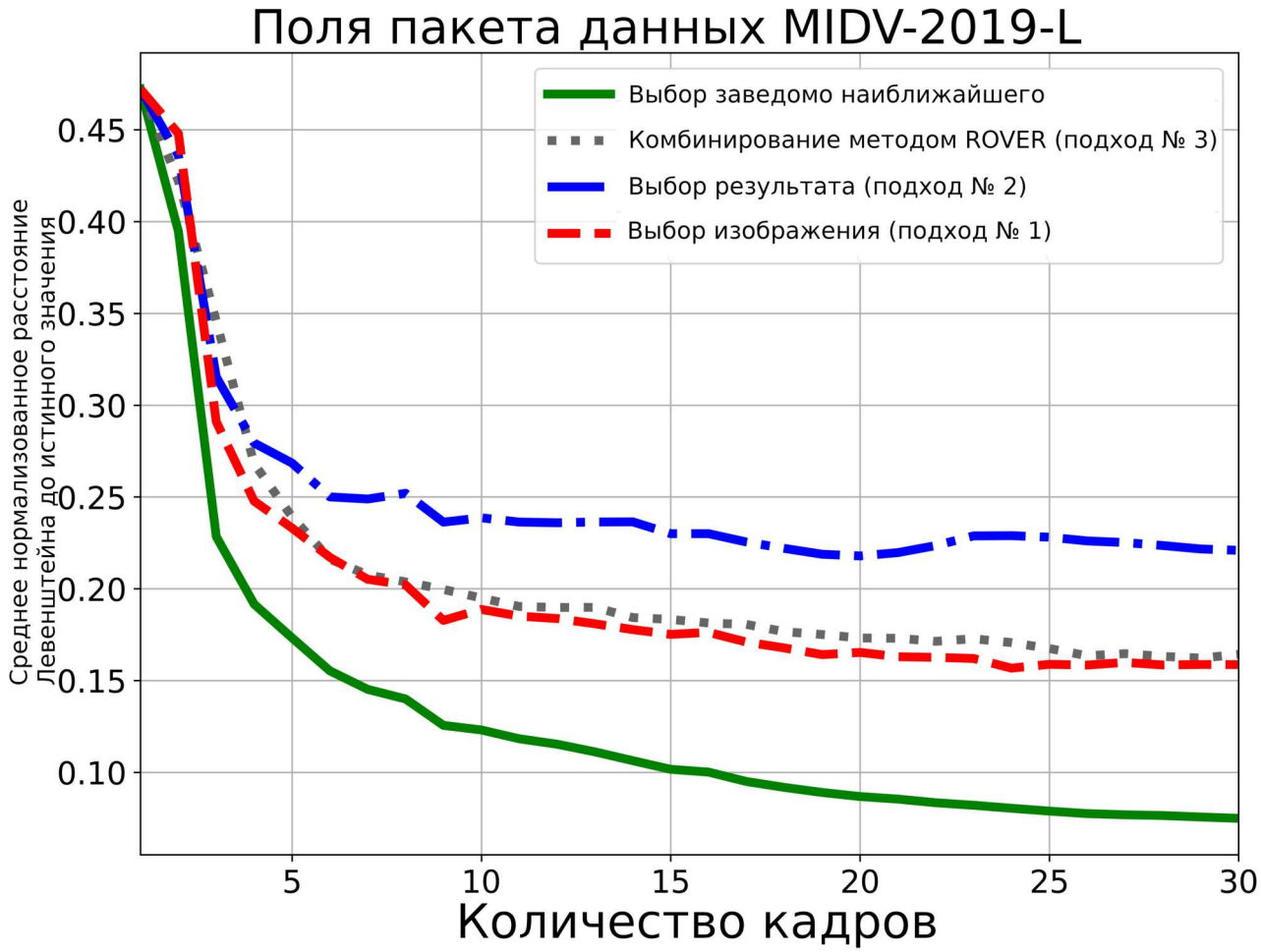


Рисунок 3.5 — Зависимость достигнутой точности распознавания строки от метода комбинирования, пакет данных MIDV-2019, подмножество видеопоследовательностей, снятых в условиях низкого освещения.

нереализуемая на практике. Это говорит о том, что все еще может существовать критерий выбора единственного наилучшего результата, который бы консистентно показывал наилучшие результаты на всех трех пакетах данных.

В литературе встречаются работы и исследования, рассматривающие и сравнивающие различные подходы к выбору изображений, определение надежности результата распознавания и, как следствие, построение априорных критериев оценки качества результата классификации объекта и различных методов слияния наборов оценок классификаторов для их ансамблирования или межкадрового комбинирования, как правило, на основе базовых статистик. Однако для решения задачи комбинирования результатов распознавания объекта в видеопотоке более содержательным образом необходимо понимать, каким образом распределяются оценки принадлежности в видеопотоке и каким образом построить эффективную предиктивную модель этих оценок.

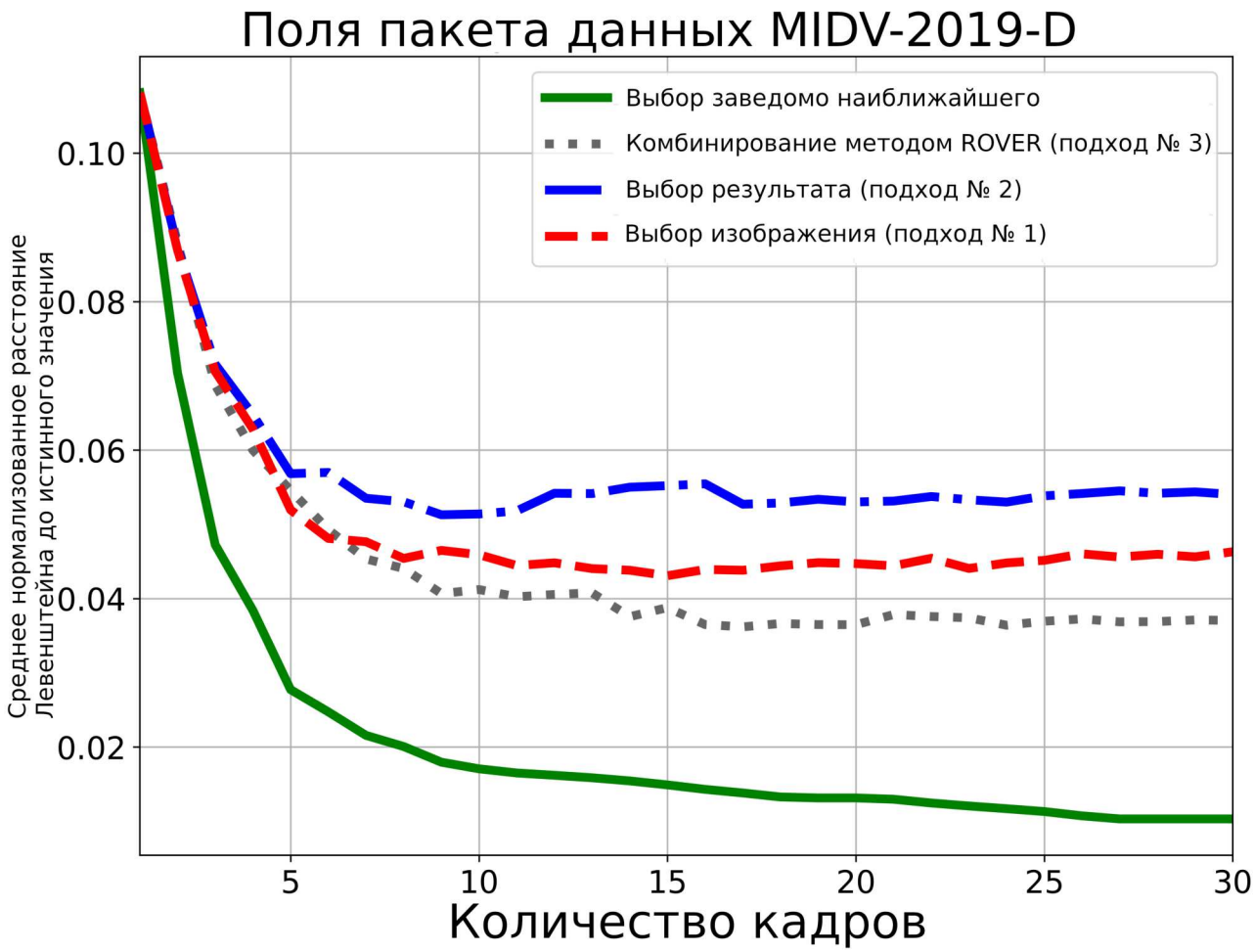


Рисунок 3.6 — Зависимость достигнутой точности распознавания строки от метода комбинирования, пакет данных MIDV-2019, подмножество видеопоследовательностей с значительными проективными искажениями.

3.3.2 Моделирование потока результатов распознавания объекта

В этом подразделе будет построена вероятностная модель, согласованная с результатами распознавания образов объектов в видеопоследовательностях, на примере задачи распознавания образов символов. Такая вероятностная модель может в дальнейшем использоваться для комбинирования результатов классификации одиночных объектов в видеопотоке как часть решения задачи распознавания в видеопотоке (3.8).

Для построения модели предположим, что целевым объектом распознавания является структурированный документ, состоящий из набора текстовых полей с заранее известными свойствами [215; 237; 238]. Для таких полей возможно отслеживание не только последовательности результатов распознавания

текстовых полей, но и анализ последовательности результатов распознавания одного знакоместа (символа) [239].

Пусть для некоторого документа есть последовательность кадров $\{I^k\}_{k=1}^K$. На каждом кадре I^k имеется поле $F(I^k)$, которое состоит, не ограничивая общности, из одного знакоместа A^k. Будем считать, что знакоместо A^k есть набор из n альтернатив $\{\langle s_i, X_i^k\rangle\}_{i=1}^n$, где s_i – код символа из некоторого алфавита Z (к примеру, $s_1 =$ «А», $s_2 =$ «Б» и т.д.), X_i^k – вероятностная оценка альтернативы (оценка принадлежности к классу), полученная при помощи классификатора одиночного символа на изображении I^k. Введем обозначение для вектора оценок $\mathbf{X}^k = (X_1^k, \ldots, X_2^k)^T$. Пусть $\mathbf{X}^k \in \mathbb{T}^n$, где $\mathbb{T}^n$ – симплекс.

$$\mathbb{T}^n = \left\{ (X_1, \ldots, X_n)^T : X_i \geqslant 0; i = 1, \ldots, n; \sum_{i=1}^n X_i = 1 \right\}. \qquad (3.16)$$

В дополнение будем предполагать, что результаты распознавания $\mathbf{X}^k$, $k = 1, \ldots, K$ есть выбора из независимых одинаково распределенных величин.

Назовем последовательность $\{\mathbf{X}^k\}_{k=1}^K$ потоком результатов распознавания объекта.

Рассмотрим теперь задачу моделирования (аппроксимации) эмпирического закона распределения потока результатов распознавания $\{\mathbf{X}^k\}_{k=1}^K$. Можно выделить четыре этапа решения поставленной задачи [240]:

1. выбор модели, т.е. выдвижение гипотезы о принадлежности выборки некоторому семейству распределений;
2. оценка параметров теоретического распределения;
3. оценка качества приближения;
4. проверка согласия между наблюдаемыми и ожидаемыми значениями с использованием статистических тестов.

В классической постановке задачи моделирования по имеющейся выборке независимых случайных величин с неизвестной плотностью распределения, принадлежащей некоторому семейству параметрических распределений, требуется построить оценки неизвестных параметров с использованием принципа максимального правдоподобия. Данная задача может не иметь решения в случае, если размерность вектора параметров оказывается большой и значительно превосходит выборочный объем. Поэтому для решения задачи будем рассматривать параметрическое семейство как комбинацию распределений с векторами

параметров меньшей размерности. Такой подход позволит получить оценки параметров при малых объемах выборок результатов распознавания.

Для реализации этого подхода приведем основные обозначения, а также некоторые определения и результаты из работы [241].

Рассмотрим два положительных случайных вектора $\mathbf{X} = (X_1, \ldots, X_n)$ и $\mathbf{Y} = (Y_1, \ldots, Y_n)$, связанных соотношениями $X_i = \dfrac{Y_i}{Y^+}$, где $Y^+ = \sum_{i=1}^{n} Y_i$. В литературе для обозначения вектора $\mathbf{X}$ используется термин *композиционные данные* (compositional data), а для вектора $\mathbf{Y}$ – базис. Заметим, что оценки $\mathbf{X}^i$ можно рассматривать как пример композиционных данных.

Довольно часто компоненты $\mathbf{X}$ могут быть сгруппированы согласно некоторому критерию однородности. В подобных случаях представлет интерес изучение суммарных показателей и относительных величин внутри каждой группы. Для формализации такого подхода принято использовать понятие *амальгамация (соединение)* и *подкомпозиция*, которые можно пояснить следующим образом. Пусть $a_0 = 0 < a_1 < \ldots < a_{C-1} < a_C = n$ набор индексов и

$$X_1, \ldots, X_{a_1} | X_{a_1+1}, \ldots, X_{a_2} | \ldots | X_{a_{C-1}+1}, \ldots, X_{a_C} \tag{3.17}$$

полное разбиение (порядка $c-1$) вектора $\mathbf{X}$ на c подмножеств. Исходя из разбиения (3.17), определим подкомпозицию с индексом i:

$$\mathbf{S}_i = (X_{a_{i-1}+1}, \ldots, X_{a_i})/X_i^+, \tag{3.18}$$

где $X_i^+ = X_{a_{i-1}+1} + \ldots + X_{a_i}, i = 1, \ldots, c$. Амальгамация представляет собой вектор $X^+ = (X_1^+, \ldots, X_c^+)$.

Известно [242], что для моделирования композиционных данных существенным является наличие у них свойства композиционной инвариантности. Базис $\mathbf{Y}$ композиционно инвариантен, если соответствующая композиция $\mathbf{X} = C(\mathbf{Y})$ не зависит от величины Y^+. Фактически, все версии понятий независимости, представленные в литературе, могут быть выражены в терминах подкомпозиций $\mathbf{S}_i, i = 1, \ldots, c$, и амальгамации X^+. Например, рассмотрим наиболее популярный случай разбиения порядка 1 ($c = 2$). Обозначим независимости символом $\perp$ и взаимную независимость переменных как $A \perp B \perp C$. Пусть $a_1 = m$. Тогда независимость разбиений означает, что $\mathbf{S}_1 \perp \mathbf{S}_2 \perp X^+$; подкомпозиционная инвариантность – $(\mathbf{S}_1, \mathbf{S}_2) \perp X^+$; нейтральность слева – $\mathbf{S}_1 \perp (\mathbf{S}_2, X^+)$; нейтральность справа – $\mathbf{S}_2 \perp (\mathbf{S}_1, X^+)$; подкомпозиционная независимость – $\mathbf{S}_1 \perp \mathbf{S}_2$.

Для моделирования композиционных данных одним из ключевых многомерных распределений является распределение Дирихле. Оно играет важную роль для представления пропорций. Это распределение имеет простой вид и обладает многими удобными математическими свойствами [243]. Вместе с тем, считается, что распределение Дирихле недостаточного гибкое. Поэтому, разными авторами были предложены обобщения распределения Дирихле [241; 242].

Случайный вектор $\mathbf{X} = (X_1, \ldots, X_n)^T \in \mathbb{T}^n$ имеет распределение Дирихле, если плотность распределения имеет следующий вид:

$$f_D(\mathbf{x}_n; \boldsymbol{\alpha}) = \frac{\Gamma(\alpha_+)}{\prod_{i=1}^n \Gamma(\alpha_i)} \prod_{i=1}^n x_i^{\alpha_i - 1}, \tag{3.19}$$

где $\boldsymbol{\alpha}$ – вектор положительных параметров, $\alpha_+ = \sum_{i=1}^n \alpha_i$.

Существует [243] простая связь между параметрами совместной плотности и маргинальных плотностей каждой из компонент $X_i \sim \text{Beta}(\alpha_i, \alpha_+ - \alpha_i)$.

Произведем следующее преобразование:

$$X_1^* = X_1, \quad X_i^* = \frac{X_i}{1 - \sum_{j=1}^{i-1} X_j}, \quad i = 2, \ldots, n-1. \tag{3.20}$$

Тогда для этих случайных величин справедливо:

$$X_1^* \sim \text{Beta}\left(\alpha_1, \sum_{j=2}^n \alpha_j\right) \text{ и}$$
$$X_i^* | X_1, \ldots, X_{i-1} \sim \text{Beta}\left(\alpha_i, \sum_{j=i+1}^n \alpha_j\right), \quad i = 2, \ldots, n-1. \tag{3.21}$$

Гибкое распределение Дирихле $FD^n(\boldsymbol{\alpha}, \mathbf{p}, \tau)$ было впервые предложено в работе [241]. Пусть $\mathbf{X} = (X_1, \ldots, X_n)^T \in \mathbb{T}^n$. Функция распределения вектора $\mathbf{X} \sim FD^n(\boldsymbol{\alpha}, \mathbf{p}, \tau)$ представляет собой конечную смесь распределений Дирихле:

$$FD^n(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{p}, \tau) = \sum_{i=1}^n p_i D^n(\mathbf{x}; \boldsymbol{\alpha} + \tau \mathbf{e}_i), \tag{3.22}$$

где $\mathbf{e}_i$ – вектор с нулевыми элементами, за исключением i-го, который равен единице, а плотность распределия имеет следующий вид:

$$f_{FD}(\mathbf{x}; \boldsymbol{\alpha}, \mathbf{p}, \tau) = \frac{\Gamma(\alpha_+ + \tau)}{\prod_{i=1}^n \Gamma(\alpha_i)} \left(\prod_{i=1}^n x_i^{\alpha_i - 1}\right) \left(\sum_{i=1}^n p_i \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i + \tau)} x_i^{\tau}\right), \tag{3.23}$$

где $\mathbf{x} \in \mathbb{T}^n$; $i = 1, \ldots, n$; $\alpha_i > 0, \alpha_+ = \sum_{i=1}^n \alpha_i$; $0 \leqslant p_i < 1, \sum_{i=1}^n p_i = 1$; $\tau > 0$.

Маргинальные распределения компонент вектора $\mathbf{X}$ можно представить следующим образом:

$$X_i \sim p_i \mathrm{Beta}(\alpha_i + \tau, \alpha_+ - \alpha_i) + (1 - p_i)\mathrm{Beta}(\alpha_i, \alpha_+ - \alpha_i + \tau), \quad i = 1, \ldots, n. \tag{3.24}$$

Будем говорить, что вектор $\mathbf{X} = (X_1, \ldots, X_n)^T \in \mathbb{T}^n$ имеет распределение Коннора-Мосиманна $CM(\alpha, \beta)$, если плотность его распределения можно представить следующим образом [244]:

$$f_{CM}(\mathbf{x}; \alpha, \beta) = \left[\prod_{i=1}^{n-1} \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} x_i^{\alpha_i - 1} \left(\sum_{j=i}^{n} x_j \right)^{\beta_{i-1} - (\alpha_i + \beta_i)} \right] x_n^{\beta_{n-1} - 1}, \tag{3.25}$$

где $\mathbf{x} \in \mathbb{T}^n$; $\alpha_i > 0, i = 1, \ldots, n, \beta_j > 0, j = 1, \ldots, n-1, \beta_0 = 0$.

Применим преобразование, аналогичное (3.20):

$$X_1^* = X_1, \quad X_i^* = \frac{X_i}{1 - \sum_{j=1}^{i-1}}, \quad i = 2, \ldots, n-1. \tag{3.26}$$

Соответственно, условные распределения определяются следующим образом:

$$X_1^* \sim \mathrm{Beta}(\alpha_1, \beta_1) \text{ и } X_i^* | X_1, \ldots, X_{i-1} \sim \mathrm{Beta}(\alpha_i, \beta_i), \quad i = 2, \ldots, n-1. \tag{3.27}$$

С распределением Дирихле тесно связано параметрическое семейство многомерных распределений Лиувилля. Для рассмотрения выберем из этого семейства распределение бета-Лиувилля. Пусть вектор $\mathbf{X} \in (0,1)^n$ имеет стохастическое представление $\mathrm{X} = R\mathrm{Y}$, где $R \perp \mathrm{Y}$, $R = \sum_{i=1}^n X_i$, $R \sim \mathrm{Beta}(a, b)$, $\mathbf{Y} = (Y_1, \ldots, Y_n)^T \in \mathbb{T}^n$, $\mathbf{Y} \sim \mathrm{Dir}(\alpha)$. Тогда будем говорить, что вектор $\mathbf{X}$ имеет распределение бета-Лиувилля. Плотность распределения бета-Лиувилля можно представить в следующем виде [245]:

$$f_{BL}(x_1, \ldots, x_n; a, b, \alpha_1, \ldots, \alpha_n) =$$

$$= \frac{\Gamma(a+b)\Gamma\left(\sum_{i=1}^n \alpha_i\right)}{\Gamma(a)\Gamma(b)\prod_{i=1}^n \Gamma(\alpha_i)} \times \left(\sum_{i=1}^n x_i \right)^{a - \sum\limits_{i=1}^{n} \alpha_i - 1} \left(1 - \sum_{i=1}^n x_i \right)^{b-1} \prod_{i=1}^n x_i^{\alpha_i - 1}. \tag{3.28}$$

Рассмотрим вектор оценок $\mathbf{X} \in \mathbb{T}^n$. Не ограничивая общности, будем считать, что выполняется $X_1 \geqslant \ldots \geqslant X_n$. Зададим некоторый критерий,

используя который можно разбить композицию $\mathbf{X}$ на две подкомпозиции $\mathbf{X}^{(1)} = (X_1, \ldots, X_m)^T$ и $\mathbf{X}^{(2)} = (X_{m+1}, \ldots, X_n)^T$. Например, к первой подкомпозиции отнесем все элементы, значения которых превосходят некоторый уровень $\mathcal{L}$, а все оставшиеся – ко второй. Обе подкомпозиции можно представить в следующем виде:

$$\mathbf{X}^{(1)} = X_+^{(1)} \frac{\mathbf{X}^{(1)}}{X_+^{(1)}}, \quad \mathbf{X}^{(2)} = X_+^{(2)} \frac{\mathbf{X}^{(2)}}{X_+^{(2)}}, \tag{3.29}$$

где $X_+^{(1)} = \sum_{i=1}^{m} X_i$; $X_+^{(2)} = \sum_{i=m+1}^{n} X_i$; $X_+^{(2)} = 1 - X_+^{(1)}$. Далее, введем новые переменные:

$$\mathbf{Z}^{(1)} = \frac{\mathbf{X}^{(1)}}{X_+^{(1)}}, \quad \mathbf{Z}^{(2)} = \frac{\mathbf{X}^{(2)}}{1 - X_+^{(1)}}, \quad R = X_+^{(1)}. \tag{3.30}$$

Тогда для $\mathbf{X}$ можно выписать выражение:

$$\mathbf{X} = \begin{pmatrix} \mathbf{X}^{(1)} \\ \mathbf{X}^{(2)} \end{pmatrix} = \begin{pmatrix} R \cdot \mathbf{Z}^{(1)} \\ (1 - R) \cdot \mathbf{Z}^{(2)} \end{pmatrix}, \tag{3.31}$$

где $\mathbf{X}^{(1)}$ – m-мерный вектор; $\mathbf{X}^{(2)}$ – $(n - m)$-мерный вектор.

Далее сформулируем предположения для переменных, входящих в первую часть выражения (3.31):

– *Предположение 1.* Пусть выполняется $\mathrm{Z}^{(1)} \perp \mathrm{Z}^{(2)} \perp R$.
– *Предположение 2.* Пусть $R \sim \mathrm{Beta}(a, b)$.
– *Предположение 3.* Пусть $\mathrm{Z}^{(2)} \sim \mathrm{Dir}(\boldsymbol{\alpha}^{(2)})$.
– *Предположение 4.* Пусть $\mathrm{Z}^{(1)} \sim \mathrm{Dir}(\boldsymbol{\alpha}^{(1)})$.
– *Предположение 5.* Пусть $\mathrm{Z}^{(1)} \sim FD(\boldsymbol{\alpha}^{(1)}, \mathbf{p}, \tau)$.
– *Предположение 6.* Пусть $\mathrm{Z}^{(1)} \sim CM(\boldsymbol{\alpha}^{(1)}, \beta^{(1)})$.

Для упрощения обозначений будем считать, что

$$\boldsymbol{\alpha} = \begin{pmatrix} \boldsymbol{\alpha}^{(1)} \\ \boldsymbol{\alpha}^{(2)} \end{pmatrix} = (\alpha_1, \ldots, \alpha_n)^T, \tag{3.32}$$

где $\boldsymbol{\alpha}^{(1)}$ – m-мерный вектор параметров; $\boldsymbol{\alpha}^{(2)}$ – $(n - m)$-мерный вектор параметров.

Используя предположения, построим три вероятностные модели распределения композиции $\mathbf{X}$.

Модель 1. Если выполняются предположения 1–4, то плотность распределения композиции $\mathbf{X}$ будет иметь следующий вид:

$$f_1(x_1,\dots,x_n;a,b,\alpha_1,\dots,\alpha_n) = \\ = \frac{\Gamma(a+b)\Gamma\left(\sum_{i=1}^m \alpha_i\right)\Gamma\left(\sum_{i=m+1}^n \alpha_i\right)}{\Gamma(a)\Gamma(b)\prod_{i=1}^n \Gamma(\alpha_i)} \times \\ \times \left(\sum_{i=1}^m x_i\right)^{a-\sum\limits_{i=1}^m \alpha_i - 1} \left(\sum_{i=m+1}^n x_i\right)^{b-\sum\limits_{i=m+1}^n \alpha_i - 1} \prod_{i=1}^n x_i^{\alpha_i - 1}. \quad (3.33)$$

Действительно, выпишем совместную плотность распределения $(\mathbf{Z}^{(1)}, R, \mathbf{Z}^{(2)})$:

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} r^{a-1}(1-r)^{b-1} \times \frac{\Gamma\left(\sum_{i=1}^m \alpha_i\right)}{\prod_{i=1}^m \Gamma(\alpha_i)} \prod_{i=1}^m [z_i^{(1)}]^{\alpha_i - 1} \times \\ \times \frac{\Gamma\left(\sum_{i=m+1}^n \alpha_i\right)}{\prod_{i=m+1}^n \Gamma(\alpha_i)} \prod_{i=m+1}^n [z_i^{(2)}]^{\alpha_i - 1}. \quad (3.34)$$

Далее необходимо перейти к координатам $x_1,\dots,x_n,x_+^{(1)}$. Вычислим значение Якобиана:

$$\mathbb{J}(z_1^{(1)},\dots,z_m^{(1)},r,z_{m+1}^{(2)},\dots,z_n^{(2)} \to x_1,\dots,x_n,x_+^{(1)}) = \\ = (x_+^{(1)})^{-m}(1-x_+^{(1)})^{-(n-m)}. \quad (3.35)$$

Таким образом, с учетом (3.30), (3.34) и (3.35), следует (3.33).

Модель 2. Если выполняются предположения 1–3 и 5, то плотность распределения композиции $\mathbf{X}$ будет иметь следующий вид:

$$f_2(x_1,\dots,x_n;a,b,\alpha_1,\dots,\alpha_n,p_1,\dots,p_n,\tau) = \\ = \frac{\Gamma(a+b)\Gamma\left(\sum_{i=1}^m \alpha_i + \tau\right)\Gamma\left(\sum_{i=m+1}^n \alpha_i\right)}{\Gamma(a)\Gamma(b)\prod_{i=1}^n \Gamma(\alpha_i)} \left(\sum_{i=1}^m x_i\right)^{a-\sum\limits_{i=1}^m \alpha_i - \tau - 1} \times \\ \times \left(\sum_{i=m+1}^n x_i\right)^{b-\sum\limits_{i=m+1}^n \alpha_i - 1} \prod_{i=1}^n x_i^{\alpha_i - 1} \left(\sum_{i=1}^m p_i \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i+\tau)} x_i^{\tau}\right). \quad (3.36)$$

Модель 3. Если выполняются предположения 1–3 и 6, то плотность распределения композиции $\mathbf{X}$ будет иметь следующий вид:

$$f_3(x_1,\ldots,x_n;a,b,\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_{m-1}) = \\ = \frac{\Gamma(a+b)\Gamma\left(\sum_{i=m+1}^{n}\alpha_i\right)}{\Gamma(a)\Gamma(b)\prod_{i=m+1}^{n}\Gamma(\alpha_i)}\left(\sum_{i=1}^{m}x_i\right)^{a-1}\left(\sum_{i=m+1}^{n}x_i\right)^{b-\sum\limits_{i=m+1}^{n}\alpha_i-1}\times \\ \times\prod_{i=1}^{m-1}\left[\frac{\Gamma(\alpha_i+\beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)}x_i^{\alpha_i-1}\left(\sum_{j=i}^{m}x_j\right)^{\beta_{i-1}-(\alpha_i+\beta_i)}\right]x_m^{\beta_{m-1}-1}\prod_{i=m+1}^{n}x_i^{\alpha_i-1}. \quad (3.37)$$

Для построения оценок распределения Дирихле применим принцип максимума правдоподобия. Будем полагать, что за оценки параметров принимаются те значения, которые обеспечивают максимум логарифмической функции правдоподобия:

$$L(\mathbf{x}_n^{(j)};\alpha) = \sum_{j=1}^{k}\log f_D(\mathbf{x}_n^{(j)};\alpha) = \\ = k\left\{\log\Gamma\left(\sum_{i=1}^{n}\alpha_i\right) - \sum_{i=1}^{n}\log\Gamma(\alpha_i) + \sum_{i=1}^{n}(\alpha_i-1)\log G_i\right\}, \quad (3.38)$$

где $G_i = \left(\prod_{j=1}^{n}x_{ji}\right)^{\frac{1}{k}}, \quad i = 1,\ldots,n.$

Известно [246], что функция L является выпуклой по α, поскольку распределение Дирихле относится к экспоненциальному семейству распределений. Это означает, что функция L является унимодальной, а максимум может быть найден простым поиском с использованием, например, метода Ньютона-Рафсона [243; 247].

Оценивание параметров гибкого распределения Дирихле будем рассматривать как задачу разделения конечной смеси распределений Дирихле, для решения которой используется EM-алгоритм [248; 249]. Предположим, что имеется k независимых наблюдений $\mathbf{x}_j$, $j = 1,\ldots,k$, каждое из которых представляет собой реализацию случайной величины с плотностью распределения, задаваемой соотношением (3.23). Далее зададим полный вектор данных $\mathbf{x}_c$:

$$\mathbf{x}_c = (\mathbf{x},\mathbf{v}) = (\mathbf{x}_1,\mathbf{v}_1,\ldots,\mathbf{x}_k,\mathbf{v}_k), \quad (3.39)$$

где вектор меток $\mathbf{v}_j = (v_{j1},\ldots,v_{jn})$ представляет неизвестные данные, $v_{ji} = 1$, если j-е наблюдение представляет собой реализацию случайной величины с

плотностью распределения, задаваемой i-й компонентой смеси распределения, и $v_{ji} = 0$, в остальных случаях.

Выпишем выражение для логарифмической функции правдоподобия с учетом (3.22) и (3.39):

$$\log L_c(\theta) = \sum_{j=1}^{k}\sum_{i=1}^{n} v_{ji}\left[\log p_i + \log f_D(\mathbf{x}_j; \alpha + \tau\mathbf{e}_i)\right], \tag{3.40}$$

где $\theta = (\alpha, \mathbf{p}, \tau)$; $f_D(\mathbf{x}_j; \alpha + \tau\mathbf{e}_i)$ – плотность распределения Дирихле.

Далее определим итерационный алгоритм типа EM решения задачи построения оценок параметров, который базируется на методе максимизации правдоподобия. Шаг алгоритма $s+1$ состоит в следующем.

E-step: при полученных на шаге s оценках параметров $\theta^{(s)} = (\alpha^{(s)}, \mathbf{p}^{(s)}, \tau^{(s)})$ и выборочных данных $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_k)$ получаем выражение для логарифмической функции правдоподобия (3.40):

$$Q(\theta; \theta^{(s)}) = \sum_{j=1}^{k}\sum_{i=1}^{n} p_i(\mathbf{x}_j; \theta^{(s)})\left[\log p_i + \log f_D(\mathbf{x}_j; \alpha + \tau\mathbf{e}_i)\right], \tag{3.41}$$

где $p_i(\mathbf{x}_j; \theta^{(s)})$ – апостериорная вероятность принадлежности наблюдения $\mathbf{x}_j$ к i-й компоненты смеси распределений при заданном значении $\theta^{(s)}$, которая определяется следующим образом:

$$p_i(\mathbf{x}_j; \theta) = \frac{p_i f_D(\mathbf{x}_j; \alpha + \tau\mathbf{e}_i)}{\sum_{r=1}^{n} p_r f_D(\mathbf{x}_j; \alpha + \tau\mathbf{e}_r)}, \quad i = 1, \ldots, n. \tag{3.42}$$

M-step: максимизируем выражение (3.41) с целью получения оценки $\theta^{(s+1)}$:

$$\theta^{(s+1)} = \arg\max_{\theta} Q(\theta; \theta^{(s)}). \tag{3.43}$$

В частности, $p_i^{(s+1)} = \frac{1}{k}\sum_{j=1}^{k} p_i(\mathbf{x}_j; \theta^{(s)})$, $i = 1, \ldots, n-1$, в то время как значения $\alpha^{(s+1)}$ и $\tau^{(s+1)}$ могут быть найдены с применением метода Ньютона-Рафсона.

Выполнение итераций происходит до тех пор, пока не будет достигнуто «достаточно малое» изменение наблюдаемого логарифмического правдоподобия (или в оценках параметров).

Получить оценки параметров распределения Коннора-Мосиманна можно с использованием свойства условных распределений (3.27) [244]. Выполним оценивание параметров бета-распределений $\alpha_r, \beta_r, r = 1, \ldots, n-1$ и полученные оценки возьмем в качестве оценок исходного распределения.

Теперь рассмотрим задачу выбора модели из множества конкурирующих моделей, которая дает наилучшее в некотором смысле приближение к характеристикам изучаемого потока результатов распознавания. В современном статистическом анализе для целей ранжирования моделей используется простой и эффективный инструмент – информационный критерий Акаике (AIC), который можно представить следующим образом:

$$Q_{AIC}(F^{(k)}, \hat{\theta}^{(k)}) = -2\sum_{j=1}^{n} \ln f^{(k)}(\mathbf{y}_j; \hat{\theta}^{(k)}) + 2q^{(k)}, \tag{3.44}$$

где $F^{(k)}$ – k-я модель; $f^{(k)}(\cdot)$ – плотность распределения для $F^{(k)}$; $\mathbf{y}_j$ – выборочные данные, $j = 1, \ldots, n$; $\hat{\theta}^{(k)}$ – вектор-параметр для $F^{(k)}$; $q^{(k)}$ – число параметров, от которых зависит $F^{(k)}$.

Выбор модели заключается в ранжировании моделей в соответствии со значениями критерия Q_{AIC} и предпочтении модели с его наименьшим значением.

Стоит заметить, что на основании критерия AIC можно построить правило разбиения композиции $\mathbf{X}$. Пусть для некоторого набора размерностей подкомпозиции $\mathcal{M} = \{m_{\min}, \ldots, m_{\max}\}$ выполняются предположения 1–3. Тогда искомую размерность для, например, модели 1 можно определить, как решение задачи

$$m^* = \arg\min_{m \in \mathcal{M}} \left(-2\sum_{k=1}^{K} \ln f_1(X_1^k, \ldots, X_n^k; \hat{a}, \hat{b}, \hat{\alpha}_1, \ldots, \hat{\alpha}_m, 1, \ldots, 1) + 2m \right). \tag{3.45}$$

Легко заметить, что полученные результаты (3.33), (3.36) и (3.37) различаются типом распределения, описывающего величину $\mathbf{X}^{(1)}/X_+^{(1)}$. Поэтому, ограничимся вычислением значений частичного критерия Q_{AIC} (см. таблицу 10).

Теперь остановимся на вопросе, насколько хорошо предложенные модели согласуются с имеющимися данными. Прежде чем перейти к построению объективных количественных оценок, полезно подвергнуть полученные модели процедуре неформальной графической диагностики, т.е., провести сравнение выборочных данных с параметрической моделью графическими методами. Для проверки характера выборочного распределения построим так называемый «график квантилей». На таком графике изображаются квантили двух распределений – эмпирического и теоретического. При хорошем соответствии

Таблица 10 — Сравнение моделей, описывающих результаты распознавания образов символов в видеопотоке

Модель	Значение Q_{AIC}
Модель 1 (распределение Дирихле)	−221,73
Модель 2 (гибкое распределение Дирихле)	−228,50
Модель 3 (распределение Коннора-Мосиманна)	−234,39

теоретического распределения проверяемым данным точки на графике располагаются вдоль прямой линии. На рисунках 3.7-3.10 приведены графики для предложенных моделей.

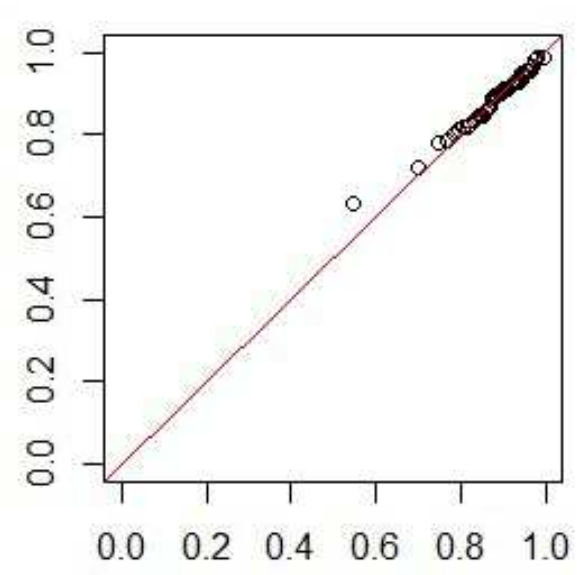


Рисунок 3.7 — График K-K для переменной $X_{+}^{(1)}$ при проверке предположения 2.

Прежде чем перейти к формулировкам и проверке гипотез о согласии, приведем формулу «стандартизации» случайной величины X, имеющей распределение $\mathrm{Beta}(a, b)$:

$$X^* = I(X; a, b) = \frac{B(X; a, b)}{B(a, b)} = \frac{1}{B(a, b)} \int_0^X t^{a-1}(1-t)^{b-1} dt, \tag{3.46}$$

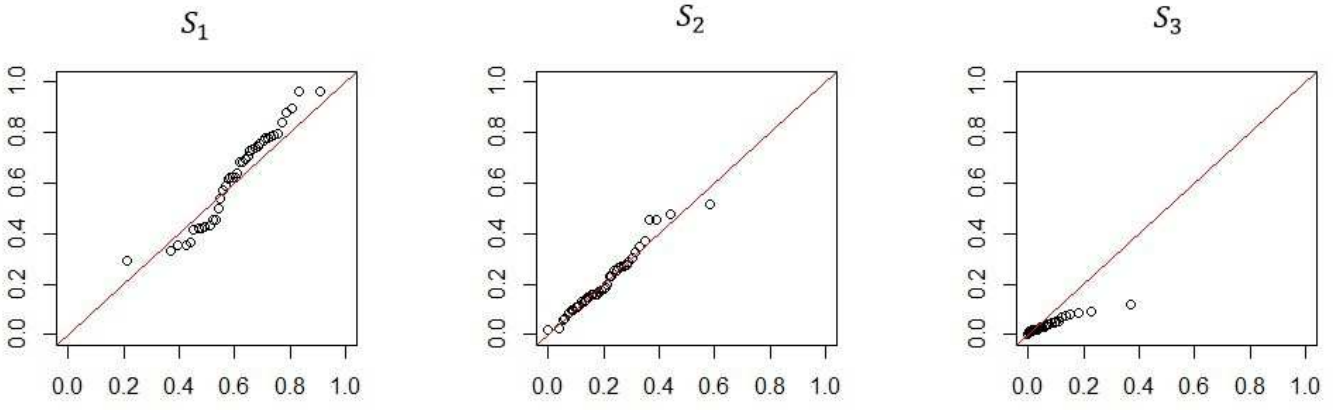


Рисунок 3.8 — График К-К для $\mathbf{X}^{(1)}$ для модели 1.

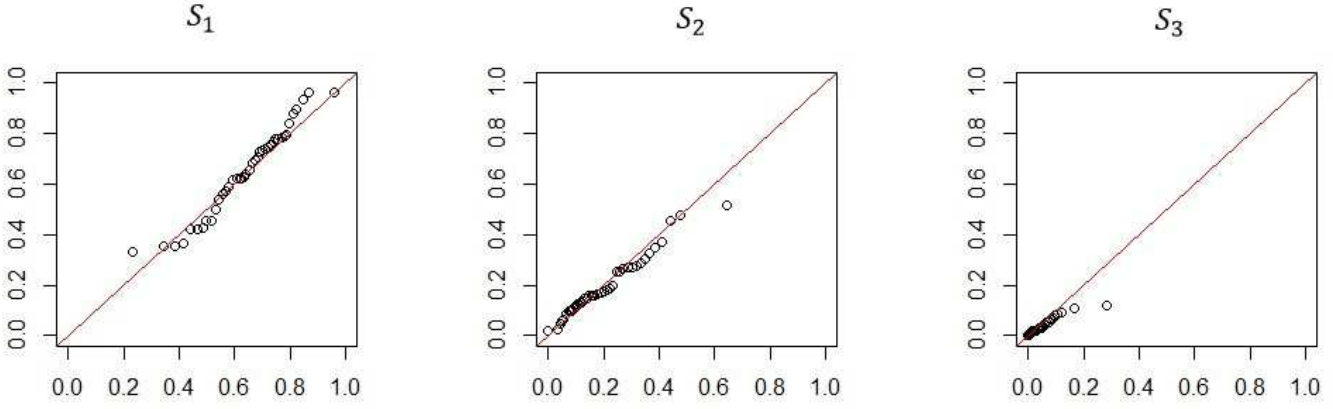


Рисунок 3.9 — График К-К для $\mathbf{X}^{(1)}$ для модели 2.

где $I(X; a, b)$ – регуляризованная неполная бета-функция; $B(X; a, b)$ – неполная бета-функция; $B(a, b)$ – бета-функция. Известно [250], что трансформированная случайная величина обладает следующим свойством:

$$X^* \sim \text{Beta}(1, 1). \tag{3.47}$$

Нам понадобится аналогичная (3.46) формула трансформации для случайной величины Y, отвечающей смеси бета-распределений

$$Y \sim p \cdot \text{Beta}(a_1, a_2) + (1 - p) \cdot \text{Beta}(b_1, b_2). \tag{3.48}$$

Введем следующее преобразование:

$$Y^* = p \cdot I(Y; a_1, a_2) + (1 - p) \cdot I(Y; b_1, b_2), \tag{3.49}$$

тогда $Y^* \sim \text{Beta}(1, 1)$.

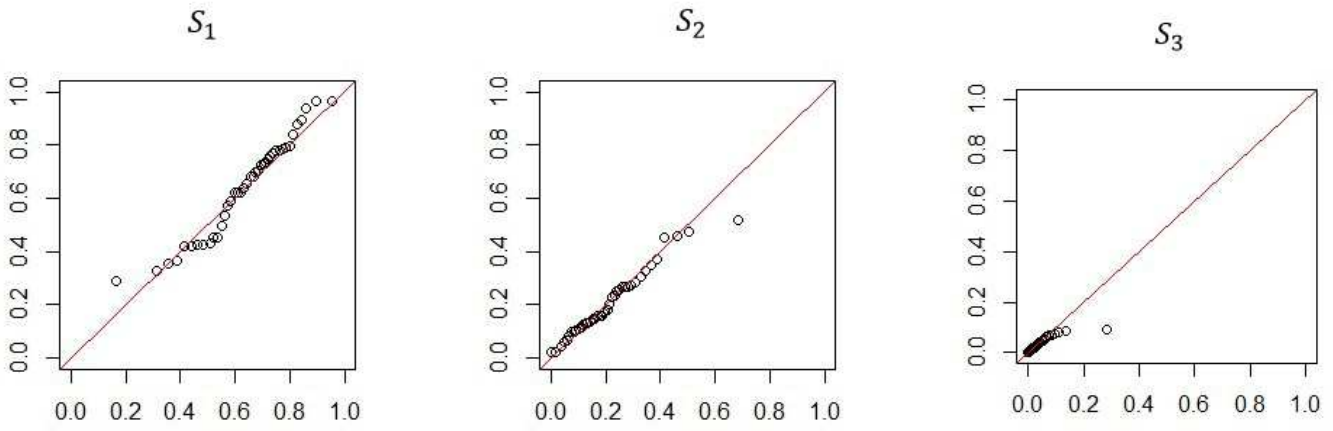


Рисунок 3.10 — График K-K для $\mathbf{X}^{(1)}$ для модели 3.

Формально, все необходимые для наших целей гипотезы согласия имеют общий вид. Пусть $X_1, \ldots, X_n$ – последовательность независимых одинаково распределенных случайных величин с функцией распределения F. Тогда проверяется нулевая гипотеза

$$H_0 : F(x) = I(x; 1, 1). \tag{3.50}$$

Для проверки соответствия выборочного распределения теоретическому закону будем использовать классический критерий согласия Андерсона-Дарлинга [251], базирующийся на статистике [252]:

$$S_\Omega = -n - 2\sum_{i=1}^{n}\left\{\frac{2i-1}{2n}\ln F(x_i, \theta) + \left(1 - \frac{2i-1}{2n}\right)\ln\left(1 - F(x_i, \theta)\right)\right\}. \tag{3.51}$$

Укажем, что большие значения статистики S_Ω указывают на плохое соответствие. Распределение статистики S_Ω быстро приближается к асимптотическому распределению, имеющему вид [253]

$$a2(S) = \frac{\sqrt{2\pi}}{S}\sum_{j=0}^{\infty}(-1)^j\frac{\Gamma(j+\frac{1}{2})(4j+1)}{\Gamma(\frac{1}{2})\Gamma(j+1)}\exp\left\{-\frac{(4j+1)^2\pi^2}{8S}\right\}\times$$
$$\times\int_0^\infty \exp\left\{\frac{S}{8(y^2+1)} - \frac{(4j+1)^2\pi^2 y^2}{8S}\right\}dy. \tag{3.52}$$

Для практических целей это распределение может использоваться при условии, что размер выборки больше 5 [251].

Наконец, рассмотрим схему проверки гипотез. Возьмем предположение 4. так как $\mathbf{Z}^{(1)} \sim \mathrm{Dir}(\boldsymbol{\alpha}^{(1)})$, то, используя свойства (3.20) и (3.21) распределения Дирихле, можно получить:

$$\begin{gathered} Z_1^* = Z_1^{(1)}, \quad Z_i^* = \frac{Z_i^{(1)}}{1 - \sum_{j=1}^{i-1} Z_j^{(i)}}, \\ Z_1^* \sim \mathrm{Beta}\left(\alpha_1^{(1)}, \sum_{j=2}^{m} \alpha_j^{(1)}\right), \\ Z_i^* | Z_1^{(1)}, \ldots, Z_{i-1}^{(1)} \sim \mathrm{Beta}\left(\alpha_i^{(1)}, \sum_{j=i+1}^{m} \alpha_j^{(1)}\right), \quad i = 2, \ldots, m-1. \end{gathered} \tag{3.53}$$

Применим к Z_i^* преобразование (3.46) и получим:

$$\begin{gathered} Y_1 = I\left(Z_1^*; \alpha_1^{(1)}, \sum_{j=2}^{m} \alpha_j^{(1)}\right), \\ Y_i = I\left(Z_i^*; \alpha_i^{(1)}, \sum_{j=i+1}^{m} \alpha_j^{(1)}\right), \quad i = 1, \ldots, m-1. \end{gathered} \tag{3.54}$$

Тогда, учитывая (3.47), можно сформулировать гипотезу согласия в следующем виде:

$$H_0 : F_i(y) = I(y; 1, 1), \tag{3.55}$$

где $F_i(y)$ – функция распределения случайной величины Y_i, $i = 1, \ldots, m-1$; $I(y; 1, 1)$ – регуляризованная неполная бета-функция с параметрами $(1, 1)$.

Результаты проверки согласия для двух проверочных альтернатив результатов распознавания текстового символа представлены в таблице 11.

Как видно, при задании уровня значимости $\alpha < 0{,}18$ нет оснований для отклонения проверяемых гипотез по критериям согласия по всем моделям.

Таким образом, были построены новые вероятностные модели, описывающие результаты распознавания символов текстовых полей документов в видеопоследовательности. Введено понятие потока результатов распознавания. Рассмотренные модели предполагают, что результат распознавания знакоместа в поле документа можно представить в виде композиции случайных величин и случайных векторов. Для различных предположений получены выражения плотностей распределений результатов распознавания знакоместа и приведены методы оценивания необходимых параметров. Для ранжирования моделей был использован информационный критерий Акаике. Проведены проверки, которые

Таблица 11 — Результаты проверки согласия моделей для двух проверочных альтернатив результатов распознавания текстового символа

Альтернатива	Модель	Значение S_Ω	p-value
«К»	1	1,4679	0,1844
	2	0,6732	0,5807
	3	0,5187	0,7273
«Е»	1	1,0981	0,3095
	2	0,5169	0,7286
	3	0,1396	0,9992

подтвердили адекватность вероятностных моделей. Полученные при моделировании потока результатов распознавания параметры можно использовать для комбинирования результатов классификации одиночных символов, решая тем самым задачу распознавания объекта в видеопотоке по схеме (3.2).

Адаптировав тот или иной алгоритм комбинирования результатов покадрового распознавания объекта, который бы аккумулировал и обновлял результат, для создания системы распознавания в видеопотоке теперь необходимо решить задачу остановки процесса распознавания – т.е. задачу принятия решения о том, что процесс захвата новых кадров можно прекратить и вывести текущий аккумулированный результат.

3.4 Проблема остановки распознавания

Решение задачи остановки процесса распознавания объекта в видеопотоке, в первую очередь, возникает при необходимости ограничить число распознаваемых изображений. Рассмотрим, к примеру, процесс распознавания объекта в видеопотоке мобильного устройства. С одной стороны, камера мобильного устройства позволяет вводить видео с высокой скоростью. С другой стороны, микропроцессор мобильного устройства не позволяет распознавать документы с такой же скоростью. Например, почти на всех современных мобильных устройствах возможна скоростью оцифровки 30 кадров секунду с разрешением

FullHD, однако скорость распознавания образов сложных объектов (к примеру, структурированных идентификационных документов) может составлять от 0,3 до 8 кадров в секунду, в зависимости от процессорной архитектуры мобильного устройства. В то же самое время в случаях отсутствия дефектов оцифровки и “дрожания” рук точное распознавание происходит на первых же кадрах видеопоследовательности. Разумеется, возможны случаи, когда последовательность из нескольких одинаково распознанных кадров перемежается множеством кадров, распознанных иначе. Возникает задача детектирования ситуации, когда можно с высокой уверенностью остановить наблюдение за распознаванием, получив надежный результат. Это можно назвать задачей ограничения наблюдений.

3.4.1 Метод ограничения количества наблюдений на основе анализа популяций

Первый подход, который можно использовать для решения такой задачи, основан на анализе популяций результатов распознавания объектов, полученных на одиночных изображениях. Для формализации этого подхода будем использовать понятие популяции, известное давно и описанное в классических работах Р. Бекхофера и Ш. Гупты [254]. Интерес к анализу популяций не ослабевает и в последнее время (см. работы [255—258]), хотя применения этого подхода к методам анализа результатов распознавания объектов в видеопотоке ранее в литературе не встречалось.

Рассмотрим процесс распознавания документа типа «Паспорт РФ» в видеопоследовательности. Пусть для каждого отдельно взятого документа есть набор из n распознаных кадров $I_1, \ldots, I_n$, с m распознанными полями $F^1(I_i), \ldots, F^m(I_i), \quad i \in \{1, \ldots, n\}$. Каждое поле представляет собор набор знакомест $F^k(I_i) = \{A_1^{k,i}, \ldots, A_l^{k,i}\}$, где $i \in \{1, \ldots, n\}$, $k \in \{1, \ldots, m\}$, а $l = l(F^k(I_i))$ – количество распознанных знакомест в поле (если поле не распознано, равно нулю). Каждое знакоместо $A_q^{k,i}$ состоит из кода символа $s_q^{k,i}$ из кириллического алфавита и оценки надежности распознавания $w_q^{k,i} \in [0, 1]$. Оценка надежности обычно базируется как на оценке принадлежности к классу, определенную классификатором, так и на факторах, априорно затрудняющих распознавание, например, на детекторах дефектов оцифровки [239].

Возможно также представление знакоместа в виде альтернатив, то есть в виде упорядоченного набора оценок соответствия всем символам алфавита распознавания. В таком представлении имеет смысл рассматривать первые и правильные альтернативы, то есть альтернативы с наибольшей оценкой и альтернативы, соответствующие известному заранее правильному коду символа.

Рассмотрим следующую постановку задачи распознавания одного поля F^k видеопоследовательности. Для последовательности кадров видеопоследовательности $I_1, \ldots, I_n$ соответствующего одному документу, для которого заранее известно содержимое полей, распознается образ текстового поля на каждом из кадров. По результатам распознавания на каждом кадре $F^k(I_1), \ldots, F^k(I_i)$, т.е. по наборам знакомест $A_1^{k,i}, \ldots, A_l^{k,i}$, где $i \in \{1, \ldots, n\}$, $l = l(F^k(I_i))$, строится комбинированный результат, т.е. комбинированный набор знакомест $A_1^k, \ldots, A_L^k$. Предположим, что в наборе распознанных кадров существуют такие кадры I_i и I_j, для которых в поле F^k совпадает количество знакомест $l^* = l(F^k(I_i)) = l(F^k(I_j))$ и для каждого знакоместа выполняется равенство:

$$s_q^{k,i} = s_q^{k,j}, \quad q \in \{1, \ldots, l^*\}. \tag{3.56}$$

Обозначим через $J = \{1, 2, \ldots, j^{(1)}\}$ множество всех индексов кадров, для которых совпадают коды символов для каджого знакоместа в поле F^k, т.е. выполняется (3.56). Объединим результаты распознавания поля для таких кадров в совокупность следующим образом. Обозначим $s_q^k = s_q^{k,j}$, для $j \in J$ и $q \in 1, \ldots, l^*$. Совокупность $G = \left\{\{s_1^k, w_1^{k,j}, j \in J\}, \ldots, \{s_1^k, w_1^{k,j}, j \in J\}\right\}$ назовем популяцией, а $l^* = l^*(G)$ – объемом популяции. Очевидно, что для конкретной видеопоследовательности может быть получена как одна единственная популяция G, так и набор популяций $G^{(1)}, \ldots, G^{(i_p)}, \ldots$.

Можно определить несколько показателей, характеризующих надежность распознавания кадров, входящих в популяцию:

$$R_1 = 1 - \prod_{j \in J} \left(1 - \min_{q=1}^{l^*} w_q^{k,j}\right). \tag{3.57}$$

$$R_2 = 1 - \prod_{j \in J} \left(1 - \prod_{q=1}^{l^*} w_q^{k,j}\right). \tag{3.58}$$

Далее рассмотрим последовательную обработку поступающих распознанных кадров. После отнесения очередного распознанного кадра к соответствующей популяции вычислим для каждой из популяций какой-либо из показателей

(например, R_1) и построим вариационный ряд $R_1^{i_1} \geqslant \ldots \geqslant R_1^{i_p}$. Вычисления прекращаются в случае, когда выполнено несколько условий, использующих заранее заданные пороги γ_1, γ_2, n_3:

1. $R_1^{i_1} - R_1^{i_2} > \gamma_1$,
2. $R_1^{i_1} > \gamma_2$,
3. объем популяции с индексом i_1 превосходит n_3.

В качестве наилучшего результата берется популяция с индексом i_1. Для ряда, содержащего только одну популяцию, проверяется выполнение двух последних условий.

Таким образом, весь набор кадров можно разбить на популяции, и, выбрав наилучшую популяцию, удовлетворяющую условиям 1–3, ограничить объем наблюдений за поступающими кадрами.

Проверим предложенный алгоритм ограничения наблюдений на тестовом наборе данных с хорошим и средним качеством видеосъемки паспортов РФ, т.е. видеопоследовательности могут иметь незначительные дефекты. Набор состоял из 50 видеопоследовательностей объемом от 17 до 50 кадров. Анализировалось распознавание семи полей (ФИО, номер и серия паспорта, дата и место рождения), одно поле (место рождения) состояло из нескольких слов, остальные поля – из одного слова.

Для оценки предложенного алгоритма ограничения наблюдений использовалось две характеристики:

- *точность* распознавания видеопоследовательности с ограничением объема наблюдений, определяемая как доля результатов распознавания, совпадающих со значением поля, полученным в отсутствии обработки всей видеопоследовательности без ограничения наблюдений;
- *число кадров*, необходимое для остановки алгоритма.

В таблице 12 приведены усредненные характеристики алгоритма на тестовом наборе. Для критериев R_1 и R_2 характеристики алгоритма практически не отличаются, что объясняется достаточной точностью распознавания последовательностей $F^k(I_i)$ с помощью механизма нейронных сетей. В основном ошибки представляют собой случаи отсутствия правильных результатов распознавания, а также ошибки поиска границ поля. Таким образом, предложенный алгоритм дает хорошую точность на данных с ограничениями по качеству распознавания.

Из таблицы 12 видно, что разработанный алгоритм терпит существенные неудачи при обработке сложных полей, состоящих из нескольких слов. В табли-

Таблица 12 — Характеристики алгоритма с критериями R_1 и R_2

Поле	Критерий R_1		Критерий R_2	
	Точность	Число кадров	Точность	Число кадров
Фамилия	0,99	5,47	0,97	7,34
Имя	0,99	4,59	1,00	5,47
Отчество	0,99	4,91	1,00	5,34
Серия паспорта	0,97	3,78	0,97	5,84
Номер паспорта	0,99	3,68	0,91	4,09
Дата рождения	0,96	4,34	0,97	3,16
Место рождения	0,91	8,87	0,85	9,41

Таблица 13 — Пример вариационного ряда для построенных популяций

Популяция G	$l^*(G)$	R_1	R_2
`Г.ПЕНЗАЛ9 ПЕНЗЕНСКОЙОБЛ`	2	0,730	0,842
`Г.ПЕНЗА-19 ПЕНЗЕНСКОЙОБ1`	3	0,127	0,318
`Г.ПЕНЗА-19 ПЕНЗЕНСКОЙОБЛ`	1	0,095	0,322
`Г.ПЕНЗА-Л9 ПЕНЗЕНСКОЙОБЛ`	1	0,092	0,417
`Г.ПЕНЗА-Л9 ПЕНЗЕНСКОЙОБ1`	3	0,058	0,276
`Г.ПЕНЗА-Л9 1П1ВЕНЗЕНСКОЙОБЛ1`	1	$5 \cdot 10^{-6}$	0,083
`Г.ПЕНЗА-Л9 ТП1ВЕНЗЕНСКОЙОБЛ1`	2	$2 \cdot 10^{-6}$	0,039
`Г.ПЕНЗА-Л9 1ПЕ1ВЕНЗЕНСКОЙОБЛ1`	22	$1 \cdot 10^{-6}$	0,113
`Г.ПЕНЗА-Л9 1ПЕ1ВЕНЗЕНСКОЙОБЛЛ`	6	0,000	0,049

це 13 приведен вариационный ряд для построенных популяций для поля «место рождения», которое можно отнести к такому классу.

Для таких случаев можно разбивать поля на отдельные слова с последующей конкатенацией выбранных наилучших популяций в качестве результата. При такой модификации алгоритма точность повышается до значений 0,97–0,98.

Описанный подход построен на формулировке постановки задачи об остановке распознавания как задачи ограничения наблюдений и обладает достаточной универсальностью. С другой стороны, его недостатком для построения систем распознавания может быть некоторая сложность в подборе оптимальных значений настраиваемых параметров (порогов γ_1, γ_2, n_3). В следующем разделе будет рассмотрен другой подход к решению этой задачи как к задаче совместной оптимизации общего функционала, включающего в себя как ожидаемую ошибку распознавания, так и ожидаемое количество обработанных кадров.

3.4.2 Метод последовательного принятия решения на основе моделирования следующего комбинированного результата

Другой подход к решению задачи остановки процесса результата распознавания является рассмотрение задачи (3.10) как задачу оптимизации функционала в процессе последовательного принятия решения. Пусть $\mathbb{X}$ обозначает множество всевозможных значений результата распознавания объекта. К примеру, в случае задачи классификации одиночного объекта множество $\mathbb{X}$ можно представить как множество всевозможных отображений из множества классов C в множество оценок принадлежности, в случае задачи распознавания текстовой строки – как множество всевозможных конечных последовательностей таких отображений. Обозначим под $x^* \in \mathbb{X}$ истинное значение объекта. Пусть на множестве $\mathbb{X}$ задана метрика ρ так, что расстояние $\rho(x, x^*)$ отражает характеристику ошибки результата x распознавания объекта. В качестве простейшей метрики ρ можно рассмотреть метрику точного соответствия:

$$\rho(a, b) = \begin{cases} 0, & \text{если } a = b; \\ 1, & \text{если } a \neq b. \end{cases} \tag{3.59}$$

Рассмотрим процесс распознавания в видеопотоке как последовательное наблюдение случайных результатов распознавания одиночных изображений $X_1, X_2, \ldots$, один результат за один шаг процесса так, что каждое наблюдение $x_i = r(I_i(x)) \in \mathbb{X}$ является реализацией X_i. Будем считать, что $X_1, X_2, \ldots$ имеют одинаковое совместное распределение с x^*. В качестве результата распознавания видеопоследовательности $R^{(n)}$ будем считать результат

комбинирования $F^{(n)}(x_1, x_2, \ldots, x_n)$, тем самым рассматривая схему распознавания (3.2) с тождественной функцией выбора S.

Процесс распознавания может быть остановлен на любом шаге $n > 0$ с некоторой функцией штрафа. Функция штрафа, соответствующая (3.10), переформулированная с помощью метрики ρ (3.59), выражается следующим образом:

$$L_n = w_e + (w_c - w_e) \cdot (1 - \rho(F^{(n)}(x_1, x_2, \ldots, x_n)), x^*)) + w_f \cdot n = \\ = w_c + (w_e - w_c) \cdot \rho(F^{(n)}(x_1, x_2, \ldots, x_n)), x^*) + w_f \cdot n, \quad (3.60)$$

где n – номер шага, на котором был остановлен процесс распознавания, w_e – стоимость ввода ошибочного результата, w_c – стоимость ввода корректного результата, w_f – стоимость распознавания одного кадра. Стоит обратить внимание, что в такой постановке в качестве метрики ρ не обязательно использовать метрику точного соответствия (3.59), но и другие метрики, к примеру, метрику Левенштейна [235], если объектом распознавания является текстовая строка.

Правило остановки может быть представлено как случайная величина N (случайный момент остановки), распределение которой зависит от входных наблюдений. Задача состоит в выборе правила остановки, доставляющего минимум функционалу ожидаемого убытка:

$$V(N) = \mathrm{E}(L_N(X_1, X_2, \ldots, X_N)), \quad (3.61)$$

где $\mathrm{E}(\cdot)$ обозначает математическое ожидание.

Особым классом задач остановки является класс *монотонных* задач, определяемый следующим образом. Пусть A_n обозначает событие $\{L_n \leqslant \mathrm{E}_n(L_{n+1})\}$, где под $\mathrm{E}_n(X)$ подразумевается условное математическое ожидание $\mathrm{E}(X|X_1 = x_1, \ldots, X_n = x_n)$. Задача остановки называется монотонной, если выполняется $A_0 \subset A_1 \subset A - 2 \subset \ldots$ (т.е. если событие A_n произошло на шаге n, то соответствующие события $A_{n+1}, A_{n+2}, \ldots$ также произойдут). Для монотонных задач остановки с конечным горизонтом (т.е. если существует некоторый шаг $n_{\max}$, на котором процесс обязан остановиться) оптимальным является «близорукое правило остановки», останавливающее процесс на шаге n, если текущее значение функции убытка не превосходит ожидаемого значения убытка при остановке на шаге $n + 1$:

$$N^* = \min\{n \geqslant 0 : L_n \leqslant \mathrm{E}_n(L_{n+1})\}. \quad (3.62)$$

Сформулируем следующее требование к семейству функций комбинирования $F^{(n)}$: ожидаемое расстояние между двумя соседними комбинированными результатами распознавания не возрастает со временем:

$$\mathrm{E}(\rho(F^{(n)}(x_1,\ldots,x_n),F^{(n+1)}(x_1,\ldots,x_{n+1}))) \geqslant \\ \geqslant \mathrm{E}(\rho(F^{(n+1)}(x_1,\ldots,x_{n+1}),F^{(n+2)}(x_1,\ldots,x_{n+2}))) \quad \forall n>0. \quad (3.63)$$

Пользуясь таким предположением о семействе функций комбинирования $F^{(n)}$ можно показать, что задача остановки (3.61) с функцией убытка (3.60) становится монотонной, начиная с некоторого шага. Действительно, обозначим через B_n событие $\{\mathrm{E}_n(\rho(F^{(n)}(x_1,\ldots,x_n),F^{(n+1)}(x_1,\ldots,x_{n+1}))) \leqslant w_f/(w_e-w_c)\}$ и рассмотрим задачу остановки, начиная с шага n, на котором событие B_n впервые произошло. События A_n, рассматриваемые в условии монотонности, принимают следующий вид:

$$A_n : \{w_c+(w_e-w_c)\cdot\rho(F^{(n)}(x_1,\ldots,x_n),x^*)+w_f\cdot n \leqslant \\ \leqslant w_c+(w_e-w_c)\cdot\mathrm{E}_n(\rho(F^{(n+1)}(x_1,\ldots,X_{n+1}),x^*))+w_f\cdot(n+1)\} = \\ = \{\rho(F^{(n)}(x_1,\ldots,x_n),x^*)-\mathrm{E}_n(\rho(F^{(n+1)}(x_1,\ldots,X_{n+1}),x^*)) \leqslant \\ \leqslant w_f/(w_e-w_c)\}. \quad (3.64)$$

При фиксированном x^*, на шаге n, пользуясь неравенством треугольника, можно получить соотношение между расстоянием от текущего результата распознавания до истинного значения, ожидаемым расстоянием до результата на следующем шаге и ожидаемым расстоянием от следующего результата до истинного значения:

$$\rho(F^{(n)}(x_1,\ldots,x_n),x^*) \leqslant \mathrm{E}_n(\rho(F^{(n)}(x_1,\ldots,x_n),F^{(n+1)}(x_1,\ldots,X_{n+1})))+ \\ +\mathrm{E}_n(\rho(F^{(n+1)}(x_1,\ldots,X_{n+1}),x^*)) \Rightarrow \\ \Rightarrow \rho(F^{(n)}(x_1,\ldots,x_n),x^*)-\mathrm{E}_n(\rho(F^{(n+1)}(x_1,\ldots,X_{n+1}),x^*)) \leqslant \\ \leqslant \mathrm{E}_n(\rho(F^{(n)}(x_1,\ldots,x_n),F^{(n+1)}(x_1,\ldots,X_{n+1}))). \quad (3.65)$$

Если правая часть неравенства (3.65) не превышает константы $w_f/(w_e-w_c)$, то и левая часть также не превышает $w_f/(w_e-w_c)$ и, следовательно, если происходит событие B_n, то и событие A_n (3.64) также должно произойти. Согласно предположению (3.63), если событие B_n произойдет, то и событие B_{n+1} также произойдет. Таким образом, $\forall n>0 : B_n \subset A_n \wedge B_n \subset B_{n+1}$.

Из этого следует, что начиная с шага n, на котором событие B_n произошло впервые, события $A_n, A_{n+1}, A_{n+2}, \ldots$ также произойдут, а значит задача остановки может рассматриваться как монотонная задача, начиная с этого шага, из чего в свою очередь следует оптимальность «близорукого правила» (3.62) среди всех правил остановки, достигающих шага n (в случае, если задача имеет конечный горизонт).

Рассмотрим теперь правило остановки, срабатывающее в случае, если произошло событие B_n:

$$N_B = \min\{n > 0 : \Delta_n \stackrel{\text{def}}{=} \mathrm{E}_n(\rho(F^{(n)}(x_1, \ldots, x_n), F^{(n+1)}(x_1, \ldots, X_{n+1}))) \leqslant \leqslant w_f/(w_e - w_c)\}. \quad (3.66)$$

Если правило N_B останавливается на шаге n, то и «близорукое правило» (3.62) остановится на этом шаге, а значит это решение является оптимальным. Более того, если $\rho(F^{(n)}(x_1, \ldots, x_n), x^*) - \mathrm{E}_n(\rho(F^{(n+1)}(x_1, \ldots, X_{n+1}), x^*)) > w_f/(w_e - w_c)$, то правило N_B не останавливается, также как и оптимальное правило (3.62). Следовательно, в случае, если предположение (3.63) верно, правило N_B никогда не остановится раньше времени и, если правило требует остановки, то решение об остановке оптимально.

Тем самым, для построения алгоритма остановки процесса распознавания с функцией убытка (3.60) можно использовать следующий подход:

1. Оценить ожидаемое расстояние Δ_n от текущего комбинированного результата распознавания объекта до (неизвестного) следующего комбинированного результата;
2. Решение об остановке на шаге n принимать путем порогового отсечения расстояния, оцененного в пункте 1, таким образом аппроксимируя поведение правила N_B (3.66).

Схема подхода представлена на рисунке 3.11.

В качестве одного из способов оценки ожидаемого расстояния Δ_n можно производить моделирование следующего комбинированного результата, исходя из предположения, что новое наблюдение $X_{n+1} = x_{n+1}$ будет близко к уже полученным на предыдущих шагах наблюдениям:

$$\Delta_n \approx \frac{1}{n+1}\left(\delta + \sum_{i=1}^{n} \rho(F^{(n)}(x_1, \ldots, x_n), F^{(n+1)}(x_1, \ldots, x_n, x_i))\right), \quad (3.67)$$

где δ – настраиваемый параметр.

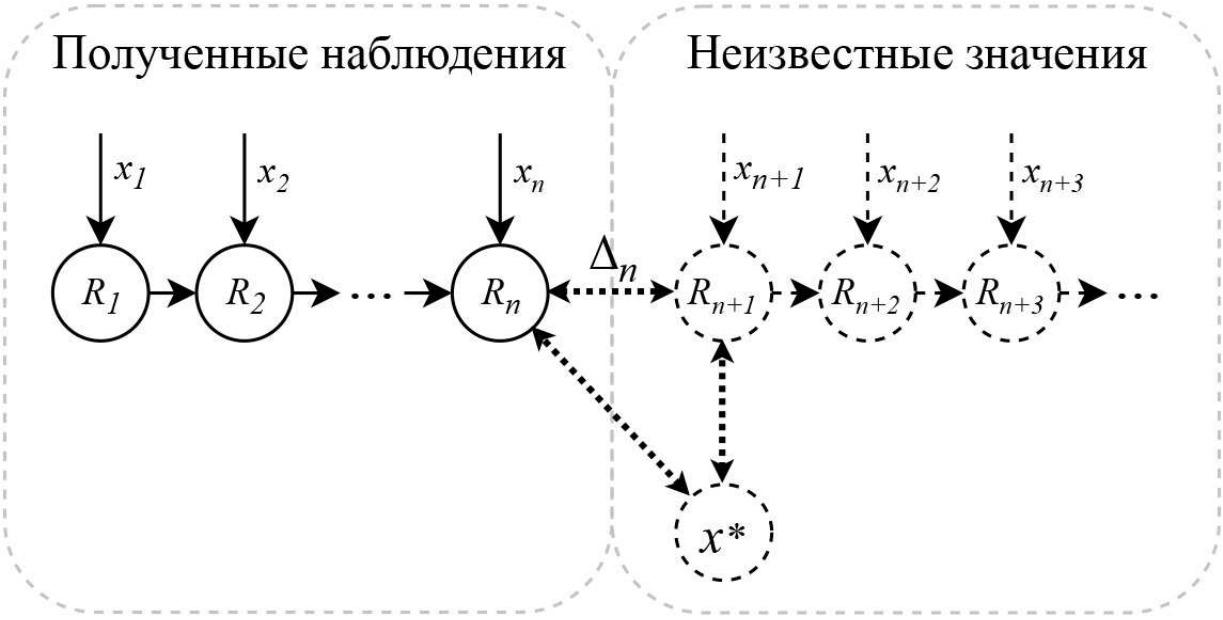


Рисунок 3.11 — Схема метода остановки процесса распознавания: оценка ожидаемого расстояния от текущего комбинированного результата до следующего.

Другим возможным способом моделирования следующего комбинированного результата для случаев, когда объектом распознавания является текстовая строка, является предсказание значений оценок принадлежности каждой альтернативы каждого знакоместа распознаваемой строки, исходя из накопленных наблюдений, путем поиска параметров распределения Дирихле или аналогов (опираясь на результаты, описанные в разделе 3.3.2).

Обратим внимание, что правило остановки в рамках описанного подхода можно выразить и в задаче, у которой стоимостью обладает не очередное наблюдение (как в задаче (3.10) стоимость распознавания одного кадра имеет величину w_f), а единица времени. Пусть w_t – стоимость продолжения процесса распознавания на одну единицу времени. Как было описано ранее в подразделе 3.2.2, пусть $I_t(x)$ – изображение объекта x, захваченное в момент времени t, и $T(t)$ – момент времени, в который может быть получен результат его распознавания. Тогда, по аналогии с (3.60), функция штрафа при остановке в момент времени $T(t)$ принимает следующий вид:

$$L_t \;=\; w_c + (w_e - w_c) \cdot \rho(F^{(n)}(x_0, x_{T^1(0)}, x_{T^2(0)}, \ldots, x_t)), x^*) + w_t \cdot T(t), \quad (3.68)$$

где $x_0, x_{T^1(0)}, x_{T^2(0)}, \ldots, x_t$ – последовательность результатов распознавания объекта на изображениях, захваченных в моменты времени $0, T^1(0), T^2(0), \ldots, t$.

Аналогичным образом построенное правило остановки (3.66) с функцией штрафа (3.68) принимает следующий вид:

$$T_B = \min\Bigg\{ T(t) > 0 : \\ : \mathrm{E}_{T(t)}(\rho(F^{(n(t))}(x_0, x_{T^1(0)}, \ldots, x_t), F^{(n(T(t)))}(x_0, x_{T^1(0)}, \ldots, X_{T(t)}))) \leqslant \\ \leqslant \frac{w_t \cdot \mathrm{E}_{T(t)}(\Delta_{T(t)})}{w_e - w_c} \Bigg\}, \quad (3.69)$$

где $T(t)$ – момент времени, в который будет доступен результат распознавания изображения, захваченного в момент t; $\mathrm{E}_{T(t)}(\cdot)$ – условное ожидание в момент времени $T(t)$; $n(t)$ – количество изображений, захваченных к моменту времени t включительно; $\mathrm{E}_{T(t)}(\Delta_{T(t)})$ – ожидаемое в момент времени $T(t)$ время, которое будет затрачено на распознавание изображения $I_{T(t)}(x)$, захваченного в момент времени $T(t)$.

Проведем экспериментальное исследование метода остановки на основе правила (3.66) на примере задачи распознавания текстовых полей документов, удостоверяющих личность, используя открытые пакеты данных MIDV-500 [207] и MIDV-2019 [208]. Исходные результаты распознавания текстовых полей получим при помощи метода [236], и в качестве метода комбинирования межкадровых результатов будем использовать метод ROVER [234]. В качестве метрики ρ будем использовать нормализованное расстояние Левенштейна [235].

Стоит обратить внимание, что как метод на основе анализа популяций, описанный в предыдущем разделе, так и метод на основе моделирования следующего комбинированного результата, обладают рядом параметров (порогов) принятия решения. В методе анализа популяций это пороги γ_1, γ_2, n_3, а в методе моделирования следующего результата – порог $w_f/(w_e - w_c)$, с которым сравнивается оценка ожидаемого расстояния до моделируемого результата, которых, хотя и выражен исходя из постановки задачи, на практике может варьироваться, поскольку значения стоимости ошибок w_e, w_c и w_f не всегда могут быть строго определены.

Для сравнения описанного метода и метода на основе анализа популяций, описанного в предыдущем подразделе, построим так называемые *профили эффективности*, отражающие, в данном случае, соотношение между средней ошибкой распознавания объекта и средним количеством использованных кадров при всевозможных значениях порогов.

Обозначим как N_{CX} правило остановки на основе метода анализа популяций, описанного в предыдущем подразделе, в котором популяции собираются на основе исходных результатов распознавания $r(I_1(x)), r(I_2(x)), \ldots, r(I_n(x))$, а также как N_{CR} – его модификацию, в которой популяции собираются на основе комбинированных результатов $F^{(1)}(r(I_1(x))), F^{(2)}(r(I_1(x)), r(I_2(x))), \ldots, F^{(n)}(r(I_1(x)), \ldots, r(I_n(x)))$.

На рисунках 3.12 и 3.13 представлены профили эффективности правил остановки N_{CX} и N_{CR}, применительно к задаче распознавания текстовых полей из пакетов данных MIDV-500 и MIDV-2019 соответственно.

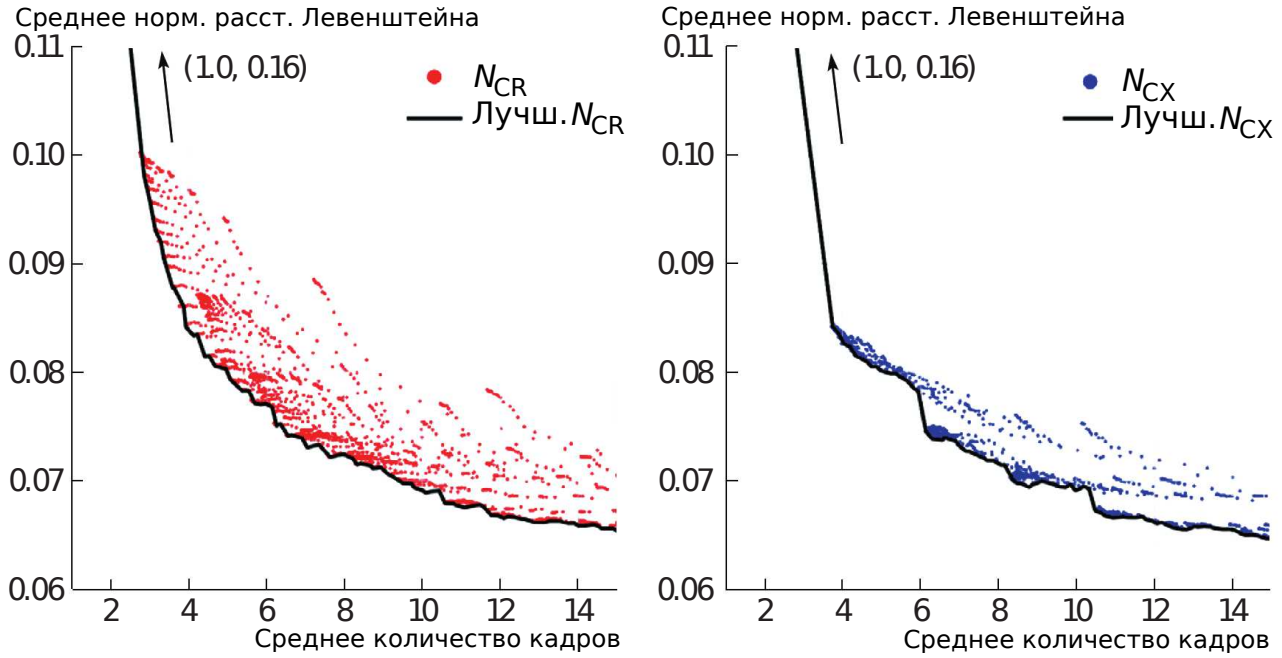


Рисунок 3.12 — Профили эффективности правил остановки N_{CR} (слева) и N_{CX} (справа) для распознавания полей идентификационных документов из пакета данных MIDV-500, точками обозначены различные варианты значений порогов, сплошной линией выделен лучший случай.

На рисунках 3.14 и 3.15 представлено сравнение профилей эффективности разных правил остановки для распознавания полей документов из пакетов данных MIDV-500 и MIDV-2019, соответственно. Серой пунктирной линией на рисунках обозначено «тривиальное» правило остановки N_K, порогом которого является количество обработанных кадров (т.е. процесс останавливается после того как были обработаны и распознаны k кадров). Синей и красной пунктирными линиями обозначены лучшие случаи правил остановки N_{CX} и N_{CR}, соответственно, и сплошной зеленой линией обозначено правило остановки N_B

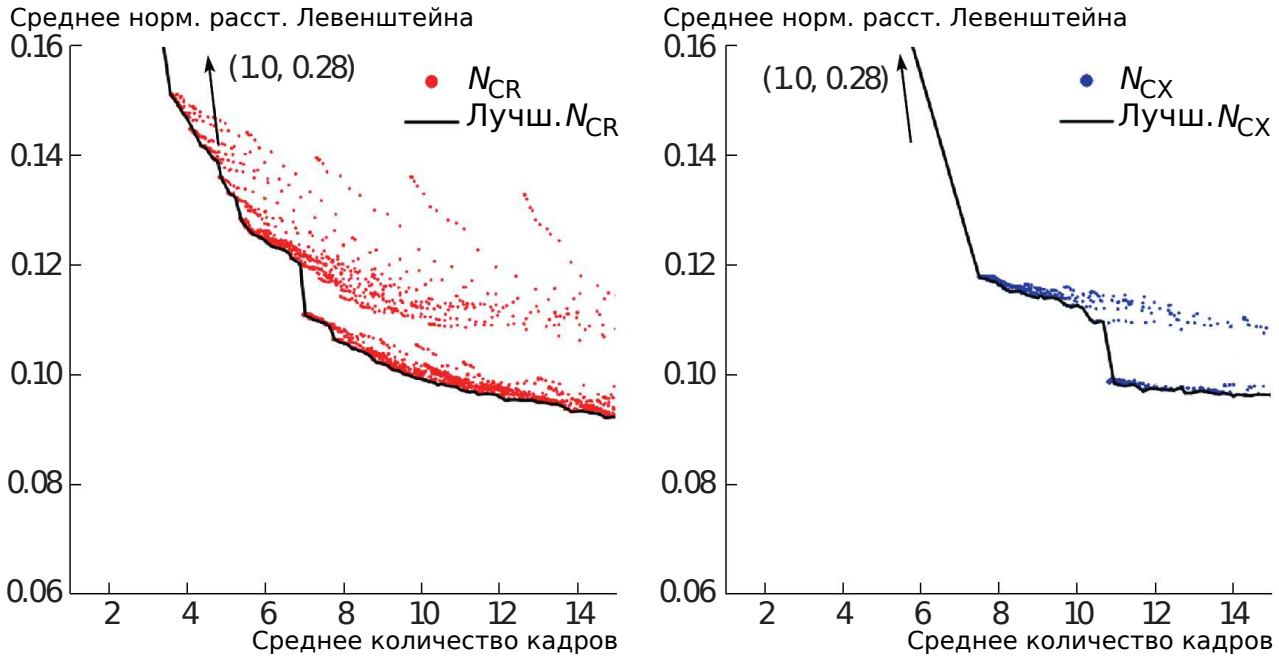


Рисунок 3.13 — Профили эффективности правил остановки N_{CR} (слева) и N_{CX} (справа) для распознавания полей идентификационных документов из пакета данных MIDV-2019, точками обозначены различные варианты значений порогов, сплошной линией выделен лучший случай.

(3.66), реализованное при помощи моделирования следующего комбинированного результата методом (3.67). Можно убедиться, что правило N_B позволяет достичь более низкого среднего уровня ошибки распознавания при том же среднем количестве кадров, по сравнению как с тривиальным правилом N_K, так и с лучшими случаями выбора порогов для N_{CX} и N_{CR}.

В таблицах 14 и 15 указаны достигнутые значения среднего расстояния от комбинированного результата распознавания до истинного ответа в момент остановки при ограничении на среднее количество обработанных кадров для текстовых полей пакетов данных MIDV-500 и MIDV-2019 соответственно.

Из полученных экспериментальных результатов можно сделать вывод, что метод остановки, основанный на моделировании следующего комбинированного результата превосходит метод анализа популяций по потенциально достигаемой точности результата распознавания в момент остановки при равном среднем количестве обработанных кадров. Однако недостатком такого метода является необходимость производить вычисления, связанные с моделированием следующего результата (к примеру, при помощи подхода (3.67)), более трудоемкие, чем вычисления, необходимые для анализа размеров накопленных популяций. Применительно к конкретным системам распознавания объектов в видеопосле-

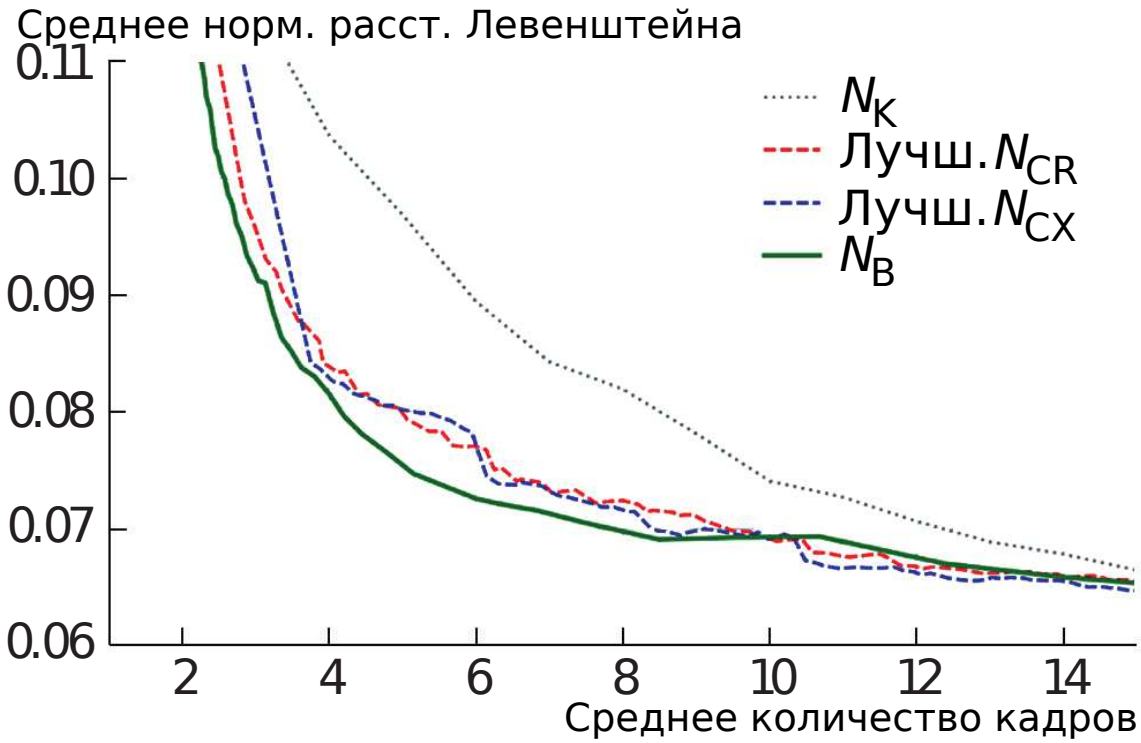


Рисунок 3.14 — Профили эффективности правил N_K, N_B, и лучших случаев правил N_{CX} и N_{CR} для распознавания полей идентификационных документов из пакета данных MIDV-500.

Таблица 14 — Достигнутые значения среднего расстояния до истинного ответа в момент остановки с ограничением на среднее количество обработанных кадров для текстовых полей пакета данных MIDV-500

Правило остановки	Ограничение среднего количество кадров							
	$\leqslant 3$	$\leqslant 4$	$\leqslant 5$	$\leqslant 6$	$\leqslant 7$	$\leqslant 8$	$\leqslant 9$	$\leqslant 10$
Лучший N_{CX}	0,161	0,084	0,080	0,078	0,074	0,072	**0,069**	**0,069**
Лучший N_{CR}	0,096	0,084	0,080	0,077	0,074	0,072	0,071	**0,069**
N_K	0,115	0,104	0,097	0,089	0,084	0,082	0,078	0,074
N_B	**0,092**	**0,082**	**0,076**	**0,073**	**0,071**	**0,070**	**0,069**	**0,069**

довательности, тем самым, имеет смысл использовать оба подхода, принимая решения исходя из специфики задачи и из доступных вычислительных ресурсов.

В разделе 3.2.2 уже описывалось, как рассматриваемые системы распознавания в видеопотоке отличаются от классических усиленным влиянием производительности алгоритмов распознавания одиночных изображений на выход системы. Это также можно проследить по выкладкам, представленным в

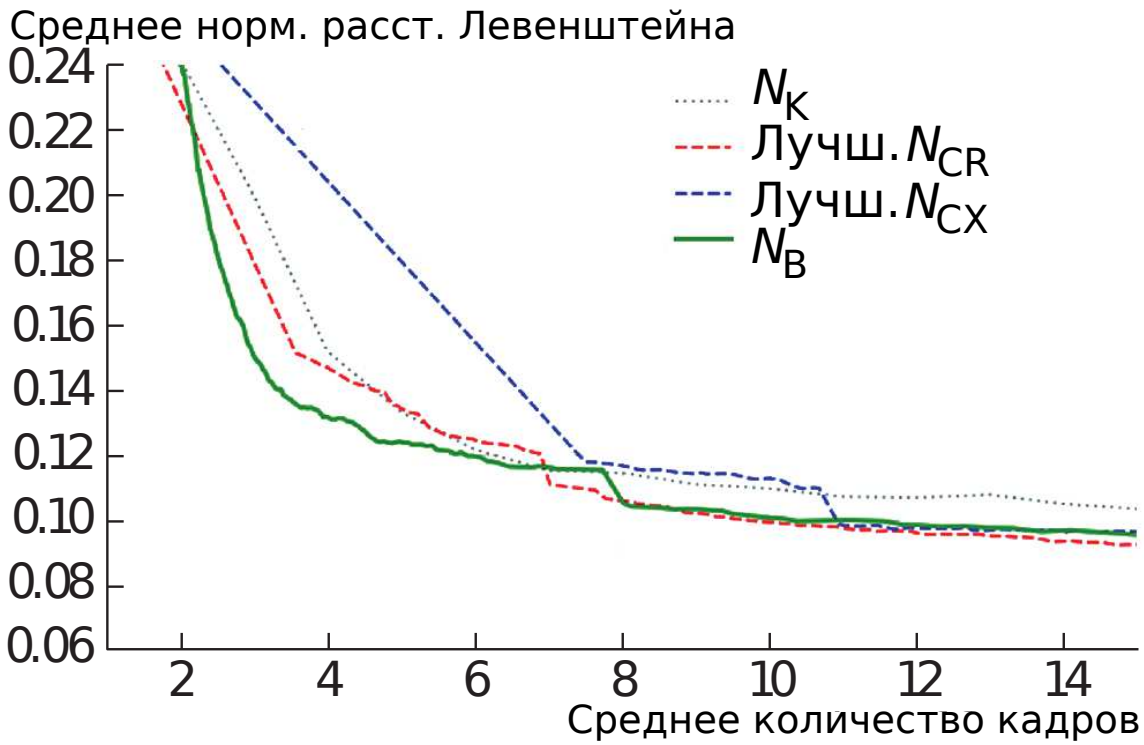


Рисунок 3.15 — Профили эффективности правил N_K, N_B, и лучших случаев правил N_{CX} и N_{CR} для распознавания полей идентификационных документов из пакета данных MIDV-2019.

Таблица 15 — Достигнутые значения среднего расстония до истинного ответа в момент остановки с ограничением на среднее количество обработанных кадров для текстовых полей пакета данных MIDV-2019

Правило остановки	Ограничение среднего количество кадров							
	$\leqslant 3$	$\leqslant 4$	$\leqslant 5$	$\leqslant 6$	$\leqslant 7$	$\leqslant 8$	$\leqslant 9$	$\leqslant 10$
Лучший N_{CX}	0,278	0,278	0,278	0,278	0,278	0,116	0,114	0,113
Лучший N_{CR}	0,278	0,147	0,136	0,125	**0,111**	0,106	0,102	**0,099**
N_K	0,200	0,152	0,133	0,122	0,115	0,114	0,111	0,110
N_B	**0,150**	**0,123**	**0,119**	**0,116**	0,115	**0,103**	**0,101**	0,100

этом разделе: уменьшение времени обработки изображения означает уменьшение стоимости наблюдения w_f (в задаче с выражением функции убытка (3.60)), что в свою очередь приводит к более поздней (с точки зрения количества обработанных кадров) остановке согласно правилу (3.66), и, как следствие, к большему количеству обработанных кадров. А как было показано в разделе 3.3.1, эмпирические данные свидетельствуют о более высокой ожидаемой точности результата в случае увеличения количества обработанных кадров. Тем

самым, в контексте систем распознавания в видеопотоке, в особенности систем, предназначенных для исполнения на устройствах с ограниченной и немасштабируемой производительностью (таких, как смартфоны или планшетные компьютеры), необходимо адаптировать методы и алгоритмы, предназначенные для обработки изображений и распознавания образов, для максимально эффективного исполнения на таких платформах.

3.5 Использование особенностей архитектур современных мобильных центральных процессоров для оптимизации вычислений в системах распознавания

Вычислительная эффективность систем компьютерного зрения, к которым также относятся системы распознавания документов, является крайне важным аспектом. Алгоритмы, использующиеся в системах распознавания, такие как алгоритмы обработки изображения, алгоритмы выделения графических примитивов, алгоритмы распознавания образов и другие, разрабатываются с огромной скоростью, при этом изменяются как их качественные характеристики, так и трудоемкость. Использование систем компьютерного зрения и распознавания образов на мобильных устройствах, маломощных терминалах, и других встраиваемых системах [149; 259; 260] накладывает дополнительные ограничения на выбор алгоритмов и на их реализацию ввиду ограниченности вычислительных ресурсов. Как следствие, важно уделять особенное внимание как трудоемкости применяемых алгоритмов обработки изображений и компьютерного зрения, так и техническим подходам, позволяющим минимизировать время исполнения в системах, где скорость решения задач обработки изображений и распознавания являются критичными [261—263].

Современные центральные процессоры предоставляют набор инструментов, предназначенных для максимально эффективной реализации определенного класса алгоритмов и разработки высокопроизводительного программного обеспечения. Одним из таких инструментов является возможность реализовывать многопоточные (многоядерные) вычисления в случае, если алгоритмы могут быть разбиты на набор задач без зависимостей по данным. Другим инструментом являются инструкции SIMD (ОКМД, Одиночный поток Команд,

Множественный поток Данных, Single Instruction Multiple Data) – набор расширений, позволяющий параллельную обработку пакетов данных с однотипными операциями, используя только одно вычислительное ядро. Каждая из SIMD-инструкций выполняется последовательно, однако на векторе данных. Таким образом, использование архитектуры SIMD позволяет повысить эффективность исполнения алгоритмов, предполагающих векторные или матричные операции [264] (см. рис. 3.16). Процессоры семейства Intel x86 предоставляют наборы инструкций MMX, SSE и AVX, процессоры семейства ARM предоставляют наборы инструкций VFP и NEON.

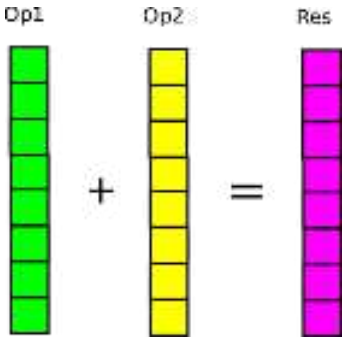


Рисунок 3.16 — Сложение двух векторов длины 8 при помощи единственной SIMD-инструкции.

Несмотря на то, что современные компиляторы производят автоматическую "векторизацию" (адаптацию программного кода, позволяющая компиляторам принимать решение о вызове SIMD-инструкций), на уровне компилятора не всегда можно однозначно принять решение о том, каким образом и какие SIMD-инструкции следует применять, что приводит к тому, что автоматическая векторизация не всегда бывает эффективна [265]. Тем самым, разработчикам приходится использовать SIMD-инструкции напрямую, при проектировании алгоритмов и при имплементации систем. Применительно к компонентам систем распознавания, для которых векторизация имеет наибольший смысл – функциям базовой обработки изображений, в рамках которых требуется обрабатывать все пиксели изображения похожим образом. Из класса таких функций, широко используемых в системах распознавания, можно выделить:

1. функции фильтрации изображений (Гауссова фильтрация, билатеральная фильтрация, фильтрация усреднением и др.);
2. поворот изображения на произвольный угол;
3. транспонирование изображения;
4. морфологические преобразования (эрозия, дилатация);

5. и т.п.

В последующих разделах будут рассмотрены подходы к реализации операций транспонирования изображения и морфологических преобразований эрозии и дилатации с помощью SIMD-инструкций применительно главным образом к архитектуре семейства процессоров ARM.

3.5.1 Алгоритм эффективного транспонирования матриц с использованием инструкций ARM NEON

Транспонирование изображения является важной и широко используемой операцией в системах компьютерного зрения и распознавания образов. Одним из применений транспонирования изображения является разбиение многоканального изображения на изображения отдельных каналов для их независимой обработки. При обработке одноканальных изображений операция транспонирования используется главным образом для эффективной реализации алгоритмов фильтрации, поскольку значительная часть таких алгоритмов сепарабельна (т.е. может быть разделена на независимые этапы обработки изображения вдоль строк изображения и вдоль столбцов изображения с одномерными фильтрами). Примерами сепарабельных алгоритмов являются Гауссова фильтрация, фильтрация усреднением с прямоугольным окном и различные виды морфологических преобразований [266]. Вертикальные фильтры применяются к столбцам изображения и обрабатывают пиксели всех строк одинаковым образом. В случае, если пиксели строк изображения расположены на соседних позициях в оперативной памяти вычислительного устройства, SIMD-инструкции могут быть использованы для увеличения производительности. Однако горизонтальные фильтры применяются к строкам изображения и одинаковым образом обрабатывают все пиксели столбцов изображения, и для оптимизации горизонтальных фильтров нельзя использовать SIMD-инструкции напрямую. Однако зачастую можно добиться увеличения производительности путем транспонирования изображения, применения ускоренного вертикального фильтра и обратного транспонирования. В таких случаях крайне важно использовать максимально эффективную реализацию транспонирования изображений.

В документации пакета инструкций ARM NEON содержится пример транспонирования матриц размера 4×4 с 16-битными элементами [267]. Для этой же цели пакет инструкций SSE2 процессоров семейства Intel x86 предоставляет макрос `_MM_TRANSPOSE4_PS`. Существуют работы, рассматривающие эффективные реализации транспонирования матриц с использованием пакетов инструкций AVX для Intel x86 [268]. В данном разделе будет описана эффективная имплементация транспонирования матриц размером 8×8 и 16×16 с использованием пакета инструкций NEON для процессоров ARM, широко используемых в мобильных устройствах.

Для использования SIMD-инструкций при разработке, как правило, используются встроенные функции компилятора (Intrinsic functions) – специальные примитивы компилятора, транслируемые в вызовы соответствующих инструкций процессора с автоматическими оптимизациями. Поскольку размер регистра ARM NEON составляет 128 бит, компиляторы, поддерживающие ARM NEON, содержат наборы встроенных функций, действующих на 128-битных регистрах, включая загрузку данных в память, выгрузку данных из памяти, арифметические операции, конвертации типов, перестановки и др. [269].

Набор инструкций архитектуры ARM включает в себя ассемблерные инструкции `VTRN.n`, которые можно использовать для реализации эффективного транспонирования матриц. Инструкции `VTRN.n` интерпретируют операнды как векторы, которые состоят из n битов, и производят транспонирование матриц размера 2×2, составленных из этих битов. Пример инструкции `VTRN.16` приведен на рис. 3.17.

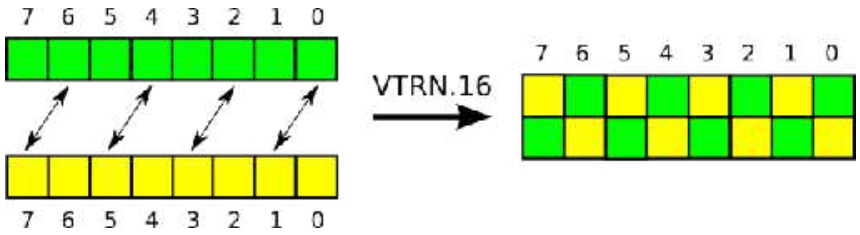


Рисунок 3.17 — Результат выполнения инструкции `VTRN.16`.

Для доступа к ассемблерным инструкциям `VTRN.n` может быть использована встроенная функция `vtrnq` из набора ARM NEON, которая транслируется компилятором в несколько вызовов инструкций `VTRN.n`.

Обозначим как `NxM.k` матрицу размером $N \times M$, элементами которой являются k-битные значения. Документация ARM NEON демонстрирует транспонирование матриц `4x4.16` при помощи трех вызовов встроенной функции

`vtrnq` [267]. Аналогичным образом матрицы `4x4.32` могут быть транспонированы при помощи четырех вызовов встроенной функции `vtrnq`: в начале транспонируются блоки `2x2.32` в первых двух строках, после чего транспонируются блоки `2x2.32` в третьей и четвертой строках, после чего транспонируются блоки `2x2.64`.

Реализовать транспонирование матриц `8x8.16` можно с использованием 64 инструкций: 16 инструкций загрузки/выгрузки, 32 инструкции перестановки и 16 дополнительных инструкций интерпретации векторов значений одного типа как векторов значений другого типа (последний тип инструкций служит для корректной компиляции и не влияет на производительность). Реализация для языка `C` приведена на листинге 3.1.

Листинг 3.1: Реализация эффективного алгоритма транспонирования матриц 8×8 с 16-битными элементами с использованием инструкций ARM NEON.

```
1  uint16x8x2_t t0 = vtrnq_u16(vld1q_u16(addr_s0), vld1q_u16(addr_s1));
2  uint16x8x2_t t1 = vtrnq_u16(vld1q_u16(addr_s2), vld1q_u16(addr_s3));
3  uint16x8x2_t t2 = vtrnq_u16(vld1q_u16(addr_s4), vld1q_u16(addr_s5));
4  uint16x8x2_t t3 = vtrnq_u16(vld1q_u16(addr_s6), vld1q_u16(addr_s7));
5
6  uint32x4x2_t x0 = vtrnq_u32(vreinterpretq_u32_u16(t0.val[0]),
7
8  vreinterpretq_u32_u16(t1.val[0]));
9  uint32x4x2_t x1 = vtrnq_u32(vreinterpretq_u32_u16(t2.val[0]),
10 vreinterpretq_u32_u16(t3.val[0]));
11 uint32x4x2_t x2 = vtrnq_u32(vreinterpretq_u32_u16(t0.val[1]),
12 vreinterpretq_u32_u16(t1.val[1]));
13 uint32x4x2_t x3 = vtrnq_u32(vreinterpretq_u32_u16(t2.val[1]),
14 vreinterpretq_u32_u16(t3.val[1]));
15 vst1q_u16(addr_d0, vreinterpretq_u16_u32(
16 vcombine_u32(vget_low_u32(x0.val[0]), vget_low_u32(x1.val[0]))));
17 vst1q_u16(addr_d1, vreinterpretq_u16_u32(
18 vcombine_u32(vget_low_u32(x2.val[0]), vget_low_u32(x3.val[0]))));
19 vst1q_u16(addr_d2, vreinterpretq_u16_u32(
20 vcombine_u32(vget_low_u32(x0.val[1]), vget_low_u32(x1.val[1]))));
21 vst1q_u16(addr_d3, vreinterpretq_u16_u32(
22 vcombine_u32(vget_low_u32(x2.val[1]), vget_low_u32(x3.val[1]))));
23 vst1q_u16(addr_d4, vreinterpretq_u16_u32(
24 vcombine_u32(vget_high_u32(x0.val[0]), vget_high_u32(x1.val[0]))));
25 vst1q_u16(addr_d5, vreinterpretq_u16_u32(
26 vcombine_u32(vget_high_u32(x2.val[0]), vget_high_u32(x3.val[0]))));
27 vst1q_u16(addr_d6, vreinterpretq_u16_u32(
28 vcombine_u32(vget_high_u32(x0.val[1]), vget_high_u32(x1.val[1]))));
```

Таблица 16 — Сравнение времени исполнения алгоритмов транспонирования матриц (Samsung Exynos 5422 с ARM NEON).

Размер матрицы	Тип данных	Время выполнения без SIMD, наносек.	Время выполнения с SIMD, наносек.
8×8	16-битное беззнаковое целое	114	**20**
16×16	8-битное беззнаковое целое	565	**47**

```
29  vst1q_u16(addr_d7, vreinterpretq_u16_u32(
30  vcombine_u32(vget_high_u32(x2.val[1]), vget_high_u32(x3.val[1]))));
```

Реализацию транспонирования матриц `16x16.8` можно реализовать аналогичным образом, используя 152 инструкции (32 инструкции загрузки/выгрузки, 72 инструкции перестановок и 48 дополнительных инструкций реинтерпретации данных).

В таблице 16 приведено сравнение времени исполнения алгоритмов транспонирования матриц `8x8.16` и `16x16.8` без ускорения с помощью SIMD-инструкций и с использованием предложенного алгоритма. Замеры производились на процессоре Samsung Exynos 5422 (поддерживающем инструкции ARM NEON). Как можно судить из приведенных замеров, предложенный алгоритм транспонирования матриц `8x8.16` превосходит по скорости исполнения в 5,7 раз реализацию без использования SIMD-инструкций, а алгоритм транспонирования матриц `16x16.8` превосходит реализацию без SIMD в 12 раз.

3.5.2 Алгоритм эффективной морфологической фильтрации изображений с использованием инструкций ARM NEON

Морфологические операции над изображениями (также называемые операциями математической морфологии) предназначены для анализа и обработки геометрических структур, присутствующих на изображении. Такие операции часто используются в алгоритмах обработки изображений документов как для фильтрации шума на изображениях, так и для решения более высокоуровневых задач, таких как точная локализация текстовых полей в зонах документа.

Как правило, каждая морфологическая операция принимает на вход два аргумента: исходное изображение и структурный элемент. Структурный элемент представляет собой фиксированную геометрическую форму, с размером, меньшим, чем размер изображения. В данном разделе будут рассматриваться прямоугольные структурные элементы с размером $w_x \times w_y$. У структурного элемента также определена его якорная точка (как правило совпадающая с центром структурного элемента, в случае, если w_x и w_y – нечетные).

Базовыми операциями математической морфологии являются операции эрозии и дилатации. При вычислении операции эрозии структурный элемент перемещается вдоль исходного изображения, и среди значений пикселей изображения, покрытых структурным элементом, вычисляется минимальное значение. Вычисленное минимальное значение записывается в результирующее изображение в позицию, соответствующую текущей позиции якорной точки структурного элемента:

$$D(x, y) = \min\{S(x - n + w_x/2, y - m + w_y/2) \mid (n, m) \subseteq T\}, \qquad (3.70)$$

где $D(x, y)$ – результирующее изображение, $S(x, y)$ – исходное изображение, $T = [0, w_x - 1] \times [0, w_y - 1]$ – прямоугольный структурный элемент с якорной точкой в центре.

Операция дилатации вычисляется аналогично эрозии, однако с вычислением максимума вместо минимума. Другие типичные операции математической морфологии, такие как открытие и закрытие изображения, морфологический градиент и т.п. могут быть выражены в виде композиции эрозии, дилатации и арифметических операций над изображениями.

Как операция эрозии, так и операция дилатации, в случае с прямоугольным структурным элементом могут быть выполнены в сепарабельном виде с уменьшением общей трудоемкости алгоритма – последовательным применением вертикального прохода по изображению (со структурным элементом размера $w_x \times 1$) и горизонтального (со структурным элементом размера $1 \times w_y$). Рассмотрим реализацию обоих проходов, предполагая, что значениями пикселей исходного изображения являются 8-битные беззнаковые целые числа.

Одним из известных подходов к имплементации горизонтального прохода эрозии (или дилатации) является алгоритм ван Херка/Гиля-Вермана [270; 271], позволяющий эффективно находить минимальное (или максимальное) значение

на отрезках константной длины w_y. Алгоритм требует выделения дополнительной памяти, в два раза превышающей размер изображения. С его помощью мы можем реализовать горизонтальный проход для всего изображения с линейной трудоемкостью от размера изображения. Пиксели каждой строки обрабатываются одинаковым образом, что позволяет использовать встроенную функцию ARM NEON `vminq_u8` (`vmaxq_u8`) для поиска минимального (максимального) значения из 16-ти пар 8-битных значений одной инструкцией.

Для сравнения мы также можем имплементировать горизонтальный проход эрозии (дилатации) за линейное время от размера окна w_y для каждого применения также с использованием встроенных функций ARM NEON, поскольку различные элементы строк изображения обрабатываются независимо. В рамках такого прохода можно заполнять две строки результирующего изображения на каждой итерации цикла по столбцам. Для каждой пары соседних строк существует $w_y - 2$ общих элемента, необходимых для вычисления минимума (максимума) на отрезке, что позволяет оптимизировать вычисления. Реализация такой линейной имплементации приведена в виде листинга 3.2.

Листинг 3.2: Линейная реализация горизонтального прохода морфологической эрозии 8-битного изображения со структурным элементом размера $1 \times w_y$ для нечетных w_y с использованием инструкций ARM NEON.

```
1  // uint8_t **src_lines — исходное изображение: двумерный массив
2  //   размера width x height с 8—битными беззнаковыми целыми значениями
3  // uint8_t **dst_lines — результирующее изображени: двумерный массив
4  //   размера width x height с 8—битными беззнаковыми целыми значениями
5  // int wing — "крыло" структурного элемента: w_y = 2 * wing + 1
6
7  for (int y = wing; y < height-wing-1; y += 2)
8  {
9    for (int x = 0; x < width; x += 16)
10   {
11     uint8x16_t val = vld1q_u8(src_lines[y-wing+1] + x);
12     for (int k = -wing + 2; k <= wing; k++)
13       val = vminq_u8(val, vld1q_u8(src_lines[y+k]+x));
14     vst1q_u8(dst_lines[y]+x,
15              vminq_u8(val, vld1q_u8(src_lines[y-wing]+x)));
16     vst1q_u8(dst_lines[y+1]+x,
17              vminq_u8(val, vld1q_u8(src_lines[y+wing+1]+x)));
18   }
19 }
```

Имплементировать вертикальный проход эрозии (дилатации) также можно при помощи алгоритма ван Херка/Гиля-Вермана с предварительным транспонированием изображения, используя метод, описанный в разделе 3.5.1. Аналогично горизонтальному проходу, также можно реализовать эффективный линейный алгоритм вертикального прохода с использованием инструкций ARM NEON – каждая строчка результирующего изображения может заполняться прямым вычислением минимума (максимума) в цикле с w_x итерациями. При помощи встроенных функций за одну итерацию будет вычисляться 16 оконных минимумов (максимумов). Края изображения обрабатываются отдельно. Реализация представлена в виде листинга 3.3.

Листинг 3.3: Линейная реализация вертикального прохода морфологической эрозии 8-битного изображения со структурным элементом размера $w_x \times 1$ для нечетных w_x с использованием инструкций ARM NEON.

```
1  // uint8_t **src_lines — исходное изображение: двумерный массив
2  //   размера width x height с 8—битными беззнаковыми целыми значениями
3  // uint8_t **dst_lines — результирующее изображени: двумерный массив
4  //   размера width x height с 8—битными беззнаковыми целыми значениями
5  // int wing — "крыло" структурного элемента: w_x = 2 * wing + 1
6
7  for (int y = 0; y < height; ++y)
8  {
9    for (int x = 0; x < width; x += 16)
10   {
11     uint8x16_t val = vld1q_u8(src_lines[y]+x—wing);
12     for (int j = x—wing+1; j <= x+wing; ++j)
13       val = vminq_u8(val, vld1q_u8(src_lines[y]+j));
14     vst1q_u8(dst_lines[y]+x, val);
15   }
16 }
```

Для сравнения производительности алгоритмов были проведены вычислительные эксперименты с морфологической эрозией исходного одноканального изображения размером 800×600, каждый пиксель которого содержал 8-битное беззнаковое целое значение, с различными размерами структурного элемента. Экспериментальные замеры производились на процессоре Samsung Exynos 5422, поддерживающем инструкции ARM NEON, реализованные алгоритмы транслировались при помощи компилятора `gcc`. Зависимость времени исполнения горизонтального прохода морфологической эрозии от размера структурного

элемента w_y представлена на рис. 3.18. Использование SIMD-инструкций позволяет ускорить реализацию при помощи алгоритма ван Херка/Гиля-Вермана более чем в 3 раза. Линейная реализация для структурного элемента размера $w_y = 3$ в 14 раз быстрее, чем реализация при помощи алгоритма ван Херка/Гиля-Вермана без использования SIMD-инструкций, однако этот выигрыш резко снижается при увеличении w_y. Можно обнаружить, что в диапазоне $0 < w_y < 70$ линейная реализация с SIMD-инструкциями является наиболее эффективной из трех сравниваемых, тогда как в случае больших структурных элементов, предпочтение следует уделять реализации алгоритмом ван Херка/-Гиля-Вермана с SIMD-инструкциями.

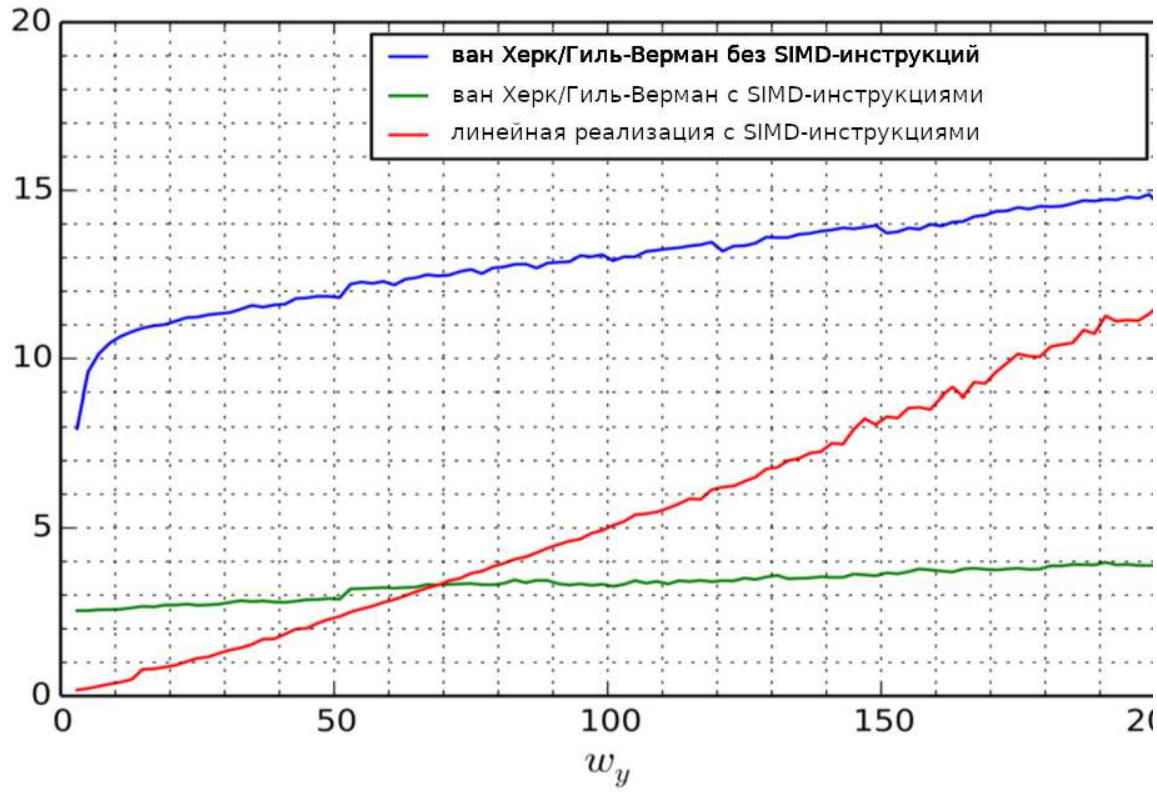


Рисунок 3.18 — Зависимость времени исполнения горизонтального прохода эрозии от размера структурного элемента.

Результаты сравнения производительности вертикального прохода морфологической эрозии представлены на рис. 3.19. Использование SIMD-инструкций позволяет ускорить реализацию, использующую алгоритм ван Херка/Гиля-Вермана почти в 3 раза при $w_x \geqslant 3$. Линейная реализация для $w_x = 3$ в 11 раз быстрее, чем реализация с алгоритмом ван Херка/Гиля-Вермана без SIMD-инструкций, и выигрыш резко снижается при увеличении w_x. Можно обнаружить, что при $0 < w_x < 60$ линейная реализация с SIMD-инструкциями

является наиболее эффективной, а при больших значениях размера структурного элемента наиболее эффективной является реализация с алгоритмом ван Херка/Гиля-Вермана и с использованием SIMD-инструкций.

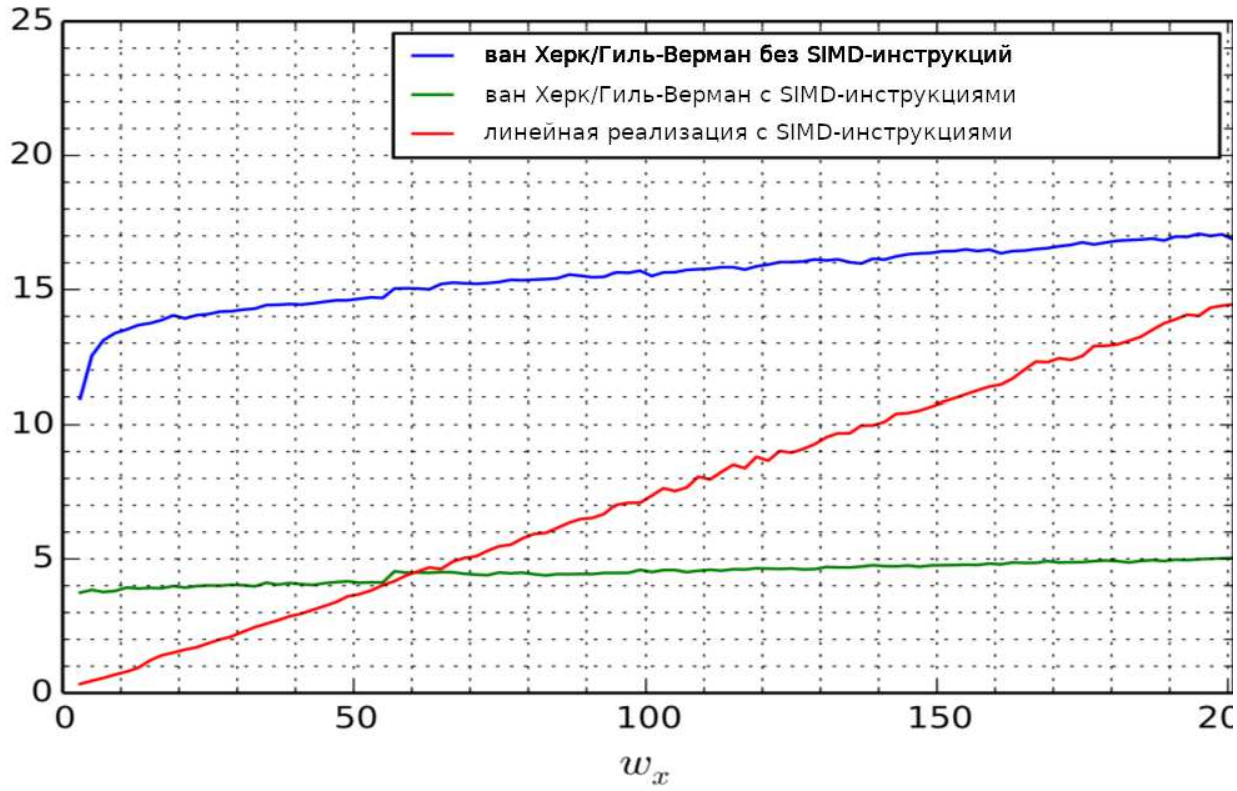


Рисунок 3.19 — Зависимость времени исполнения вертикального прохода эрозии от размера структурного элемента.

Таким образом, на основе полученных экспериментальных результатов можно сделать вывод, что наиболее эффективной имплементация морфологической эрозии (дилатации) для исполнения на процессорах семейства ARM с поддержкой инструкций NEON является комбинация рассмотренных методов. Для горизонтального прохода следует использовать линейную имплементацию (см. листинг 3.2) при $w_y < 70$ и алгоритм ван Херка/Гиля-Вермана с SIMD-инструкциями в остальных случаях. Для вертикального прохода следует использовать линейную имплементацию (см. листинг 3.3) при $w_x < 60$ и алгоритм ван Херка/Гиля-Вермана с SIMD-инструкциями в остальных случаях. Различия в пороговых значениях w_x и w_y обусловлено различиями в том, каким образом в алгоритмах горизонтального и вертикального проходов структурирован доступ к памяти устройства.

3.6 Выводы по главе

Сформулирована задача распознавания в видеопотоке. Предложено рассматривать распознавание объекта в видеопоследовательности в схеме с остановкой как задачу минимизации общего функционала стоимости. Проведено исследование трех методов комбинирования на двух открытых пакетах данных MIDV-500 и MIDV-2019. Показано, что на сравнительно хороших данных наилучшее качество показывает метод выбора лучшего (оценка по фокусировке) кадра. А на зашумленных изображениях с сильными проективными искажениями выигрывает метод комбинирования на основе подхода ROVER. Наихудший результат показал метод выбора наилучшего результата распознавания. Стоит также заметить, что стратегия выбора одного результата заведомо наиближайшего к правильному, нереализуемая на практике, дала наилучший результат.

Были построены новые вероятностные модели, описывающие результаты распознавания символов текстовых полей документов в видеопоследовательности. Введено понятие потока результатов распознавания. Рассмотренные модели предполагают, что результат распознавания знакоместа в поле документа можно представить в виде композиции случайных величин и случайных векторов. Для различных предположений получены выражения плотностей распределений результатов распознавания знакоместа и приведены методы оценивания необходимых параметров. Экспериментально проверено, что можно рассматривать распределение оценок как распределение Дирихле. Для ранжирования моделей был использован информационный критерий Акаике. Таким образом, полученные при моделировании потока результатов распознавания параметры можно использовать для комбинирования результатов классификации одиночных символов, решая тем самым задачу распознавания объекта в видеопотоке.

Предложены и исследованы два подхода к решению задачи автоматического останова распознавания в видеопотоке. Первый подход основан на анализе популяций результатов распознавания объектов, полученных на одиночных изображениях. Второй предложенный подход основан на совместной оптимизации общего функционала, включающего в себя как ожидаемую (моделируемую) ошибку распознавания, так и ожидаемое количество обработанных кадров. Результаты эксперимента показали, что метод остановки, основан-

ный на моделировании следующего комбинированного результата, превосходит метод анализа популяций по потенциально достигаемой точности результата распознавания в момент остановки при равном среднем количестве обработанных кадров.

Продемонстрировано на примере оптимизации операций методической морфологии для мобильных процессоров ARM с поддержкой инструкций NEON, что использование эффективных имплементаций алгоритмов обработки изображений и базовых алгоритмов компьютерного зрения, опирающихся на особенности современных мобильных архитектур, могут позволить минимизировать время обработки одного изображения в системах распознавания документов, что в свою очередь позволяет повышать точность распознавания документов в видеопотоке и открывает возможность построения систем распознавания документов, работающих в реальном времени.

Глава 4. Пакеты данных для оценки качества и обучения систем распознавания документов

4.1 Введение

Одним из самых слабо исследованных и острых вопросов в создании систем распознавания документов, удостоверяющих личность, является формирование пакетов данных для обучения и тестирования систем в целом и отдельных методов. Это связано с природой информации, содержащейся в документах: в большинстве стран законодательно запрещен сбор и публикация персональных и биометрических данных. Приходится констатировать практически полное отсутствие таких пакетов в открытом доступе для исследователей, что приводит к невозможности объективного сравнения различных методов и подходов к решению задач в этой области. В этой главе будет предложена методика, сформированы пакеты публичных данных для проведения исследований в этой области и предложения по их использованию.

4.2 Пакеты данных для обучения систем распознавания

4.2.1 Методы синтеза данных для обучения и настройки алгоритмов распознавания

Качество распознавания документов, в том числе качество распознавания текстовых реквизитов или графических объектов, напрямую зависит от качества обучения отдельных классификаторов. Примером подобного классификатора может являться классификатор печати, либо классификатор отдельного текстового символа. Начальным этапом построения классификатора служит формирование обучающего множества: к примеру, применительно к классификатору отдельно символов обучающее множество может состоять из образов символов различного вида (черно-белых (бинарных), полутоновых, цветных) и

сответствующих символам атрибутов (код символа в некотором заранее определенном алфавите, признаки шрифта (жирность, курсивность, гарнитура)). Совокупность атрибутов определяет *алфавиты* обучения и классификации. Результативность обучения классификатора, т.е. достижения высокой точности распознавания и монотонность оценок надежности распознавания, сильно зависит от объема обучающего множества и от точности соответствия установленных атрибутов символов.

Вообще говоря, все указанное также относится и к тестовому множеству, необходимому для проверки качества обучаемого классификатора, однако в рамках данного раздела мы будем рассматривать только обучающие множества. Процесс обучения, как правило, рассматривается как итерационная процедура, т.е. для первичного обучения и последующих сеансов дообучения используются различные обучающие множества.

Принципы и способы построения обучающих множеств

При построении обучающего множества нужно особое внимание обратить на репрезентативность данных. В работе [272] уделено внимание основным принципам формирования обучающего множества:

- *достаточность* – число обучающих примеров должно быть достаточным для надежного обучения. Разумеется, число примеров может быть различным для разных моделей обучения. Например, для нейронной сети необходимо, чтобы число обучающих примеров было в несколько раз больше, чем число весов межнейронных связей, в противном случае модель может не приобрести способности к обобщению [272]. В реальности достаточность оценивается с помощью характеристик обученного метода, например, исследованием зависимости точности распознавания от числа обучающих примеров в предположении, что график этой зависимости монотонен.
- *разнообразие* – большое число разнообразных возможных комбинаций признаков в обучающих примерах. Этот принцип тесно связан с предыдущим, он усиливает требования к числу обучающих примеров, в которых явно оцениваются используемые в классификаторе признаки

и их комбинации. Оценка разнообразия может быть проведена с помощью кластеризации, использующей представления образа в виде набора признаков, разнообразие оценивается, например, зависимостью числа получившихся кластеров от числа обучающих примеров.

– *распределение частот классов* – в зависимости от типа используемого метода классификации, примеры различных классов зачастую должны быть представлены в обучающей выборке примерно в пропорциях, соответствующих пропорциям классов в тестовой выборке. Преобладающие классы будут определены как более вероятные для новых наблюдений. Так, к примеру, при создании классификатора стандартных текстов определенного языка разумно ориентироваться на частотность распределения встречаемости отдельных символов [273].

Набор образов символов для обучения классификатора может быть сформирован различными способами: сгенерирован искусственно, извлечен из изображений (тех, которые будут распознаваться, или из похожих изображений). Далее необходима «разметка» образов символов, состоящая в приписывании каждому образу его атрибутов, как минимум кода символа (также в литературе встречается термин «аннотирование»). Разметка может проводиться как автоматически, так и вручную, при создании классификаторов для распознавания документов, ручная разметка обязательна, если требуется обучение на примерах, извлекаемых из последовательности случайных образов документов. Ручная разметка может быть основана как на предъявлении оператору (лицу, осуществляющему разметку) отдельного символа, так и на предъявлении части образа документа с контекстным окружением этого символа. Последний способ разметки является более точным, нежели первый, но он требует больше времени оператора на анализ каждого символа.

Задача формирования обучающего множества образов символов состоит в получении как можно больше надежно размеченных разнообразных образов. Другими словами, задача сводится к следующему:

– оценка объема множества и, возможно, оценка количества комбинаций признаков;
– минимизация ошибочно размеченных образов, а также оценка доли ошибочно размеченных образов;
– оценка соответствия распределения частот классов заранее заданному распределению, в случае необходимости.

Применительно только к одной из задач классификации, обязательной в системах распознавания документов, в том числе документов, удостоверяющих личность – к задаче классификации печатных символов, задача построения множества образов для обучения, может уже быть довольно трудоемкой и затратной [274; 275]. В процессе формирования обучающего множества из реальных данных приходиться решать подзадачи предварительной обработки, такие как поиск символов, удаление шумов и посторонних объектов. Сложность построения множества образов отдельных символов отмечена в работе [276], в которой отмечается, что качество обучающего множества напрямую зависит от точности алгоритма поиска границ символа: каждый символ должен быть строго центрирован, одни и те же символы должны иметь одинаковые размеры.

В работах [274; 275] указано на создание специализированных форм, при заполнении которых необходимо придерживаться определенных правил. Указанный подход к формированию обучающего множества требует больших человеческих затрат. На этапе заполнения формы, как было отмечено ранее, важно получить как можно более широкий диапазон вариантов написания каждого класса образов. Эта особенность требует большего числа респондентов, что делает этот подход к созданию обучающего множества дорогим и неэффективным.

В работе [277] рассмотрены различные способы получения базы графических символов. Отмечено явное преимущество сохранения образов символов непосредственно из программы распознавания документов. Например, при такой схеме создания множества образов достаточно точно определяются границы каждого символа. Естественно, возникает вопрос о надежности распознавания отдельного символа.

Модель процесса формирования обучающего множества из реальных данных

При использовании реальных данных (т.е. не искусственно синтезированных образов распознаваемых объектов, а извлеченных из объектов реального мира) для создания репрезентативных обучающих множеств большого объема, требуется значительное количество ресурсов для осуществления разметки. В

случае применения ручной разметки возникает отдельная задача: каким образом при имеющихся ресурсах операторов за ограниченное время создать обучающее множество наибольшего объема. В этом подразделе будет рассматриваться модель процесса формирования обучающего множества большого объема на примере задачи формирования из реальных данных пакета данных для обучения классификатора одиночных текстовых символов.

Пусть существует источник образов, из которого поступают как образы символов, так и образы «несимволов» (т.е. образы объектов, которые должны определяться классификатором как не принадлежащие ни к одному заранее установленному классу символов). Рассмотрим два механизма разметки образцов символов, поступающих из некоторого источника: автоматический классификатор образов, построенный по принципу оптического распознавания символов (OCR) и разметка операторами образов, которые могут быть как предварительно классифицированными, так и не классифицированными.

Автоматическая разметка проводится быстро, но сопряжена с ошибками. Ручная разметка является более точной, но ограничена скоростью работы операторов.

Рассмотрим подробнее процедуру разметки набора образов символов. Разметка может включать в себя следующие операции:

- изменение образа символа;
- изменение границ символа;
- проверку границ символа с возможной *отбраковкой* (удалением образа из обучающего множества);
- ввод кода символа;
- проверку кода символа с возможной отбраковкой.

Приведенные операции упорядочены по убыванию затраченного на операцию времени. Время на выполнение операции проверки кода символа с возможной отбраковкой может быть уменьшено, если оператору подаются на разметку ранее классифицированные символы с одинаковым кодом. Необходимо отметить, что функция отбраковки уменьшает время работы оператора, но в то же самое время может уменьшать разнообразие признаков в образах обучающего множества.

Как уже отмечалось выше, оператору может предъявляться как отдельный образ, так образ в контексте соседних символов, в последнем случае *точность* разметки, определяемая как отношение количества ошибочно разме-

ченных образов к общему количеству образов, повышается за счет увеличения расхода времени оператора на анализ группы символов.

Время выполнения операций варьируется от $0{,}3 - 0{,}5$ секунды для операции проверки кода символа с возможной отбраковкой (в случае, когда предъявляются однородные образы, которые заранее отсортированы по коду символа и иным признакам) до 30 и более секунд для операции изменения образа символа.

Рассмотрим разбиение множества M, извлеченного из некоторого источника образов, на три подмножества $M_a \cup M_v \cup M_e$, где

- M_a – множество уверенно классифицированных образов символа – эти образы не подлежат ручной проверке;
- M_v – множество образов, требующий ручной проверки, во-первых, факта принадлежности к символам, и, во-вторых, правильности классификации.
- M_e – множество образов, которые не могут быть классифицированы автоматически и которые оператор при ручной проверке должен классифицировать заново.

Зададимся оценками времен t_v и t_e обработки одного образа оператором из множеств M_v и M_e соответсвенно. Тогда общее время обработки оператором множеста M определится как

$$t = |M_v| \cdot t_v + |M_e| \cdot t_e. \tag{4.1}$$

Способ разбиения множества M задаст время обработки t. Для больших объемов множеств время обработки почти всегда ограничено. Например, для $|M| = 1\ 000\ 000$ образов, при $t_e = 0{,}5$ сек, общее время обработки каждого символа составит примерно 18 дней. Отсюда следует, что для описанной работы не удастся ограничиться одним оператором, и что необходимы средства автоматизации процесса формирования обучающего множества.

Нередко возникают затруднения при классификации похожих символов, например, необходимость различать буквы «O» и цифры «0». Для решения этой проблемы необходимо обратиться к образу текстового поля, из которого был получен символ. Оператору должна быть доступна исходная текстовая строка с указанием текущего символа в текстовом поле. Таким образом, контекст поля существенно повышает качество обучающего множества.

Из вышесказанного следует, что способ разбиения множества M на подмножества M_a, M_v, M_e и способ представления элементов этих множеств позволяют минимизировать время, затраченное на обработку оператором подмножеств M_v, M_e, и минимизировать количество ошибок классификации образом множества M. Отметим, что способ разбиения множества M на подмножества также может включать ручные операции, которые необходимо учесть при оценке общих затрат времени.

Способ формирования множества образов символов в процессе эксплуатации OCR-системы

В задаче распознавания документов, например, при сохранении в архиве потока образов документов, ошибочно распознанные образы должны быть исправлены или, как минимум помечены как ненадежно распознанные. Этапы верификации и редактирования результатов распознавания обусловлены бизнес-логикой системы распознавания документов [278]. Эти этапы проводятся силами операторов из организации, эксплуатирующей OCR-систему.

Достаточно часто результаты распознавания документов, то есть документы в цифровом виде, не могут быть переданы разработчикам OCR-системы из организации, эксплуатирующей OCR-систему, прежде всего, по требованиям информационной безопасности (в особенности, если речь идет о документах, удостоверяющих личность). Однако результаты распознавания, состоящие из множества $M_a \cup M_v \cup M_e$, не позволяют восстановить исходные документы, и могут быть переданы разработчикам OCR-системы для повторного обучения классификаторов.

То есть процесс разбиения на $M_a \cup M_v \cup M_e$ осуществляется на технических средствах организации, эксплуатирующей OCR-систему. Несмотря на использования для разбиения результатов верификации и редактирования, операторам не приходится делать никаких новых специальных действий.

Предлагаемый в данном подразделе способ формирования множества образов символов основан на использовании результатов распознавания текстовых полей и тестовых строк, подтвержденных оператором.

Рассмотрим задачу посимвольного сопоставления результата распознавания строки (набор альтернатив с весами для каждого образа) и соответствующей последовательностью символов, подтвержденной оператором на этапах редактирования и верификации. То есть каждому символу текстовой строки нужно соотнести образ символа распознаваемой строки.

Решение задачи, то есть сопоставление результата распознавания с набором символов, будем производить методом динамического программирования с метрикой Левенштейна [279]. Возьмем за основу базовые принципы алгоритма MCHSR [280]. MCHSR является одним из методов контекстной обработки результатов распознавания. Этот алгоритм был разработан для поиска вхождения фрагмента текста в строке результатов распознавания. Мы же рассматриваем задачу полного наилучшего сопоставления подтвержденной строки с результатом распознавания.

На первом шаге алгоритма построим таблицу, в ячейках которой будет указано качество классификатора символа, если символ совпадает с одной из альтернатив для текущего знакоместа. Если текущий символ отсутствует в списке альтернатив, то клетку таблицы оставим пустой.

На втором шаге алгоритма найдем наилучший путь (путь наибольшего веса) из левой нижней точки таблицы в правую верхнюю точку. Разрешены следующие переходы (см. рис. 4.1):

- вверх по ребру ячейки таблицы – случай, когда среди результатов распознавания отсутствует альтернатива, соответствующая введенному оператором символу (образ символа был не распознан). Для простоты изложения будем считать стоимость перехода равную 0.
- вправо по ребру ячейки таблицы – случай, когда результату распознавания не соответствует ни один из символов строки (фрагмент «мусора» распознан как символ). Аналогично, будем считать стоимость перехода равную 0.
- переход по диагонали ячейки – сопоставление символа с одной из альтернатив знакоместа. Стоимость перехода положим равной значению ячейки.

В результате вышеописанного алгоритма будет сформирован набор соответствий: символ тестовой строки – образ символа. Возможны случаи, когда для символа не найден растровый образ и, наоборот, для растрового образа не найден символ.

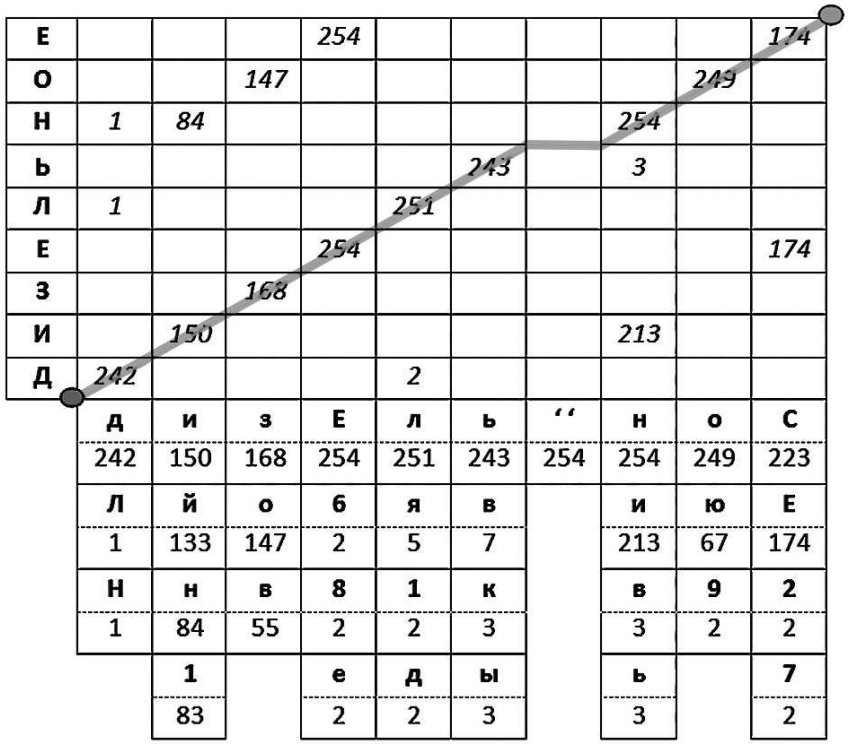


Рисунок 4.1 — Модель сопоставления результата распознавания и текстовой строки, введенной оператором: построение таблицы и поиск лучшего соответствия растровых образов и символов (указано жирной серой линией).

После сопоставления каждый растровый образ можно отнести к одному из трех видов:

– уверенно распознанный образ символа – для данного образа наилучшая альтернатива символа (альтернатива с наибольшим весом) соответствует символу из строки и сама альтернатива имеет высокое качество распознавания;
– образ, требующий подтверждения – для данного образ символа одна из второстепенных альтернатив (не наилучшая альтернатива) соответствует символу из строки;
– «неправильно» распознанный образ символа – образ символа, для которого не найден соответствующий символ из строки.

Таким образом, мы получили три множества образов символов M_a, M_v, M_e.

Сохраняемые символы могут быть представлены как бинарными, так и полутоновыми и цветными образами. В последних случаях для обучения могут понадобиться не только образы как таковые, но и параметры отделения полутоновых и цветных образов букв от фона. В простейшем случае порог отделения фона может быть взят из результатов бинаризации группы символов,

составляющих строку, и уточнен адаптивными алгоритмами расчета порога бинаризации, например, при помощи метода Ниблэка [63].

Проверим предложенную модель и способ экспериментально. Для проведения эксперимента было получено более 2 000 000 символов, среди которых более 82% образов были отнесены к множеству M_a. Доля символов, требующая дополнительную проверку оператором, составила менее 10%, что позволяет существенно ускорить процесс создания обучающего множества. Множество символов M_e составило около 8%. Примеры символов из разных множеств представлены на рис. 4.2.

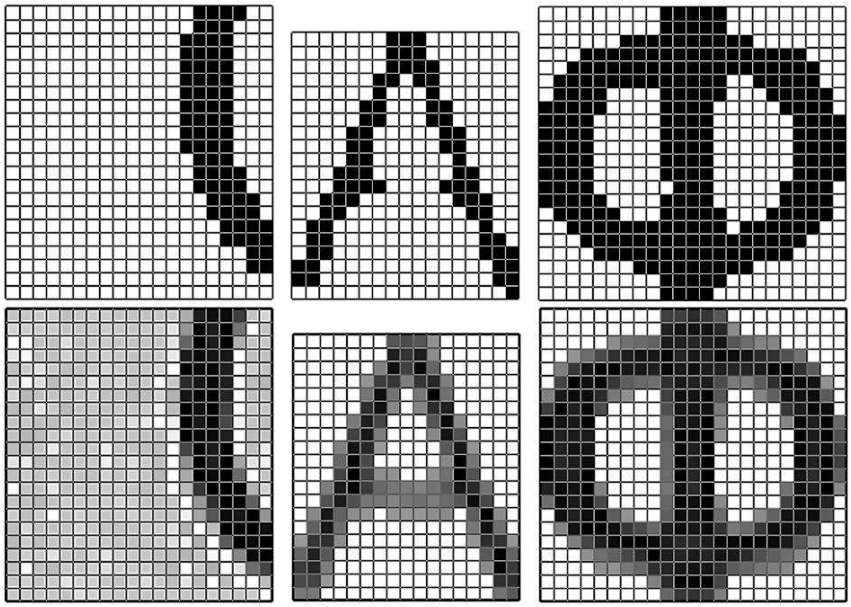


Рисунок 4.2 — Примеры символов из разных множеств. Слева: пример образа, не являющегося символом; справа: символ из множества уверенно распознанных образов; в центре; образ символа, требующий дополнительную проверку оператором.

На этапе отнесения образа символа к одной из трех групп предложено считать символ надежно распознанным, если первая альтернатива имеет высокое качество. Происследуем размер множества образов M_v от зафиксированного качества символа первой альтернативы Q_0. Иными словами, символ не нужно дополнительно подтверждать, если качество первой альтернативы $q > Q_0$, иначе символ попадает в множество сомнительно распознанных образов. На рис. 4.3 представлены результаты эксперимента для документов «Паспорт РФ».

Центральное место в задаче формирования обучающего множества занимает надежность классификации образов. Проанализируем зависимость количества ошибок множества M_a от качества символа первой альтернативы.

Рисунок 4.3 — Зависимость размера (в % от общего числа образов M) базы M_v от минимально допустимого значения качества образа из множества M_a для документов «Паспорт РФ».

Для этого подсчитаем количество ошибок для каждого промежутка значений качества альтернативе на выборке документов типа «Паспорт РФ». Уменьшение доли ошибок с ростом значения альтернативы представлен на рис. 4.4.

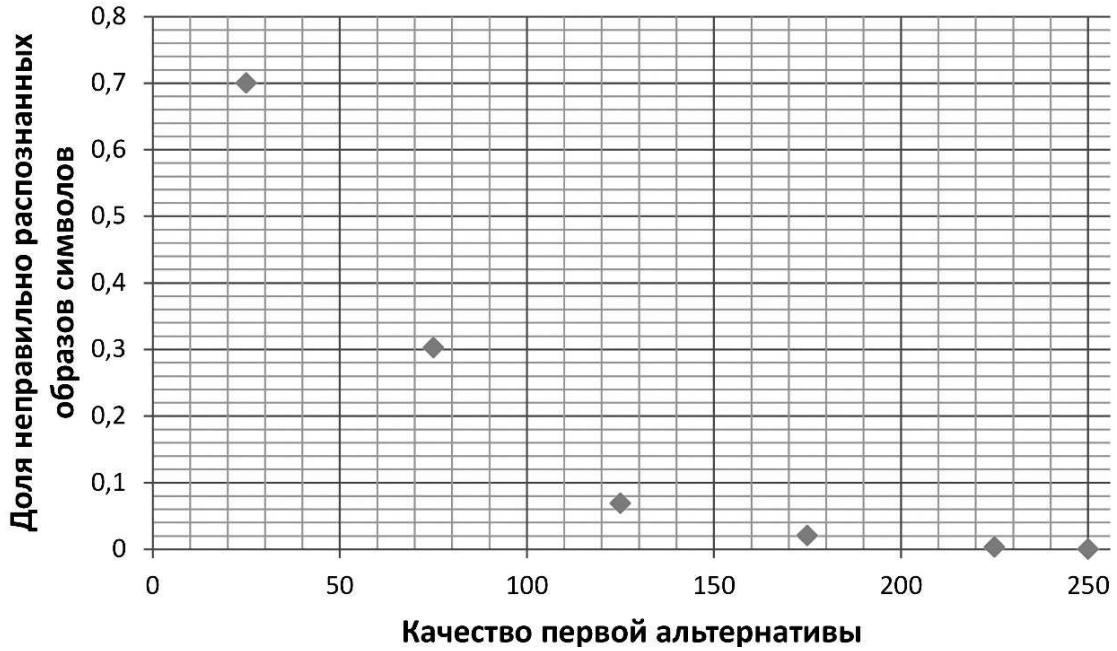


Рисунок 4.4 — Зависимость количества неверно распознанных символов от качества первой альтернативы.

Возникает вопрос, можно ли создать множество M_a, где доля ошибочно классифицированных образов не более заданного заранее числа p? Исследование возникшей проблемы показало (см. рис. 4.5), что с ростом первой

альтернативы Q_0 доля ошибочно классифицированных образов в множестве M_a монотонно убывает. Таким образом, для заданного числа ошибочно классифицированных образов p найдется значение Q_0, при котором множество M_a содержит менее p ошибочно классифицированных образов.

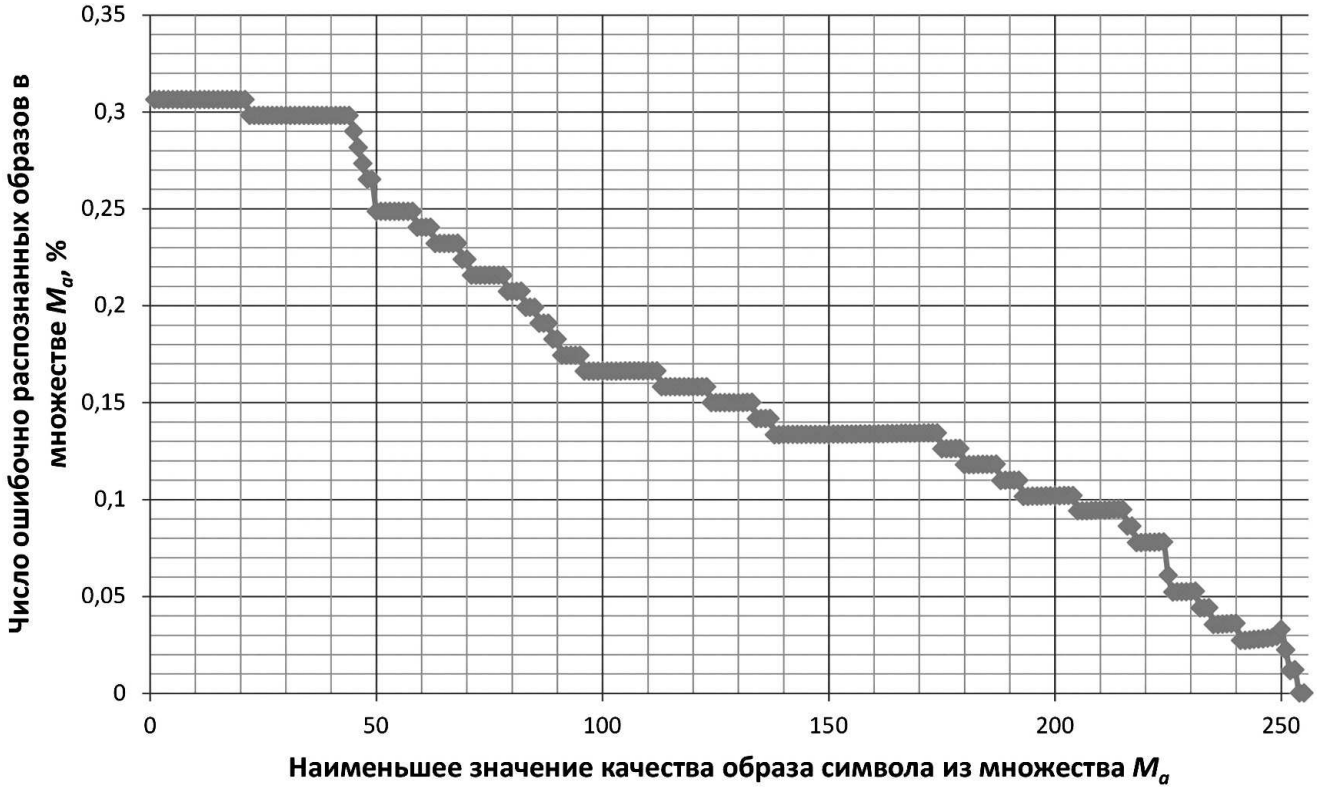


Рисунок 4.5 — Зависимость доли ошибочно распознанных образов в множестве M_a от качества символов, составляющих это множество.

Посчитаем требуемое время на создание множества графических образов предложенным способом, и сравним его с временем классификации каждого символа множества M. Будем производить расчет для образов символов, полученных на выборке документов «Паспорт РФ». Предположим, что необходимо разметить 1 000 000 образов символов, с долей ошибочно классифицированных символов не более 0,1%. По изложенным выше расчетам, разметка всех символов составит 18 дней.

Множество M_a будет содержать менее 0,1% ошибочно классифицированных образов при $Q_0 = 200$ (см. рис. 4.5). Для полученного значения Q_0, размер множества M_v составит 100 000 образов (см. рис. 4.4). Следовательно, время разметки 1 000 000 образов символов составит около $t = |M_v| \cdot t_v \approx 2$ дня.

Не будем забывать, что реальный оператор не может работать 8 часов в день с одинаковой производительностью, что приведет к пропорциональному увеличению затрат времени на разметку обоими способами. Тем не

менее, проведенный расчет показал, что предложенный способ позволяет значительно ускорить процедуру формирования обучающего множества (в 9 раз на приведенном примере). Предложенный способ позволяет совершенствовать классификатор системы OCR, используемой в некоторой организации, в основном, за счет разметки, предусмотренной регламентом работы операторов в процессе этой системы.

4.2.2 Проблемы синтеза искусственных обучающих выборок

Для практической реализации систем распознавания документов, поддерживающих мультиязыковое распознавание, распознавание текстовых полей сложной структуры и т.п. зачастую подход построения множеств образов символов для обучения из реальных данных затруднен или вовсе невозможен. Помимо проблем, связанных с доступностью достаточного количества реальных примеров документов, которые могли бы использоваться как источник образов символов, значительным недостатком также может являться недостаточная репрезентативность имеющихся пакетов реальных данных – к примеру, недостаточное покрытие специальными символами, пунктуацией, или недостаточное покрытие имеющихся реальных данных различными вариантами контекстного окружения символов.

Для решения подобных проблем существует набор подходов, связанных с синтезом искусственных обучающих выборок символов (здесь и далее – генерация *синтетических* примеров). Потенциальные объемы синтетических выборок теоретически не ограничены [281], и такой подход обладает набором преимуществ, связанных с более точным контролем репрезентативности обучающих данных.

Наиболее очевидным подходом к генерации синтетических примеров является автоматическое создание изображений с множеством случайно выбранных символов из некоторого алфавита [282—284] (см. рис. 4.6). Такой подход достаточно прост в имплементации и адаптации к различным языкам. Однако, прямое его применение не принимает в расчет языковую модель, т.е. не позволяет контролировать контекстное окружение символов, порождаемое особенностями построения слов и предложений в естественном языке, что может в

дальнейшем негативно влиять как на качество классификации отдельных символов, так и в целом на качество распознавание текста.

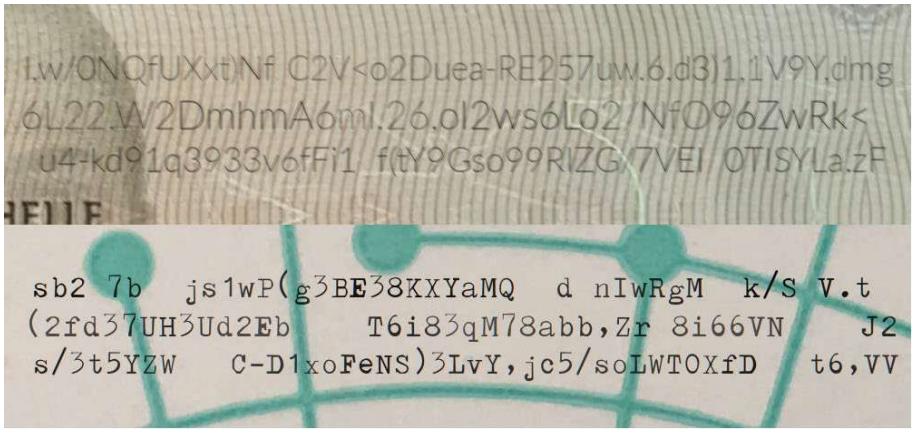


Рисунок 4.6 — Примеры синтезированных изображений для обучения систем распознавания текстовых строк.

Естественным развитием подхода автоматической генерации изображений со случайными образами символов является модификация процедуры генерации таких изображений путем ее дополнения набором правил и ограничений, регулирующих контекстное окружение символов. Примером таких правил могут являться ограничения на написание нескольких одинаковых символов подряд, ограничение на непосредственных соседей символов пунктуации (включая символ «пробела»), запрет на чередование заглавных и строчных букв в рамках слова и т.п.

Реализация подобных правил позволяет повысить репрезентативность синтезированных цепочек символов по отношению к целевому языку. Однако внедрение таких модификаций значительно усложняет механизм генерации обучающей выборки, а сложность наборов правил может привести к трудоемкости или невозможности масштабирования такой системы, к примеру, с целью адаптации для других типов целевых документов или для других языков.

Перспективным направлением в этой области является использование обучаемых заранее языковых моделей для дальнейшей генерации синтезированных изображений [285; 286]. Языковая модель, представленная, к примеру, в виде нейронной сети, может быть использована для построения абстракции над частотными характеристиками символов, слов, контекста, синтаксических и семантических связей. Схожесть сгенерированных таким образом текстов с естественными позволяет повысить репрезентативность синтезированных выборок

и, как следствие, качество итоговых классификаторов и систем распознавания. Этот подход, однако, также обладает рядом недостатков, а именно:

- «запоминание» близких синтаксических и семантических связей может быть нежелательно в случае, если в целевых объектах распознавания таких связей нет или они обладают слабовыраженным характером;
- для обучения языковых моделей снова требуются уже свои обучающие наборы данных, при этом – уже реальных данных, и в случае если имеющиеся обучающие наборы нерепрезентативны, это естественным образом может привести к нерепрезентативности финальных классификаторов;
- затрудняется возможность ручного контроля синтаксических и семантичесских правил.

Согласно эмпирическим данным, качественные характеристики обучающих наборов, построенных таким образом, все же превосходят таковые у наборов, построенных при помощи ручных модификаций [287], что может свидетельствовать о том, что вектор развития области синтеза обучающих данных для распознавания текста будет двигаться в сторону обучаемых языковых моделей.

Другим важным аспектом процедур синтеза искусственных обучающих выборок является *аугментация* – процесс увеличения объема обучающего множества образов путем применения геометрических искажений объекта или всего образа, добавления пиксельных шумов с различными параметрами и пр. Аугментация является особенно ценным подходом для генерации синтетических обучающих баз для обучения нейросетевых классификаторов. Так, в работе [288] авторы используют параллельный перенос образа, горизонтальное отражение и модификации цветовых компонент пикселей для увеличения объема обучающей базы классификации изображении произвольного вида, в работе [289] авторы смогли достичь наилучшего (на момент публикации) качества классификации символов публичной базы MNIST [117] с помощью применения эластичных искажений образов из обучающей выборки, другие коллективы применяют различные комбинации трансформации изображений [290], специальные нейросетевые модели для генерации искаженных обучающих данных [41] и другие подходы. Отдельная область исследований посвящена методам построения обучающих выборок с генерацией обучающей базы по одному исходному примеру каждого класса с последующей аугментацией [291; 292].

Главным вопросом при применении метода аугментации при синтезе обучающих выборок является выбор техник и параметров искажений, вносимых при генерации данных. Естественно полагать, что для достижения наибольшей репрезентативности синтетической обучающей базы, при выборе наборов и параметров искажений нужно руководствоваться параметрами целевого объекта и параметрами тех возможных искажений, которым может быть подвержен образ, который при эксплуатации системы будет подаваться на распознавание. Тем самым возникает задача определения моделей искажений объектов, подлежащих классификации, применительно к конкретным системам распознавания, или к конкретным целевым объектам – и хотя в литературе достаточно много внимания уделяется классификации таких объектов как печатные символы, в промышленных системах распознавания документов, так же как и в системах распознавания документов, удостоверяющих личность, возникают задачи детектирования, локализации и классификации других типов объектов, таких, как печати, штрих-коды специального вида, гербовые элементы и т.п., для которых задачу синтеза обучающих выборок в целом, и, в частности, подбора параметров искажений при аугментации, нельзя считать решенной.

Способ аугментации образов оттисков печатей для построения классификаторов и детекторов

В данном подразделе будет описан способ генерации и аугментации множества изображений оттисков круглых печатей для обучения детекторов и классификаторов. В первую очередь следует определить, что в рамках конкретной задачи построения обучающей выборки, следует подразумевать под термином «изображение оттиска круглой печати», и каким набором инвариантов они обладают. Одним из очевидных геометрических свойств оттисков круглых печатей является инвариантность к повороту вокруг своего центра. Поскольку при выполнении оттиска печати могут использоваться чернила различного цвета, оттиски могут обладать разными цветовыми характеристиками, а также ввиду деградации чернил или ввиду некачественного оттиска, яркость образа печати может варьироваться и быть неравномерной. Примеры оттисков круглых печатей приведены на рис. 4.7.

Рисунок 4.7 — Примеры оттисков круглых печатей.

В рамках предлагаемого способа ограничимся методов моделирования искажений, связанных с геометрическими преобразованиями и различиями яркости оттисков, при этом не принимая во внимание цветовой тон чернил. Здесь и далее будем считать, что пример образа оттиска круглой печати можно признать корректным, если видны и различимы не менее 75% точек оттиска (т.е. не более 25% объема чернил перестали быть различимы вследствие деградации носителя или некачественным приложением печати).

Для последующего анализа степени влияния различных моделируемых искажений и их комбинаций, введем преобразования A_R, A_G, и A_F, где A_R – преобразование вращения, A_G – глобальное преобразование интенсивности пикселей и A_F – моделирование локального изменения контраста.

Преобразование вращения A_R можно задать следующим образом:

$$A_R(x, y) = R(x - x_c, y - y_c) + (x_c, y_c), \tag{4.2}$$

$$R = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}, \tag{4.3}$$

где θ – угол вращения, и (x_c, y_c) – центр вращения.

Для определения A_G предлагается использовать случайное монотонное преобразование интенсивности [293] для моделирование изменений яркости оттиска печати. Пусть $f_j(h)$ – семейство монотонных кусочно-линейных функций с изломами в $h = 0, \Delta_j, 2\Delta_j, 3\Delta_j, \ldots$, где $\Delta_j = 2^{-j}$ и $j = 0,1,\ldots,M$. Функции $f_j(h)$ обладают следующими свойствами:

$$f_j(0) = f_{j-1}(0) = 0, \quad f_j(1) = f_{j-1}(1) = 1; \tag{4.4}$$

$$f_j(2k\Delta_j) = f_{j-1}(2k\Delta_j), \quad \forall k \in \{0, 1, \ldots, 2^{j-1}\}; \tag{4.5}$$

$$f_j((2k+1)\Delta_j) = r(f_j(2k\Delta_j), f_j((2k+2)\Delta_j)), \quad \forall k \in \{0, 1, \ldots, 2^{j-1} - 1\}, \tag{4.6}$$

где r – ограниченное нормальное распределение, такое, что:

$$r(a,b) = \max\{\mu - S, \min\{\mu + S, N_j\}\}, \tag{4.7}$$

$$S = \psi_2(b-a), \quad \psi_2 \in [0,1], \tag{4.8}$$

$$N_j \sim \mathcal{N}(\mu, \sigma^2), \tag{4.9}$$

$$\mu = \frac{b+a}{2}, \quad \sigma = \psi_1(b-a), \quad \psi_1 \geqslant 0, \tag{4.10}$$

и где ψ_1, ψ_2, M – параметры алгоритма построения семейства функций. Примеры построенных функций такого вида представлены на рис. 4.8.

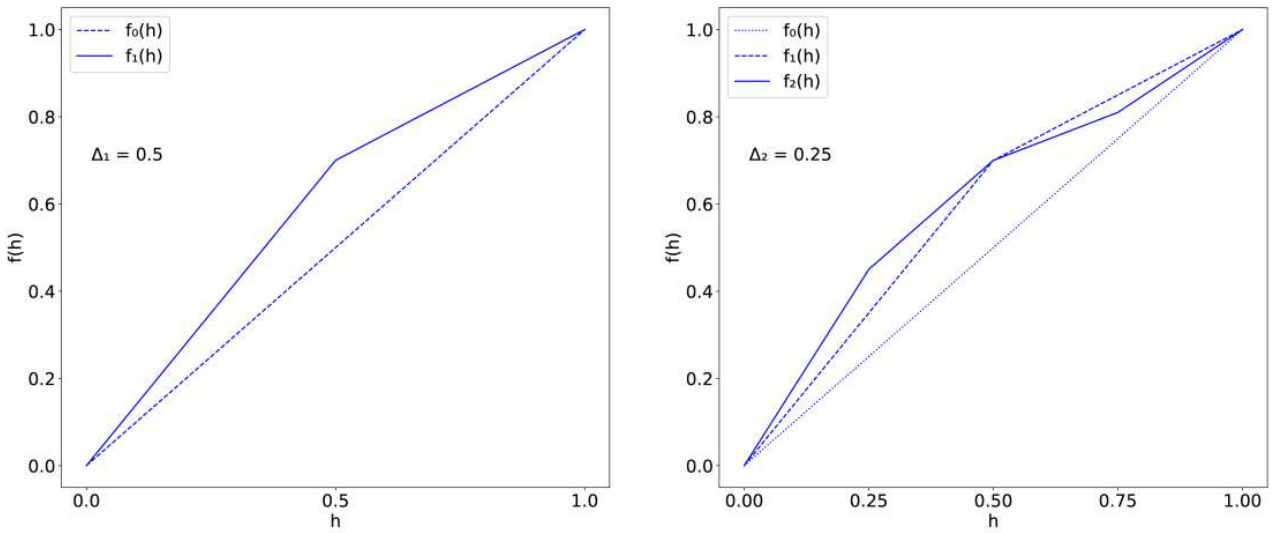


Рисунок 4.8 — Примеры пошагового построения семейства случайных монотонных преобразованийй интенсивности пикселей.

Глобальное преобразование интенсивности пикселей теперь можно задать в виде:

$$A_G(x,y) = f_M(h(x,y)), \tag{4.11}$$

где $h(x,y) \in [0,1]$ – значение интенсивности пикселя с координатами (x,y).

Для моделирование локального изменения яркости вследствие некачественного оттиска или деградации носителя, предлагается использовать следующий подход, опирающийся на учет расстояния от точки приложения силы. Пусть (x_c, y_c) – центр круглой печати, r – ее радиус. Определим точку приложения силы (ξ_x, ξ_y):

$$\begin{cases} \xi_x, \xi_y \sim \mathcal{U}_{[0,2r]} \\ (\xi_x - x_c)^2 + (\xi_y - y_c)^2 \leqslant r^2 \end{cases}. \tag{4.12}$$

Преобразование локального изменения яркости определим в следующем виде:

$$A_F(x, y) = h(x, y) \cdot w_{\text{obj}} + w_{\text{bg}} + \tau, \tag{4.13}$$

$$w_{\text{obj}} = 1 - \frac{d}{2r}, \quad w_{\text{bg}} = 1 - w_{\text{obj}}, \tag{4.14}$$

где $h(x, y) \in [0, 1]$ – значение интенсивности пикселя с координатами (x, y), d – Евклидово расстояние между точками (ξ_x, ξ_y) и (x, y), и τ – случайный шум.

Для анализа влияния описанных методов аугментации обучающей базы образов оттисков круглых печатей, проведем экспериментальное исследование с использованием открытого пакета данных «SPODS» [294], который содержит изображения документов с различными графическими атрибутами, такими как подписи, печати и логотипы. Примеры изображений из пакета данных «SPODS» представлены на рис. 4.9.

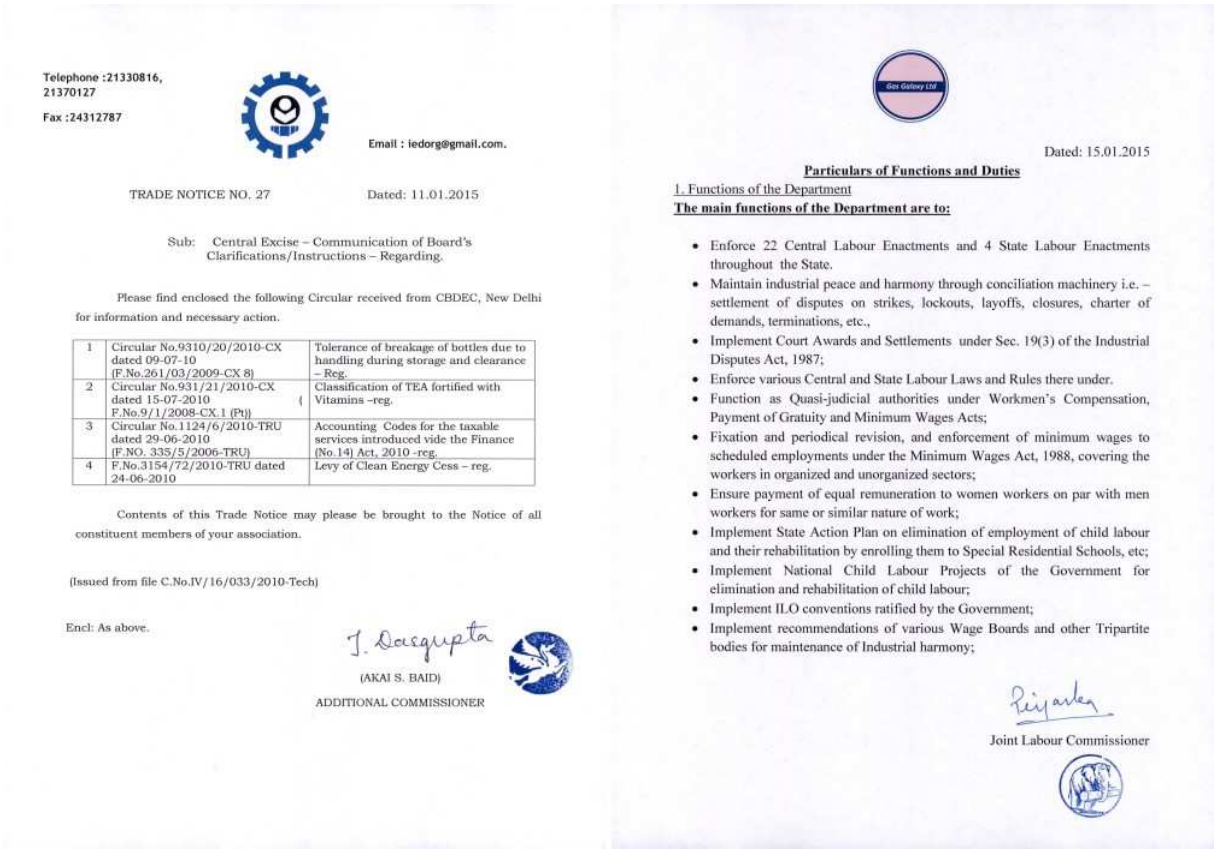


Рисунок 4.9 — Примеры изображений документов из открытого пакета данных «SPODS» [294].

Из открытого пакета данных «SPODS» можно выделить 163 изображения, с оттисками круглой печати. Для исходного множества образов обучающей базы выберем 5 классов оттисков, используя 158 в качестве тестового подмножества.

Множество исходных обучающих образов, тем самым, состоит из 5 печатей, представленных на рис. 4.10.

Рисунок 4.10 — Исходное обучающее множества для детектирования оттиска круглых печатей.

В качестве примера обучаемого метода детектирования оттисков печатей рассмотрим классификатора из семейства Виолы и Джонса – одного из широко используемых подходов для решения задач детектирования и классификации графических объектов, применяемых, в том числе, для анализа оттисков печатей [97; 295]. Для контроля качества детектирования будем использовать гармоническое среднее точности и полноты детектирования:

$$F_{\text{mean}} = \frac{2 \cdot \text{Recall} \cdot \text{Precision}}{\text{Recall} + \text{Precision}}, \tag{4.15}$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \tag{4.16}$$

где TP – количество корректно задетектированных оттисков, FP – количество ошибочно задетектированных оттисков, FN – количество пропущенных изображений оттисков, и где оттиск считается задетектированным корректно, если найденный прямоугольник, содержащий оттиск печати, не отклоняется от истинного больше чем на 0,5 в терминах коэффициента Жаккара (отношение между пересечением и объединением прямоугольников).

При обучении детектора Виолы и Джонса только на исходном обучающем множестве из 5-ти образов (см рис. 4.10), значение качества детектирования составляет $F_{\text{mean}} = 0{,}025$, что свидетельствует о нерепрезентативности обучающей выборки.

Для аугментации исходного набора образов для обучения с применением преобразования вращения A_R (4.2) использовались повороты на углы $\theta_j = j \cdot \pi/36$, где $j = 0{,}1, \ldots, 72$. При обучении детектора на аугментированной таким образом обучающей базе качество детектирование составило $F_{\text{mean}} = 0{,}947$, что свидетельствует о значительном влиянии преобразований

вращения в методах аугментации. Однако среди тестового набора образов существовало значительное количество примеров, не задетектированных полученным методом (см. рис. 4.11) ввиду высокой вариативности интенсивности пикселей.

Рисунок 4.11 — Примеры незадетектированных оттисков печатей детектором, обученным на базе с аугментацией преобразованием вращения.

Для аугментации исходного набора образов с применением глобального преобразования интенсивности пикселей A_G (4.11) были использованы параметры $\psi_1 = 0{,}3$, $\psi_2 = 0{,}9$ и $M = 15$. Для каждого исходного изображения были синтезированы 5 изображений с глобальным преобразованием интенсивности. При обучении детектора на обучающей базе, аугментированной только преобразованием A_G (4.11) с указанными параметрами, качество детектирование составило $F_{\text{mean}} = 0{,}182$. Примеры, не задетектированные таким детектором, представлены на рис. 4.12. Ключевой особенностью этих примеров можно считать отсутствие некоторых частей объекта и отсутствие целостности границ.

Рисунок 4.12 — Примеры незадетектированных оттисков печатей детектором, обученным на базе с аугментацией глобальными преобразованиями интенсивности пикселей.

Для аугментации исходного набора образов с применением моделирования локального изменения яркости пикселей A_F (4.13) данное преобразование было применено один раз для каждого входного изображения. Примеры полученных

Таблица 17 — Качество детектирования оттисков круглых печатей с аугментациями и их комбинациями

Показатель	Исх.	A_R	A_G	A_F	A_R,A_G	A_R,A_F	A_G,A_F	A_R,A_G,A_M
Precision	1,000	0,986	0,889	1,000	0,949	0,993	0,909	0,969
Recall	0,013	0,911	0,101	0,044	0,943	0,924	0,127	0,987
F_{mean}	0,025	0,947	0,182	0,085	0,946	0,957	0,222	**0,978**

преобразованных изображений представлены на рис. 4.13. При обучении детектора на обучающей базе, аугментированной только преобразованием A_F (4.13), качество детектирование составило $F_{\text{mean}} = 0{,}085$.

Рисунок 4.13 — Примеры изображений оттисков печатей после применения локального преобразования яркости пикселей A_F (4.13).

Качество детектирования оттисков круглых печатей при обучении детектора Виолы и Джонса с применением описанных преобразований аугментации и всеми их комбинациями, представлено в таблице 17.

Можно заметить, что наибольший прирост качества детектирования достигается при обучении детектора на базе с применением аугментации преобразованием вращения A_R (4.2), что объяснимо природой признаков Хаара, применяемых в классификаторах и детекторах семейства Виолы и Джонса. Наилучшие качественные показатели детектирования оттисков печатей достигаются при обучении с применением всех трех преобразований аугментации (см. последний столбец таблицы 17).

Таким образом, как было показано в подразделе, подход аугментации для увеличения объема выборки для обучения классификаторов и детекторов имеет важное значение, в случае, если достаточно репрезентативное обучающее множество невозможно составить из имеющихся реальных данных. Как было показано на примере задачи детектирования оттисков круглых печатей на документах, наиболее высокое качество обучаемых классификаторов и детекторов

можно достич, используя комбинации различных преобразований при генерации синтетической обучающей базы.

4.3 Оценка качества работы систем распознавания идентификационных документов

Одной из наиболее важных задач при разработке системы распознавания является объективная количественное оценивание результатов распознавания таких систем. Правильный выбор способа оценивания результатов распознавания имеет важное значение как для разработчиков мобильных систем распознавания, так и для конечных пользователей.

Методология оценки качества работы систем распознавания идентификационных документов, представленная в данной главе, задается критериями (сущности, которые необходимо оценить), соответствующими показателями (конкретное значение, определяемое для выбранного критерия) и методам (способ определения соответствующего показателя).

Ключевым критерием распознающих систем является точность распознавания. Несмотря на кажущуюся простоту, данный критерий требует детального раскрытия. Так, при определении точности распознавания необходимо оценить как точность распознавания текстовых полей, так и точность выделения графической информации (изображения документа, фотографии держателя документа, подписей и печатей при их наличии и т. п.). В рамках настоящей диссертационной работы подробно разобраны указанные критерии, приведены соответствующие показатели, которые позволяют получить объективную оценку качества работы систем распознавания идентификационных документов.

4.3.1 Оценка точности локализации документа

Локализация документа на изображении — один из первых основных шагов любой системы распознавания документа, суть которой заключается в определении координат многоугольника, обрамляющего документ.

В рамках данной работы рассмотрим случай, когда документ плоский, прямоугольной формы с заранее известным соотношением сторон, выполненный из твердого материала. При этом углы документа могут быть как прямоугольными, так и скругленными (что характерно для большинства

документов, удостоверяющих личность, соответствующих международной спецификации ISO). Пусть четырехугольник t с вершинами $(0;\ 0)$, $(w-1;\ 0)$, $(w-1; h-1)$ и $(0;\ h-1)$, где w – ширина документа, а h - высота документа, прямоугольник, описывающий шаблон документа. Пусть на исследуемом изображении у нас определены истинное расположение документа (обозначим соответствующий четырехугольник за m), а также найденное с помощью рассматриваемого алгоритма локализации документа (обозначим за q). Тогда расстояние между этими четырехугольниками можно определить, опираясь на меру Жаккара следующим образом:

$$\mathrm{D}_{\mathrm{Jaccard}}\left(q,m\right) = 1 - \frac{\mathrm{area}\left(q \cap m\right)}{\mathrm{area}\left(q \cup m\right)}.$$

В качестве альтернативы можно рассматривать также расстояние на базе меры Жаккара, вычисленное в системе координат идеального расположения документа:

$$\mathrm{D}^{\mathrm{gt}}_{\mathrm{Jaccard}}\left(q,m,t\right) = 1 - \frac{\mathrm{area}\left(Mq \cap t\right)}{\mathrm{area}\left(Mq \cup t\right)},$$

где M обозначает гомографию такую, что $Mm = t$.s

Несмотря на то, что мера Жаккара является одной из наиболее часто встречающихся мер при оценке качества локализации документа, она обладает рядом недостатков. Во-первых, небольшое смещение найденного четырехугольника документа с точки зрения меры Жаккара может быть идентичным неправильно найденному углу документа. Эти ошибки локализации документа принципиально различны для последующих шагов распознавания документа: последняя ошибка приведет к гораздо большему «перекосу» документа после исправления проективных искажений, что далее может отразится на качестве распознавания текстовых полей и выделении графической информации. Во-вторых, в соответствии с определением меры Жаккара нет разницы в между смещением найденных границ наружу или внутрь, тогда как в последнем случае может быть потеряна часть значимой информации. Поэтому наряду с мерой Жаккара используют другие способы оценки расстояния между четырехугольников, опирающиеся на эвклидово расстояние между соответствующими вершинами документа:

$$\mathrm{D}^{0}_{\mathrm{corner}}\left(q,m,t\right) = \max_{\mathrm{i}} \frac{\left\| t_i - Hm_i \right\|_2}{P\left(t\right)}\ ,$$

где H — гомография, такая что $Hq = t$, а $P(t)$ — значение периметра шаблона документа. Принимая в расчет, что четырехугольник документа может

быть обнаружен с точностью до перенумерации вершин, формула определения расстояния между двумя четырехугольников выглядит следующим образом:

$$\mathrm{D}^4_{\mathrm{corner}}\left(q,m,t\right) = \min_{q^{(i)}\in Q} \mathrm{D}^0_{\mathrm{corner}}\left(q^{(i)},m,t\right),$$

где Q — множество перенумераций вершин $\{[a,b,c,d]\,;[b,c,d,a]\,;[c,d,a,b]\,;[d,a,b,c]\}$.

4.3.2 Оценка точности определения типа документа

Задача определения типа документа представляет, по сути, представляет собой задачу многоклассовой классификации. При этом, по своей природе, значимость отдельных классов не выделяется. В такой постановке оптимальным является использование доли верно определенных типов документов, которая определяется следующей формулой:

$$accuracy = \frac{T}{N},$$

где T – количество верно классифицированных документов, а N – общее количество документов, подвергались которые классифиции.

4.3.3 Оценка точности распознавания текстовых полей

Оценка точности распознавания текстовых полей является одной из самых важных оценок любой распознающей системой. Для случая распознавания документов одной из наиболее простых, но в то же время крайне показательных оценок, является доля верно распознанных текстовых полей (PSR, per-string recognition rate), вычисляемая следующим образом:

$$\mathrm{PSR} = \frac{L_{correct}}{L_{total}},$$

где L_{total} задает общее количество текстовых полей на документе, а $L_{correct}$ представляет собой количество верно распознанных полей на изображении документа.

В ряде случаев оказывается полезным перейти от подсчета верно распознанных текстовых полей к более детальной оценке верно распознанных символов. В таких случаях в качестве первого шага необходимо определить способ подсчета расстояния между двумя текстовыми строками. На практике для этой цели широко используется расстояние Левенштейна, которое также встречается в литературе как «редакционное расстояние» или «дистанция редактирования». Фактически, оно определяется как минимальное количество односимвольных операций (вставки, удаления или замены одного символа на другой), необходимых для превращения одной последовательности символов в другую. Тогда доля верно распознанных символов определяется следующим образом:

$$\mathrm{PCR} = 1 - \frac{\sum_{i=1}^{L_{total}} \min\left(\mathrm{lev}\left(l_{i_{ideal}}, l_{i_{recog}}\right), \mathrm{len}\left(l_{i_{ideal}}\right)\right)}{\sum_{i-1}^{L_{total}} \mathrm{len}\left(l_{i_{ideal}}\right)},$$

где $\mathrm{len}\left(l_{i_{ideal}}\right)$ – количество символов в i-ом текстовом поле, а $\mathrm{lev}\left(l_{i_{ideal}}, l_{i_{recog}}\right)$ – расстояние между истинным и распознанным значениями для i-го текстового поля.

4.3.4 Определение точности выделения графических полей

Помимо текстовой информации документы, удостоверяющие личность, также содержат важные графические области. Примерами таких областей являются область фотографии держателя документа, область подписи, печать выдающего органа и т. п. Промышленная распознающая система должна помимо текстовых данных обеспечивать возврат запрашиваемых графических регионов. При этом способ локализации таких графических регионов может быть в некотором смысле примитивным (фактически «вырезать» запрашиваемый регион по заданным координатам) или опираться на алгоритмы компьютерного зрения и машинного обучения (например, выполнять поиск лица на документе с помощью нейросетевых алгоритмов).

Ошибки локализации графической информации могут привести к серьезным проблемам при интеграции систем распознавания в сложные высокоуровневые бизнес-процессы. Например, ошибка локализации фотографии держателя документа может фактически остановить процесс двухфакторной верификации

личности, в рамках которой распознается не только удостоверяющий документ, но и выполняется сравнение селфи с фотографией в документе. Поэтому оценка точности выделения графических полей является важным критерием оценки точности распознавания.

Математически данная задача полностью идентична задаче локализации документа на изображении. Действительно, заданные графические поля определяются на шаблоне документа указанием обрамляющего четырехугольника. Далее необходимо оценить точность нахождения заданных четырехугольников. Следовательно, все методы оценки точности поиска четырехугольника, рассмотренные параграфе оценки точности локализации документа, применимы и в данном случае.

4.3.5 Определение качества проверок подлинности

Оценку качества алгоритмов проверок подлинности документа имеет смысл проводить используя проблемно-ориентированный подход, т.е. оценивая относительные доли ошибок первого и второго родов, оценивая вероятности истинных и ложных срабатываний детекторов тех или иных аномалий на изображениях документов. Подробно вопросы оценки качества проверок подлинности документов были изложены ранее в главе 2 (параграф 2.8).

Ключевым вопросом в объективной оценке способности тех или иных алгоритмов и систем анализа документов выявлять аномалии и проверять подлинность документа, представленного на изображении или предъявляемого в видеопоследовательности, является моделирование типов атак, на обнаружение которых направлены алгоритмы анализа, и подготовка специальных тестовых пакетов данных для объективного сравнения различных подходов.

4.4 Пакеты данных для оценки качества работы систем распознавания

Системный подход к оценке качества работы алгоритмов и систем распознавания объектов основывается на использовании пакетов данных, позволяющих сформировать понимание генеральной совокупности объектов, которые предполагается анализировать. Как результат, существует большое количество различных пакетов данных (в научном сообществе также используется термин «датасет»), состоящих из аннотированных изображений и видеоклипов и содержащих некоторые целевые объекты: отдельные символы или строки текста, лица, животные и т.п. В случае, когда информация, представленная в таких пакетах данных, не содержит конфиденциальных материалов, датасет может быть опубликован в открытом доступе, что позволяет активно использовать его в научном сообществе.

Однако документы, удостоверяющие личность, являются особыми в том смысле, что они содержат конфиденциальную личную информацию. Это создает сложности в нескольких аспектах. Во-первых, хранение любых персональных данных представляет угрозу безопасности в случае их утечки, что приводит к значительному финансовому ущербу как для соответствующих владельцев документов, так и для стороны, ответственной за утечку данных. Во-вторых, люди понимают риск, связанный с утечкой, и их не так легко убедить «поделиться» своими настоящими документами с реальными личными данными с кем-то. В-третьих, непосредственно физический образец документа является собственностью владельца документа, что существенно затрудняет процесс сбора данных для таких пакетов. В-четвертых, общедоступных образцов каждого документа, удостоверяющего личность, очень мало, и большинство из них защищены законами об авторском праве, что делает их непригодными для использования в отдельных задачах, в том числе в научных исследованиях. Наконец, во многих странах является незаконным не только распространение персональных данных, но даже их сбор и хранение без специального разрешения.

Как следствие, хотя в открытом доступе существуют несколько пакетов данных, содержащих изображения документов, удостоверяющих личность (к примеру, часть пакета данных SmartDoc [72] или пакет данных с документа-

ми Бразилии BID Dataset [113]), их репрезентативность на сегодняшний день оставляет желать лучшего. В настоящей работе представлен изложен системный подход к созданию таких пакетов данных, а также приведено описание конкретных открытых пакетов данных, предназначенных для использования в рамках задач распознавания документов, удостоверяющих личность:

- пакет данных MIDV-500 [207], сформированный на базе открытых примеров, представляющий собой коллекцию документов, снятых в стандартных условиях;
- пакет данных MIDV-2019 [208], сформированный на базе открытых примеров, представляющий собой расширение пакета данных MIDV-500 в части условий съемки документов;
- пакет данных MIDV-2020 [296], содержащий полностью синтетические данные, представляющий собой полностью обширную коллекцию клипов, фотографий и сканов документов, удостоверяющих личность, снятых в различных условиях;
- пакет данных MIDV-LAIT [297], содержащий синтетические данные, предназначенный для использования в задачах распознавания документов, содержащих текстовые поля, выполненные с использованием персидско-арабской, тайской и индийской письменностью;
- пакет данных MIDV-Holo, содержащий видеопоследовательности искусственно созданных документов с пленочными дифракционными оптическими элементами защиты, предназначенный для оценки алгоритмов анализа элементов защиты документов, изменяющих свои оптические характеристики в видеопотоке;
- пакет данных DLC-2021 [298], содержащий модели атак на предъявление документов, включая случаи предъявления цветной неламинированной копии документа, полутоновой копии документа, и съемки изображения документа с экрана.

4.4.1 Пакет данных MIDV-500

Для формирования пакета данных MIDV-500 использовалось 50 различных типов документов, удостоверяющих личность: 17 видов идентификацион-

Таблица 18 — Условия съемки и соответствующие идентификатора клипов в пакете данных MIDV-500

Идентификатор	Описание условий съемки
TS, TA	«На столе» – простейший случай, документ лежит на столе однородной текстуры, снято с использованием Samsung Galaxy S3 (GT-I9300) и Apple iPhone 5 соответственно
KS, KA	«На клавиатуре» – документ лежит на клавиатуре, что осложняет детекцию прямолинейных границ, снято с использованием Samsung Galaxy S3 (GT-I9300) и Apple iPhone 5 соответственно
HS, HA	«В руках» – документ держится руками, снято с использованием Samsung Galaxy S3 (GT-I9300) и Apple iPhone 5 соответственно
PS, PA	«Частично виден» – на отдельных кадрах часть документа или весь документ целиком не видны, снято с использованием Samsung Galaxy S3 (GT-I9300) и Apple iPhone 5 соответственно
CS, CA	«Помехи» – сцена съемки помимо самого документа содержит большое количество посторонних элементов, снято с использованием Samsung Galaxy S3 (GT-I9300) и Apple iPhone 5 соответственно

ных (ID) карт, 14 видов паспортов, 13 видов водительских удостоверений и 6 других типов.

На рисунке 4.14 приведены примеры отдельных кадров, представленных в представленном пакете данных, снятые при различных условиях.

Рисунок 4.14 — Примеры отдельных кадров для 5 различных условий (слева направо): на столе, на клавиатуре, в руках, частично виден, помехи.

Таким образом, пакет данных MIDV-500 состоит из 500 клипов, причем продолжительность каждого составляла не менее 3 секунд. Первые 3 секунды каждого клипа были раскадрованы с частотой 10 кадров в секунду, что привело в итоге к созданию 15 000 кадров. Каждый кадр обладал разрешением 1080×1920 пикселей.

Каждый кадр был аннотирован вручную путем указания координат углов документа. Если углы документа не видны на кадре, соответствующая координатная точка экстраполируется за пределы кадра (если документ вообще не виден на кадре, все четыре точки будут лежать за границами кадра). Формат аннотирования приведен на рисунке 4.15.

```
{
  "quad": [ [0, 0],     [111, 0],
            [111, 222], [0, 222] ]
}
```

Рисунок 4.15 — Формат аннотирования каждого кадра в пакете данных MIDV-500

Кроме того, для каждого документа было также проведено вручную аннотирование с указанием координат текстовых и графических полей, а также истинным значением всех имеющихся полей. Формат аннотирования приведен на рисунке 4.16.

Всего в пакете данных было размечено 48 областей с фотографией держателя документа, 40 областей с подписью и 546 текстовых полей. Помимо полей, написанных латинскими буквами с диакритическими знаками, пакет данных содержит поля, написанные кириллицей, а также с использованием греческого, китайского, японского, арабского и персидского алфавитов.

4.4.2 Пакет данных MIDV-2019

Как и в случае с пакетом данных MIDV-500, новый пакет данных MIDV-2019, содержит видеоклипы 50 различных типов документов, удостоверяющих личность, включая 17 удостоверений личности, 14 паспортов, 13

```
{
  "field01": {
    "value": "Erika",
    "quad": [ [983, 450], [1328, 450],
              [1328, 533], [983, 533] ]
  },
  "photo": {
    "quad": [ [78, 512],   [833, 512],
              [833, 1448], [78, 1448] ]
  }
}
```

Рисунок 4.16 — Формат аннотирования документа в пакете данных MIDV-500

Таблица 19 — Новые условия съемки и соответствующие идентификаторы дополнительных клипов в пакет данных MIDV-2019

Идентификатор	Описание условий съемки
DG, DX	«Перекошенный» – документ снят при сильной проективном искажении с использованием Samsung Galaxy S10 (SM-G973F/DS) и Apple iPhone XS Max соответственно
LG, LX	«Слабое освещение» – документ снят в условиях слабого освещения с использованием Samsung Galaxy S10 (SM-G973F/DS) и Apple iPhone XS Max соответственно

водительских удостоверений и 6 других типов документов. Те же самые бумажные образцы, которые использовались в качестве источника пакета данных MIDV-500, также использовались для создания пакета данных MIDV-2019. Для каждого распечатанного документа были сняты видеоклипы в двух новых условиях съемки с использованием двух мобильных устройств. Таким образом, было получено по 4 новых видеоклипа на каждый документ (всего 200 новых видеофрагментов). Новые идентификаторы клипов описаны в таблице 19. Образцы изображений добавленных условий представлены на рисунке 4.17.

Опишем детальнее необходимость съемки документов в новых условиях. Начнем с условий «Перекошенный», при котором документы снимались с сильными проективными искажениями. Требование к системам распознавания документов работать в неконтролируемом режиме иногда приводит к тому, что

пользователи фиксируют документы с большими проективными искажениями — например, чтобы избежать бликов на отражающих поверхностях документа. Хотя методы, выполняющие локализацию документа, пытаются исправить изображение документа перед обработкой, необходимо иметь правильные примеры соответствующих изображений документов в пакетах данных, чтобы объективно оценить применимость таких методов локализации документов.

Рисунок 4.17 — Примеры изображений документов, снятых при новых условиях съемки в пакет данных MIDV-2019: «перекошенные» слева и снятые при «слабом освещении» – справа.

Ко второй партии новых клипов относятся клипы, снятые в условиях слабого освещения, без использования вспышки. В качестве практически важного случая использования подобных условий съемки можно привести пример распознавания удостоверяющих документов пассажиров в поездах дальнего следования проводниками. На полученных таким образом изображениях документов текст по-прежнему виден и может быть различим человеком, однако OCR-системы не всегда успешно справляются с этой задачей. Примеры изображений текстовых полей, вырезанных из отдельных кадров, снятых в условиях низкой освещенности, представлены на рисунке 4.18.

Все клипы сняты в разрешении Ultra HD (2160x3840 пикселей) продолжительностью не менее 3 секунд. Дополнительно первые 3 секунды каждого клипа были раскадрованы с частотой 10 кадров в секунду. Как и в пакете данных MIDV-500, для каждого раскадрованного изображения вручную было выполнено аннотирование.

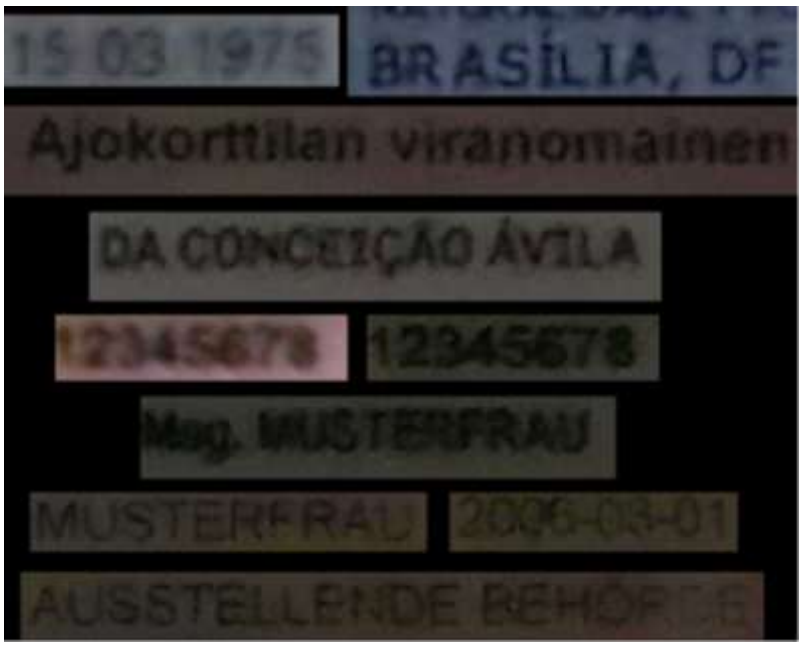


Рисунок 4.18 — Примеры текстовых полей, вырезанные из клипов, снятых в условиях «Слабое освещение».

4.4.3 Пакет данных MIDV-2020

Пакет данных MIDV-2020 базируется на 10 типах документов, которые использовались в ранее описанных пакетах данных MIDV-500 и MIDV-2019. Но, в отличии от них, пакет данных MIDV-2020 ставит своей целью задачу обеспечения вариативности текстовых полей, лиц и подписей при сохранении реалистичности набора данных.

Виды документов, удостоверяющих личность, пакета данных MIDV-2020, перечислены в таблице 20. Для каждого из 10 типов документов, представленных в пакете данных, было создано по 100 образцов документов.

Для создания уникальных образцов документов исходные образцы изображений, полученные из Викисклада (Wikimedia Commons, общее централизованное виртуальное хранилище для изображений, звукозаписей, видеороликов и других мультимедийных файлов, включаемых в страницы проектов Фонда Викимедиа, которые можно свободно распространять, изменять и использовать в любых целях, в том числе и за пределами проектов Викимедиа) были использовали те же исходные изображения, что и при создании пакета MIDV-500, которые впоследствии были с помощью фоторедакторов «очищены» от персональных данных (фамилия, имя, номер документа и т. п.) подписей

Таблица 20 — Типы документов в пакете данных MIDV-2020

№	Код типа документа	Описание	Код PRADO	Код MIDV-500
1	alb_id	Идентификационный документ Албании	ALB-BO-01001	01
2	aze_passport	Паспорт Азербайджана	AZE-AO-02002	05
3	esp_id	Идентификационный документ Испании	ESP-BO-03001	21
4	est_id	Идентификационный документ Эстонии	EST-BO-03001	22
5	fin_id	Идентификационный документ Финляндии	FIN-BO-06001	24
6	grc_passport	Паспорт Греции	GRC-AO-03003	25
7	lva_passport	Паспорт Латвии	LVA-AO-01004	32
8	rus_internalpassport	Внутренний паспорт РФ	—	39
9	srb_passport	Паспорт Сербии	SRB-AO-01001	41
10	svk_id	Идентификационный документ Словакии	SVK-BO-05001	42

и персональных фотографий. Затем новые сгенерированные значения таких персональных данных были добавлены к изображениям с использованием начертания шрифта похожего на исходный шрифт образца документа. Дополнительно для каждого документа была сгенерирована фиктивная подпись, отдаленно напоминающая написание фамилии, а также сгенерировано уникальное изображение лица.

Значения таких полей как пол, дата рождения, дата выдачи и срок действия документа были сгенерированы в соответствии со спецификой стран-эмитентов и заданным распределением возрастных и гендерных параметров:

– 80% документов соответствуют взрослым владельцам (в возрасте от 18 до 60 лет), 10% документов соответствуют пожилым людям (в возрасте от 60 до 80 лет) и 10% – детям и подросткам (до 17 лет) в зависимости

от минимального возраста, допустимого в соответствии со спецификой типа документа;

– гендерное распределение было равномерным: 50% сгенерированных документов соответствуют женщинам, а оставшиеся 50% – мужчинам.

Имена и адреса были сгенерированы с использованием баз данных имен и примеров адресов, доступных в Интернете, с использованием списков имен в Википедии [299] и онлайн-генераторов имен [300].

Искусственно сгенерированные изображения лиц были выполнены с помощью сервиса Generated Photos [301], который в свою очередь использует генеративно-состязательную нейронную сеть StyleGAN [302] в качестве инструмента для создания искусственных изображений лица. Изображения делались либо в цвете, либо в оттенках серого, в зависимости от целевого типа документа. Снимки подбирались с учетом возраста держателя.

Полученные изображения шаблонов сохранены в исходном разрешении и доступны в файле «templates.tar». В общей сложности, файл «templates.tar» содержит 1000 изображений сгенерированных описанным образом уникальных документов. На рисунке 4.19 представлен пример сгенерированного изображения.

Рисунок 4.19 — Пример сгенерированного документа в пакете данных MIDV-2020 (изображение alb_id/00).

Для каждого сгенерированного изображения документа подготовлен соответствующий файл аннотации в формате JSON, выполненный с помощью VGG Image Annotator v2.0.11 [303]. На рисунке 4.20 приведен пример аннотации сгенерированного изображения документа. Для каждого изображения аннотация

включает в себя прямоугольные ограничивающие рамки фотографии владельца документа («фотография»), ограничивающую рамку овала лица («лицо»), ограничивающую рамку поля подписи («подпись») и прямоугольники, соответствующие позициям текстовых полей. Для каждого текстового поля указано его точное значение. Верхняя и нижняя границы прямоугольника текстового поля соответствуют верхней и нижней базовой линии соответственно. Кроме того, для каждого текстового поля предоставляется дополнительная информация, указывающая, содержит ли поле строчные буквы, а также литеры с нижним или верхними выносными элементами. Наконец, для типов документов, в которых присутствует вертикально ориентированные текстовые поля, предусмотрен дополнительный атрибут ориентации, указывающий угол поворота поля против часовой стрелки, выраженный в градусах.

Рисунок 4.20 — Пример аннотации сгенерированного изображения документа. 1 – фотография лица держателя документа, 2 – подпись, 3–13 – значимые текстовые реквизиты, 14 – окаймляющий прямоугольник овала лица.

Подготовленные таким образом изображения-шаблоны распечатывались на глянцевой фотобумаге, ламинировались и обрезались в соответствии со спецификой типа документа. При необходимости, углы документа закруглялись с помощью 4-миллиметрового скруглителя углов. Далее изготовленные бумажные образцы сгенерированных документов использовались для создания сканов, фотографий и видеоклипов документов.

Сканы документов Каждый образец бумажного документа был отсканирован с помощью сканеров Canon LiDE 220 и Canon LiDE 300 в «прямом» и

«повернутом» на некоторый произвольный угол режимах. Все отсканированные изображения имеют разрешение 2480 × 3507 пикселей.

Отсканированные изображения были первоначально сохранены в формате TIFF, а затем преобразованы в JPEG с использованием ImageMagick 7.0.11 с параметрами сжатия по умолчанию. Соответствующие изображения в формате JPEG, включая их аннотации, размещены в пакете данных в архивах «scan_upright.tar» и «scan_rotated.tar». Оригиналы изображений в формате TIFF размещены в архивах «scan_upright_tif.tar» и «scan_rotated_tif.tar». Всего имеется 1000 «прямых» сканов и 1000 «повернутых» сканов. Имена отсканированных изображений соответствуют именам шаблонного изображения, из которого был создан соответствующий бумажный документ.

Аннотации к отсканированным изображениям предоставляются в формате JSON, выполненным с помощью VGG Image Annotator v2.0.11, и содержат ограничивающие рамки овала лица (отмечены меткой «face») и четырехугольника документа (отмечены меткой «doc_quad»). Первая вершина четырехугольника всегда соответствует верхнему левому углу физического документа, а остальные вершины располагаются по часовой стрелке.

Фотографии документов Для каждого физического образца документа была сделана фотография в различных условиях и на два смартфона. Половина фотографий была сделана с помощью Apple iPhone XR, а другая половина – с помощью Samsung S10. При захвате изображений были использованы следующие условия съемки:

- условия низкой освещенности (по 20 документов каждого типа);
- использование клавиатуры в качестве фона (по 10 документов каждого типа);
- естественное освещение, фотографирование документа на улице (по 10 документов каждого вида);
- использование письменного стола в качестве фона (по 10 документов каждого типа);
- использование ткани различной фактуры в качестве фона (по 10 документов каждого вида);
- использование бизнес-документов в качестве фона (по 10 документов каждого типа);

– сильные проективные искажения документа в момент съемки (по 20 документов каждого вида);
– наличие засветов от солнца или лампы, которые скрывают часть документа (по 10 документов каждого типа).

На рисунке 4.21 представлены примеры фотографий документов, снятых при причисленных выше условиях.

Рисунок 4.21 — Примеры фотографий документов в пакете данных MIDV-2020: (a) условия низкой освещенности; (b) использование клавиатуры в качестве фона; (c) естественное освещение, фотографирование документа на улице; (d) использование письменного стола в качестве фона; (e) использование ткани различной фактуры в качестве фона; (f) использование бизнес-документов в качестве фона; (g) сильные проективные искажения документа в момент съемки; (h) наличие засветов от солнца или лампы, которые скрывают часть документа.

Все фотографии сохранены в формате JPEG и имеют разрешение 2268 × 4032 пикселей. Непосредственно изображения с соответствующими аннота-

циями расположены в архиве «photo.tar» пакета данных MIDV-2020. Имена изображений соответствуют именам шаблона, из которого был создан физический документ. Всего 1000 фотографий. Формат аннотаций использовался тот же, что и для аннотирования сканов.

Видеоклипы документов Для каждого образца документа был снят видеоклип с использованием того же распределения условий съемки, что и для фотографий. Каждый клип снимался вертикально, в разрешении 2160 × 3840 пикселей, со скоростью 60 кадров в секунду. Всего было подготовлено 1000 роликов разной длины.

Исходные клипы были разделены на кадры с помощью ffmpeg версии n4.4 с параметрами по умолчанию. Аннотирование подвергался каждый 6-ой кадр данного клипа. Для этого была выполнена соответствующая раскадровка видеоклипа, в рамках которой был сохранен каждый 6-ой кадр и сгенерированы файлы вида «000001.jpg», «000007.jpg», «000013.jpg» и т. д. В самом маленьком видеоклипе в итоге получилось 38 аннотированных кадров, в самом большом — 129 аннотированных кадров. Всего набор данных включает 68409 аннотированных видеокадров.

Кадры с соответствующими аннотациями сохранены в пакете данных в архиве «clips.tar». Аннотация каждого кадра имеет тот же формат, что и для фотографий и отсканированных изображений. Исходные видеофайлы находятся в архиве «clips_video.tar». Все видеоклипы были лишены звукового канала.

4.4.4 Пакет данных MIDV-LAIT

Главной особенностью этого пакета данных является то, что все созданные удостоверяющие документы содержат текстовые поля либо на персидско-арабском языке, либо на тайском языке, либо выполнены с использованием какой-либо индийской письменности. Так, в названии MIDV-LAIT буква «L» означает латиницу (в основном, это английский язык), «A» — персидско-арабскую письменность, «I» — разновидность индийской письменности, а «T» — латиницу. для тайского. Пакет данных MIDV-LAIT содержит изображения удостоверяющих документов помимо латиницы еще в двенадцати письменностях.

Таблица 21 — Список представленных письменностей и языков в пакете данных MIDV-LAIT

Письменность	Язык	Письменность	Язык
Тайская	Тайский	Бенгальская	Бенгальский
Персидско-арабская	Арабский	Гуджарати	Гуджарати
Персидско-арабская	Персидский	Одия (Ория)	Одия (Ория)
Персидско-арабская	Урду	Каннада	Каннада
Персидско-арабская	Урду (Насталик)	Малаялам	Малаялам
Персидско-арабская	Пушто	Деванагари	Хинди
Гурмукхи	Панджаби	Деванагари	Маратхи
Тамильская	Тамильский	Деванагари	Непальский
Телугу	Телугу	Латинская	Английский

Для персидско-арабского письма текстовые строки выполнены в почерках насх и насталик. Полный список систем письма и языков приведен в таблице 21. Важно также отметить, что языки урду и урду (насталик) хотя и одинаковы с точки зрения языка, они отличаются по внешнему виду и написанию, поэтому ряд методов распознавания текстовых строк, эффективных для насха, не работают для насталика. Здесь ситуация очень похожа на распознавание печатных и рукописных текстов — хотя сам язык может быть одним и тем же, методы, которые эффективно справляются с задачами, будут разными. Таким образом, было решено разделить подобные документы на две группы.

Для создания пакета данных MIDV-LAIT, аналогично ранее представленным в данной главе пакетам данных, использовались образцы изображений документов, доступные по лицензии Creative Commons. Аналогично ранее представленному пакету данных MIDV-2020, были использованы синтезированные значения текстовых полей, а также синтезированы лица владельцев документов. С гендерной точки зрения распределение документов равномерное. На рисунке 4.22 представлены примеры синтезированных документов, удостоверяющих личность, включенных в пакет данных MIDV-LAIT.

Подготовленные синтезированные изображения документов были распечатаны на цветном принтере, заламинированы. Полученные бумажные образцы документы были отсняты с помощью Apple iPhone X Pro, в результате чего бы-

Рисунок 4.22 — Примеры синтезированных документов пакета данных MIDV-LAIT.

ли сгенерированы видеоролики продолжительностью не менее 2 секунд. Первые две секунды каждого видеоролика были раскадрированы с частотой 10 кадров в секунду, в результате чего было сгенерировано по 20 кадров для каждого видеоролика.

Каждый полученный кадр был аннотирован указанием координт четырехугольника документа, присутствующего на кадре. Первая вершина четырехугольника всегда соответствует верхнему левому углу физического документа, а остальные вершины располагаются по часовой стрелке. Аналогично ранее представленным пакетам данных, аннотация сохранена в формате JSON.

Для каждого сгенерированного документа произведено аннотирование данных. В частности, указаны: значения и координаты текстовых полей, а также письменность каждого поля. На рисунке 4.23 приведен пример аннотирования текстового поля на арабском языке сгенерированного документа в пакете данных MIDV-LAIT.

Известно, что в персидско-арабской письменности написание выполняется справа налево. При аннотировании наших сгенерированных документов мы строго следовали этому правилу. Однако, некоторые текстовые редакторы не позволяют отобразить такую письменность корректно и инвертируют написание персидско-арабского текста, отображая символы слева направо.

В общей сложности пакет данных MIDV-LAIT состоит из 180 уникальных сгенерированных документов, удостоверяющих личность, на базе которых

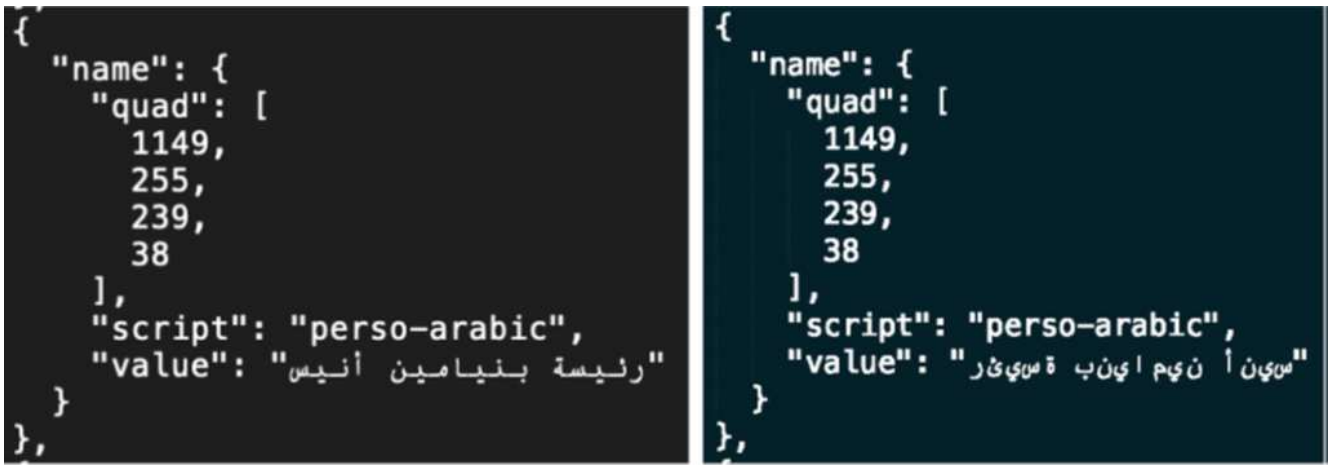


Рисунок 4.23 — Пример аннотирования текстового поля на арабском языке сгенерированного документа в пакете данных MIDV-LAIT.

сгенерировано 3600 изображений. Сгенерированные документы относятся к документам 14 стран. Статистика по странам, документы которых сгенерированы для пакета данных, представлена в таблице 22. Еще раз важно отметить, что документы с арабской письменностью разделены на два класса по почерку: насх или насталик. Насталик используется только в удостоверяющих документах Пакистана. Несмотря на то, что основная цель настоящего пакета данных состоит в создании документов с арабскими текстовыми полями, мы также разметили все поля на латинице. Большинство удостоверяющих документов имеет дублирующие поля на латинице. Например, все паспорта содержат дублирующие поля на английском языке (см. рисунок 4.24). Именно поэтому количество размеченных латинских фактически совпадает с количеством размеченных текстовых полей, выполненных в персидско-арабской письменности.

Рисунок 4.24 — Пример документов, содержащих текстовые поля, выполненные латиницей и персидско-арабским письмом.

Что же касается персидско-арабских текстовых полей, то тут также присутствует большая вариативность в шрифтах. Так, на рисунке 4.25 проиллюстрированы написание имени «Мухxамад» на иранском удостоверении

Таблица 22 — Количество документов и текстовых полей в пакете данных для каждого вида письменности

Письменность	Количество документов	Количество текстовых полей
Тайская	10	1200
Персидско-арабская (Насх)	70	9700
Персидско-арабская (Насталик)	10	400
Гурмукхи	10	600
Тамильская	10	600
Телугу	10	560
Бенгальская	10	600
Гуджарати	10	600
Одия (Ория)	10	600
Каннада	10	640
Малаялам	10	600
Деванагари	10	900
Латинская	165	23 000

личности, выполненное различным шрифтом. Как можно заметить, визуально написание одного и того же имени очень сильно отличается друг от друга.

Рисунок 4.25 — Различные способы написания «Мухаммад» на иранском удостоверяющем документе.

Говоря о цифровых полях, стоит отметить, что в пакете данных MIDV-LAIT такие поля представлены четырьмя видами: западно-арабские, восточно-арабские, персидские цифры и цифры деванагари), как показано на рисунке 4.26. Кроме того, разные удостоверяющие документы в MIDV-LAIT содержат даты в пяти различных календарях: исламском календаре (лунная хиджра), иранском календаре (солнечная хиджра), тайском солнечном календаре, бикрам самбат (официальный календарь Непала) и григорианском календаре.

Рисунок 4.26 — Различные варианты чисел в пакете данных MIDV-LAIT.

4.4.5 Пакет данных MIDV-Holo

Помимо задачи непосредственно распознавания документа, удостоверяющего личность, включающей в себя определение типа документа, извлечение значений его атрибутов, важной задачей является автоматическая проверка действительности. Часть защитных элементов, используемых при производстве документов, удостоверяющих личность, специально предназначена для защиты бланков от копий или воспроизведения. Одним из подходов к защите документов, удостоверяющих личность, является использование OVD (Optically Variable Devices) – элементов с изменяющимися оптическими характеристиками. В ряде таких элементов большую нишу занимают плеточные дифракционные оптические элементы, часто называемые «голограммами», хотя в строгом физическом смысле они голограммами, чаще всего, не являются. Использование таких элементов, к примеру, рекомендовано Европейским Союзом [304] для защиты документов от копирования, поскольку воспроизводить такие элементы защиты невозможно легкодоступными полиграфическими методами.

Оптический эффект пленочных дифракционных оптических элементов достигается за счет взаимодействия света с периодическими рельефными структурами малых периодов (порядка 1 мкм), организованными, как правило, в виде пикселей размером 5–50 мкм. При освещении точечным немонохроматическим источником света эти структуры создают цветовые, динамические и другие характерные оптические эффекты [305]. Пример «голограмм» на документах, удостоверяющих личность, приведен на рис. 4.27.

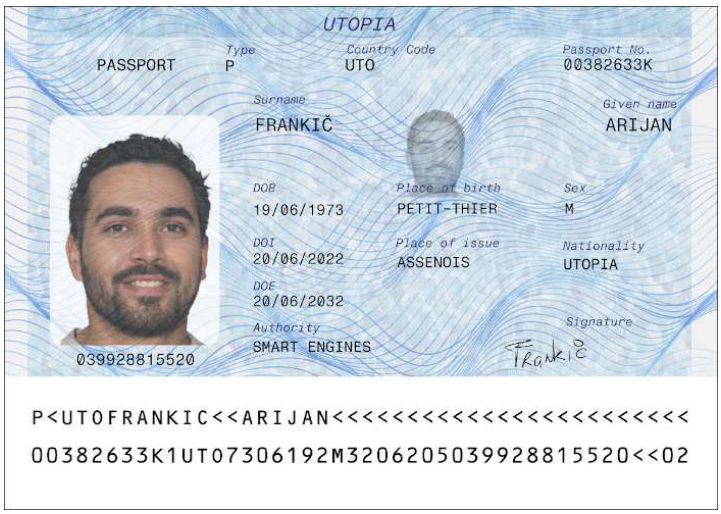
Одной из особенностей задачи автоматической верификации наличия или отсутствия голограммы на документе является необходимость анализировать видеопоток, поскольку одного кадра или фотографии недостаточно для проверки изменяющихся характеристик защитных элементов. В литературе предложено несколько методов определения присутствия голограмм на документах, удостоверяющих личность, однако из-за отсутствия открытых для

Рисунок 4.27 — Пример искусственно созданного паспорта с дифракционными оптическими элементами защиты бланка, сфотографированного под двумя разными углами. Можно заметить, что голографический рисунок изменяет свои цветовые характеристики в зависимости от угла обзора.

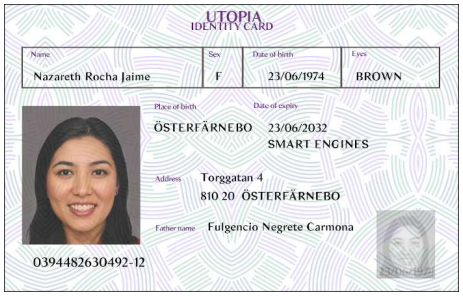
исследователей пакетов данных, содержащих документы с защитными элементами подобного рода, не представляется возможным проводить объективные эмпирические оценки таких методов. В связи с этим, в рамках диссертации был также создан пакет данных MIDV-Holo, содержащий видеопоследовательности искусственных документов, удостоверяющих личность, с голографическими элементами защиты.

Для создания пакета данных MIDV-Holo были предварительно подготовлены пустые шаблоны документов «Утопии» – 10 типов шаблонных паспортов размером 125×88 мм и 10 типов идентификационных карт размером 85.6×54 мм. Размеры были выбраны согласно международным стандартам как наиболее распространенные в мире. Изображения для фона бланка документа были получены из магазина стоковых изображений [306]. Изображения были предобработаны для создания фонов документов в различных цветоных тонах и с различными узорами.

Графические (фотография лица владельца и подпись) и текстовые данные были получены путем автоматической генерации, аналогично тому, как это делалось при создании пакета данных MIDV-2020, описанного выше в секции 4.4.3. Примеры полученных документов представлены на рис. 4.28.



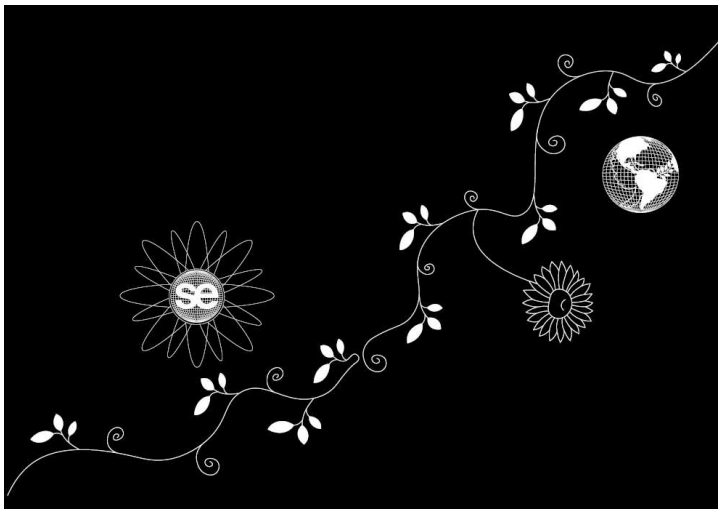
a) Паспорт «Утопии»



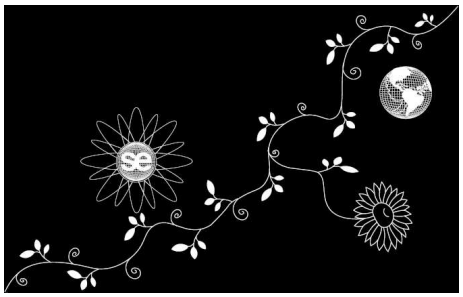
b) ID-карта «Утопии»

Рисунок 4.28 — Примеры полученных шаблонов искусственных паспортов и идентификационных карт.

Для всех шаблонов паспортов и идентификационных карт был создан единый шаблон голографического рисунка. Бинарные маски созданного голографического рисунка представлены на рис. 4.29.



a) Маска для паспортов



b) Маска для ID-карт

Рисунок 4.29 — Бинарные маски созданного голографического рисунка для паспортов и идентификационных карт (белыми пикселями обозначены точки, в которых находится изменяющийся голографический узор)

Пленочные дифракционные оптические элементы с созданным рисунком были изготовлены Центром Тонких Оптических Технологий [307] методом электронно-лучевой литографии, с использованием слоя сульфида цинка под термопластичным материалом с дифракционными рельефными структурами,

для получения прозрачной пленки. Шаблоны изображений паспортов и идентификационных карт были распечатаны на фотобумаге, заламинированы, и покрыты пленкой с голографическими элементами.

Для моделирования сценариев съемки документа, не содержащего голографических элементов защиты, также были подготовлены примеры документов без голографической пленки и примеры документов с имитацией голографический пленки нескольких видов. Примеры изображений «скомпрометированных» документов, соответствующих исходному оригиналу, представленному на рис. 4.27, приведены на рис. 4.30.

Полученные документы были сняты на видео при помощи мобильных устройств iPhone 12 и Samsung Galaxy S10 в различных условиях освещения, общий размер пакета данных составил 700 видеопоследовательностей. Для кадров видеопоследовательности были вручную размечены точные координаты документа на изображении, аналогично другим пакетам данных семейства MIDV. Примеры кадров полученных видеопоследовательностей представлены на рис. 4.31.

Созданный набор данных MIDV-Holo является уникальным в своем роде, поскольку в открытом доступе больше не существует пакетов экспериментальных данных изображений и видеопоследовательностей документов, удостоверяющих личность (или их макетов), которые бы содержали пленочные дифракционные элементы защиты, и которые тем самым позволяли бы объективно анализировать работу алгоритмов автоматической верификации наличия защитных голограмм при анализе видеопоследовательностей документов. Публикация созданного пакета данных MIDV-Holo позволит исследователям в области анализа и распознавания документов вывести уровень разрабатываемых методов и алгоритмов на новый уровень.

4.4.6 Пакет данных DLC-2021

В рамках систем анализа и распознавания документов, удостоверяющих личность, помимо верификации наличия защитных элементов бланка (таких, как дифракционные оптические элементы, задействованные в пакете данных MIDV-Holo, описанном в предыдущей секции), важной задачей является обна-

a) b)

c) d)

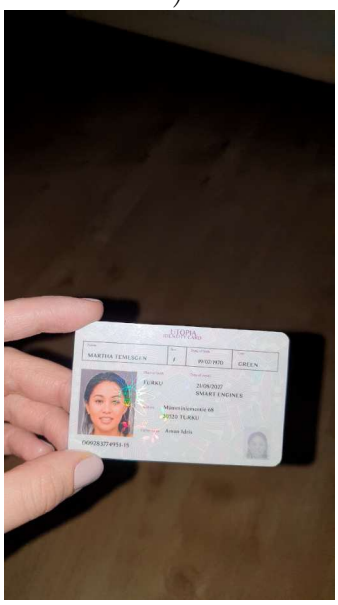
Рисунок 4.30 — Примеры 4х типов моделей «скомпрометированных» документов: a) копия шаблона документа без голографической пленки; b) копия шаблона документа с искусственно добавленным рисунком голограммы; c) распечатанная цветная фотография «оригинального» документа; d) оригинальный документ с наклеенной фотографией другого лица, без голографических элементов.

ружение атак на предъявление документа, где в качестве объекта предъявление может служить не только физическая подделка документа, но и его репродукция, к примеру, на экране. Примеры таких атак возникают в случаях, когда злоумышленники заполучают фотографию подлинного документа и видоизменяют эту фотографию путем ретуширования, после чего печатают при помощи принтера или перефотографируют экран монитора. В литературе такие ситуации называются атаками на предъявление (presentation attack) или атаками ретрансляции (rebroadcast attack). Для оценки алгоритмов, детектирующих

1)

2)

3)



4)

5)

6)

Рисунок 4.31 — Примеры кадров видеопоследовательностей пакета данных MIDV-Holo: 1) «оригинальный» паспорт с офисным освещением и без бликов; 2) «оригинальный» паспорт с офисным освещением; 3) «оригинальная» идентификационная карта с офисным освещением; 4) «оригинальная» идентификационная карта с включенной подсветкой мобильного устройства; 5) «копия» с имитацией голограммы с включенной подсветкой; 6) «копия» с имитацией голораммы с уличным освещением.

такие атаки, был создан открытый пакет аннотированных видеопоследовательностей DLC-2021.

В пакете данных DLC-2021 были рассмотрены три варианта атак: показ копии документа в градациях серого, показ цветной неламинированной копии, и показ фотографии документа на экране монитора. Основой пакета данных DLC-2021 служат документы, изготовленные в рамках работы над пакетом данных MIDV-2020, описанным выше в секции 4.4.3. Примеры изображений паспорта Греции, представленные в пакете данных DLC-2021, изображены на рис. 4.32.

Рисунок 4.32 — Примеры изображений пакета данных DLC-2021. а) оригинальный документ MIDV-2020; б) неламинированная цветная копия; в) неламинированная полутоновая копия; г) съемка документа с экрана.

В общей сложности в пакете данных DLC-2021 представлено 80 различных документов (10 типов, по 8 экземпляров уникальных документов каждого из типов), опубликованных в виде 1424 аннотированных видеопоследовательностей. Характеристики представленных видеопоследовательностей, представленных в DLC-2021, представлены в таблице 23.

4.5 Анализ использования пакетов данных MIDV в научных исследованиях

Пакеты данных семейства MIDV использовались различными исследовательскими группами для проведения экспериментов в области анализа и

Таблица 23 — Количество видеопоследовательностей для различных устройств и вариантов съемки, представленных в пакете данных DLC-2021

Устройство	Разрешение	Кадров в секунду	Вариант съемки				Всего
			Оригинал	Цвет. копия	Полутон. копия	Фото экрана	
Samsung S10	3840 × 2160	30	140	283	121	200	744
iPhone XR	3840 × 2160	60	70	201	51	200	522
Samsung S10	1920 × 1080	30	40	—	39	—	79
iPhone XR	1920 × 1080	30	40	—	39	—	79
Всего			290	484	250	400	1424

распознавания документов, либо в качестве основных экспериментальных наборов данных, либо в комбинации с закрытыми пакетами или с другими, также публично доступными. Спектр задач, решаемых исследователями, включал задачи детектирования и точной локализации документа на изображении, идентификацию типа изображения, распознавание текстовых реквизитов, точный поиск лица держателя документа, межкадровое комбинирование результатов распознавания, и другие.

4.5.1 Поиск и ректификация документов

В работах [308—310] описан алгоритм поиска четырехугольника документа на фотографии или кадре видеопоследовательности, использующий анализ контуров объектов, присутствующих на изображении и ранжирование гипотез об истинном четырехугольнике исходя из известных заранее возможных соотношениях сторон искомого документа. В этих работах эксперимент проводился с использованием MIDV-500 в качестве основного пакета данных, в котором содержатся документы с тремя различными соотношениями сторон, наиболее присущими идентификационным документам. В работах [201; 311] пакеты данных MIDV-500 используются для анализа другого типа алгоритмов детектирования документов на изображении, уже зависящих непосредственно от графического представления документа. В этих работах применяется подход с предварительным выделением и дескрибированием ключевых точек на изображении, и поиском оптимального геометрического преобразования, кото-

рое бы переводило идеальный шаблон документа в область на изображении. В работе [311] рассматриваются вопросы построения ограничений в оптимизационной задаче поиска этого графического представления, а в работе [201] рассматривается система эффективного детектирования и локализации жестко структурированных документов (на примере документов, удостоверяющих личность), пригодная как для работы на фотографиях и кадрах видеопотока, так и на сканированных изображениях. Поскольку в оригинальном наборе данных MIDV-500 не содержалось фотографий и сканированных изображений, в работе [201] также публикуется соответствующее расширение, базирующееся на тех же исходных шаблонов документов и с форматом разметки, аналогичной оригинальному. В работе [312] исследуется полный набор потенциальных признаков (включая ключевые точки, исчезающие точки, машиночитаемую зону, границы, углы, лицо держателя, логотипы, статические тексты и т. п.), которые могут использоваться для точной локализации и идентификации документа, удостоверяющего личность, на изображении. Авторы предлагают комбинированный метод, выбирающий из множества гипотез, построенных по всевозможным признакам, и используют пакет MIDV-500 как один из двух контрольных пакетов данных.

В работе [313] предлагается метод точной локализации и идентификации документа, основанный на предварительной грубой локализации при помощи поисковой нейронной сети, и последующего уточнения на основе контурных методов. В качестве одного из контрольных пакетов данных используется MIDV-500, однако в работе отмечается, что хотя пакеты данных MIDV-500 и MIDV-2019 достаточно хорошо описывают процесс мобильной съемки документа в процедурах удаленной идентификации личности, вариативность данных, представленная в них, недостаточна для обучения нейросетевых моделей (стоит отметить, что эта работа была проделана и опубликована до публикации пакета данных MIDV-2020, содержащего множество синтезированных примеров документов). Подобный недостаток также отмечается в работах по нейросетевому детектированию текста на произвольных изображениях [314]. Тем не менее, пакет данных MIDV-500 все же использовался для обучения и тестирования нейросетевых моделей поиска точек схода с использованием слоев нейронных сетей на основе преобразования Хафа в работах [315; 316] и для обучения нейросетевых моделей для поиска документа на изображении, выражающих задачу как задачу семантической сегментации [94] – в задачах, для которых вариатив-

ность текстовых данных и лица держателя документа не является критичным фактором. Метод ректификации изображения документа, использующий предварительный поиск точек схода также рассматривается в работе [317]. В работе [318] рассматривалась задача поиска произвольного документа на изображении (не обязательно документа, удостоверяющего личность) и пакет данных MIDV-500 использовался в качестве одного из пяти открытых пакетов данных, использующихся для обучения глубокой нейросетевой модели DeepLab (в качестве тестовой выборки использовался отдельный, также публично доступный пакет данных [319], содержащий изображения документов произвольной природы).

4.5.2 Поиск отдельных объектов и геометрических примитивов

Помимо задач детектирования и локализации документов также ведутся исследования в области более частных задач, связанных с устойчивым и эффективным детектированием геометрических примитивов, таких как отдельные линии или отрезки [320], генерацией устойчивых описаний локальных особенностей изображений [321; 322], и использования множественных признаков [323] для локализации и идентификации документов. Разницы в постановках задач, которые ставят авторы таких исследований, приводят к различным требованиям к пакетам данных и к предоставляемой разметке. Так, в работе [322] авторы конструируют устойчивый бинарный дескриптор локальной области изображения, для использования в методах детектирования и точного поиска документов, удостоверяющих личность. Поскольку целевой задачей является именно поиск документов, удостоверяющих личность, пакеты данных MIDV-500 и MIDV-2019 были использованы как основные пакеты как для обучения, так и для тестирования, и авторы показывают значительное преимущество построенного дескриптора перед более «универсальными» аналогами, применительно к целевой задаче. Напротив, авторы работы [320], исследующие универсальные методы устойчивого детектирования прямых линий на произвольных изображениях, отмечают, что в рамках широкой постановки задачи пакеты данных MIDV-500 не могут быть использованы, поскольку не обладают достаточно подробной геометрической разметкой, и присутствующие в размет-

ке прямые линии относятся только к границам документов. Это приводит к тому, что авторы вынуждены проводить экспериментальные исследования на полностью синтетических пакетах данных, моделируя изображения с заранее известным положением искомых объектов.

Различия в постановках задач меняет требования не только к глубине разметки, предоставляемой вместе с открытыми пакетами данных, но и к специфике определений размечаемых объектов. Так, в работе [324] пакет данных MIDV-500 был использован в качестве основного экспериментального набора данных для решения задачи поиска лиц на изображениях документов, удостоверяющих личность. Разметка лица владельца, присутствующая в пакетах данных MIDV-500 и MIDV-2019 представляла собой четырехугольник поля документа, в который вклеена или пропечатана фотография владельца, что имеет смысл с точки зрения построения систем анализа документов, предполагающих предварительное детектирование и ректификацию страниц документа [325], однако не является типичным среди исследователей, занимающихся детектированием лиц на произвольных изображениях – более типичным является разметка окаймляющего прямоугольника овала лица человека со сторонами, параллельными сторонам изображения. В связи с этим авторы работы [324] не могли использовать консистетные метрики качества детектирования лиц и были вынуждены готовить специальную разметку необходимых окаймляющих прямоугольников на первом и последнем кадре каждой видеопоследовательности пакета данных MIDV-500 с последующей интерполяцией. Высокая заинтересованность соавторов работы [324] в появлении объемных пакетов данных для объективного сравнения методов детектирования лиц на документах также мотивировало их принять участие в работе над пакетом данных MIDV-2020, в котором для каждого из 72 409 изображений документов представлена разметка как фотографии владельца как поля документа, так и окаймляющего прямоугольника овала лица (см. рис. 4.20).

4.5.3 Распознавание текстовых реквизитов на изображениях и на видеопоследовательности

Одна из наиболее важных задач анализа документов является точное распознавание его текстовых реквизитов. В работе [236] пакет данных MIDV-500 используется в качестве основного тестового набора для построения и анализа метода распознавания текстовых полей идентификационных документов на изображениях, полученных с камеры мобильных устройств (с обучением на искусственно синтезированных данных). В работе [326] предлагается имплементация квантизованных 4-битных нейронных сетей для распознавания символов, обучаемая на синтезированных данных и тестируемая на текстовых полях пакета данных MIDV-500. Вопросы генерации синтетических наборов обучающих данных также находятся в центре внимания работы [287], где пакеты данных MIDV-500, MIDV-2019 и MIDV-LAIT используются в качестве обучающей выборки для построения модели генерации правдоподобных последовательностей искусственных обучающих изображений для распознавания машиночитаемых зон идентификационных документов.

В работе [327] рассматривается задача детектирования и распознавания машиночитаемой зоны документов, удостоверяющих личность. Авторы предлагают двухэтапный нейросетевой подход MRZNet и проводят сравнительное исследование семи различных методов (включая авторский) с использованием трех пакетов данных – собственный пакет данных, не являющийся публично доступным, синтетический пакет данных SyntheticMRZ [237] и пакет MIDV-500, который содержит изображения международных паспортов и идентификационных карт, на которых присутствуют машиночитаемые зоны. Примечательно, что точность детектирования машиночитаемых зон на закрытом пакете данных, отмечаемая авторами, для собственного метода составила 100% (при точности ближайшего сравниваемого метода 74.15%), тогда как на пакете данных MIDV-500 точность авторского метода составила только 73.94%, и показала лишь третье место среди сравниваемых методов. Авторы объясняют такое несоответствие присутствием проективных и нелинейных искажений и бликов на изображениях пакета MIDV-500.

В работе [71] исследуется влияние предварительной бинаризации изображений документов на качество распознавания текстовых реквизитов, с

использованием пакета данных MIDV-500. Авторами показано, что современные широко используемые методы распознавания текстовых строк показывают более высокую точность распознавания при обработке исходных изображений, нежели чем при обработке предварительно бинаризованных изображений, даже при использовании лидирующих методов бинаризации, однако при использовании идеальной бинаризации (при которой для каждого конкретного изображения порог бинаризации в каждом локальном окне подбирается человеком) точность методов распознавания все же возрастает. Для проведения такого исследования авторам пришлось подготовить специальную трудоемкую разметку идеальной бинаризации документов для исходных изображений-шаблонов, кадры видеопоследовательностей пакета данных MIDV-500 в работе не рассматривались.

Поскольку целевой задачей при подготовке пакетов данных семейства MIDV являлось распознавание документов, удостоверяющих личность, в видеопоследовательности, эти пакеты данных стали основными для ряда исследований по повышению точности распознавания текстовых реквизитов путем межкадрового комбинирования результатов распознавания [328; 329], анализа влияния на качество распознавания текстовых полей искажений отдельных кадров, таких, как дефокусировка [330] или поворот в трехмерном пространстве [331] и построения решающего правила, позволяющего автоматически останавливать процесс распознавания объектов в видеопотоке [332—335].

4.5.4 Анализ качества изображения и поиск компрометирующих признаков

В автоматических системах анализа и распознавания документов, удостоверяющих личность, помимо непосредственно распознавания текстовых полей и извлечения значимой графической информации (такой, как фотография лица держателя документа), важной задачей является детальный анализ качества изображения документа и выявление признаков, свидетельствующих о том, что изображение или сам документ могли быть подделаны.

В работе [336] рассматривается задача определения качества фотографии лица владельца, извлекаемой из изображения документа, и ее пригодности для

дальнейшей обработки (к примеру, для автоматической сверки с фотографией лица, полученной из другого источника). Качество фотографии в рамках этого исследования определяется экспертом, что требует дополнительной разметки изображений пакета данных. В работе [337] рассматривается задача определения качества целикового изображения документа. В этой работе используется пакет данных MIDV-2020 – исходные изображения шаблонов документов рассматриваются как «эталонные» изображения с максимальным качеством, что позволяет анализировать артефакты их сканирования и фотографирования.

В работах [338—340] рассматривается задача автоматического детектирования чувствительной информации (к примеру, персональных данных) в потоках изображений – задача информационной безопасности, требующая в том числе применение современных методов компьютерного зрения. В этих работах пакет данных MIDV-500 использовался в качестве положительной выборки – набора изображений, на которых заведомо присутствуют чувствительные данные, которые необходимо автоматически детектировать. Безусловное наличие на изображениях пакета данных MIDV-500 информации, структурированной аналогично персональным данным, также использовалось в работе [341] для построения интеллектуальных моделей лингвистического анализа платежной информации.

Пакет данных MIDV-500 также использовался в качестве позитивной выборке в работе [342] для построения автоматического метода детектирования нарушений в текстуре фона документов, удостоверяющих личность, для выявления подозрительных манипуляций над изображениями. Отрицательная выборка (т. е. выборка изображений, над которыми проводились манипуляции) генерировалась автоматически, используя изображения исходных шаблонов документов из MIDV-500. Напротив, в работе [343] пакет MIDV-500 используется в качестве отрицательной выборки – в данной работе авторы предлагают метод автоматической верификации присутствия голографических элементов защиты на идентификационном документе. В качестве положительной выборки авторы использовали закрытый пакет данных действительных документов, а в качестве «подделок» – пакет MIDV-500, в котором не содержится никаких элементов защиты бланка.

Исследование методов автоматического детектирования компрометирующих признаков является крайне важным направлением для повышения безопасности процесса удаленной идентификации личности, и для проведения

таких исследований необходимо наращивать базу пакетов данных самой различной природы. Как отмечают авторы работы [344], пакетов данных MIDV-500, MIDV-2019 и MIDV-2020 недостаточно для комплексного анализа систем верификации действительности идентификационных документов, поскольку в них не представлены примеры распространенных атак, таких как повторная съемка документа, отображаемого на экране. Именно по этой причине были подготовлены такие пакеты данных, как MIDV-Holo (секция 4.4.5) и DLC-2021 (секция 4.4.6), которые включают в себя моделирование различных видов атак для объективной оценки алгоритмов, их детектирующих.

4.6 Методология создания открытых пакетов данных документов, удостоверяющих личность

Для поддержания современных темпов развития методов и подходов к задачам компьютерного зрения, в особенности учитывая тренд использования глубоких нейросетевых моделей, требует создания больших объемов данных для обучения и тестирования. Применительно к конкретным задачам, таким как анализ и распознавание документов, удостоверяющих личность, это создает трудности – либо связанные с особенностями целевого объекта (такие, как соблюдение юридических и этических норм, связанных с сохранностью чувствительных данных), либо связанные с ограниченным размером сообщества, занимающимся тем или иным рядом задач. Однако, как показывает опыт использования пакетов данных семейства MIDV, даже в случае, если основной экспериментальный пакет данных не публикуется, ограниченный открытый пакет данных, такой как MIDV-500, может использоваться в качестве дополнительного набора данных для публикации объективной оценки качества того или иного метода, пригодного для сравнения с другими.

Авторы ряда открытых пакетов данных заранее разделяют публикуемые выборки на обучающие, валидационные и тестовы, однако, как можно заметить из практики использования пакетов семейства MIDV, такое разделение может не иметь смысла – разнообразие подходов к решению тех или иных задач настолько велико, что исследователи вынуждены самостоятельно выбирать протоколы разделения пакетов данных для обучения и проверки, а иногда, для

решения более высокоуровневых задач, используют комбинацию нескольких пакетов данных из разных доменов (как, к примеру, в работе [318]).

Гораздо более важным аспектом построения открытых пакетов данных является структурирование пакетов таким образом, чтобы исследователи могли выделять строго определенным образом подмножества, объединяемые теми или иными признаками. К примеру, авторы проанализированных работ раздельно анализировали подмножества пакетов данных по освещенности и степени выраженности проективных искажений [328], по количеству видимых в кадре углов документа [308; 310], по наличию и целостности определенных видов текстовых полей или машиночитаемых зон [327; 332] и т. п. Поскольку одной из целей публикации открытых пакетов данных является предоставление возможности различным исследовательским группам сравнивать разрабатываемые методы и подходы на общей экспериментальной базе, необходимо предоставить возможность в явном виде указывать те подмножества пакета данных, которые используются в том или ином эксперименте.

Наиболее важным вопросом в построении и развитии открытых пакетов данных является вопрос их расширения либо новыми данными (с целью увеличения общего объема, либо с целью добавления данных, обладающих новыми, не представленными ранее признаками), либо добавлением специфической, проблемно-ориентированной разметки (к примеру, окаймляющих прямоугольников овалов лиц [324], идеальной бинаризации изображений документов [71], экспертной разметкой качества изображений или их фрагментов [336] и др.), либо производных сгенерированных изображений (таких, как изображения с нарушениями текстур фона документа [342]). Зачастую такие расширения используются авторами работ для конкретных исследований, но не выкладываются в публичное пространство для повторного воспроизведения результатов или для сравнительного анализа. В связи с этим важным аспектом открытых пакетов данных является предоставление инструментария для пользовательского расширения этих пакетов данных новыми изображениями, либо новой разметкой так, чтобы другие исследовательские группы могли выбирать не только интересующие их подмножества опубликованных данных, но и специфические производные изображения или уже подготовленную разметку, тем самым уменьшая несоответствия при сравнении методов и упрощая методологию исследований [345].

Исходя из опыта создания открытых пакетов данных для оценки систем анализа и распознавания документов, удостоверяющих личность, а также на основе анализа их использования в научном сообществе, можно сформулировать основные этапы и методологические принципы создания подобных пакетов данных:

1. В качестве исходных изображений документов, удостоверяющих личность, следует использовать изображения примеров документов, доступные в открытых хранилищах, таких как Wikimedia Commons. Большая часть изображений примеров идентификационных документов в таких хранилищах могут быть использованы для любых целей, поскольку согласно локальному законодательству изображения, являющиеся частью нормативных документов, не могут быть защищены авторским правом. Следует также отметить, что на значительной части подобных изображений на документе пропечатано слово «ОБРАЗЕЦ» на том или ином языке, что может создавать трудности при распознавании документа и не соответствует целевому сценарию использования системы распознавания документов, следовательно, такие элементы следует ретушировать. В отдельных случаях имеет смысл создавать полностью искусственные шаблоны идентификационных документов, не привязанных ни к какой конкретной стране или органу выдачи, как было сделано для пакета данных MIDV-Holo (описанного в секции 4.4.5), пользуясь при этом международными стандартами и рекомендациями по структуре таких документов.
2. Для генерации случайных значений текстовых реквизитов можно использовать находящиеся в открытом доступе автоматические генераторы случайных имен на различных языках. Для генерации случайных изображений лица держателя документа можно использовать существующие сервисы генерации искусственных лиц, такие как Generated Photos [301], основанные на генеративных состязательных нейронных сетях.
3. Подготовленные шаблоны документов в искусственными данными следует публиковать как часть пакета данных, поскольку для таких исходных изображений существует возможность предоставить полную разметку значений реквизитов и их геометрического расположения относительно шаблона документа. Разметку следует производить из-

вестными и апробированными инструментами, формат хранения данных которых уже известны научному сообществу. Примером такого инструмента разметки является VGG Image Annotator [303]. Также, при публикации исходных шаблонов у научного сообщества появляется возможность дополнять пакеты данных новыми фотографиями или видеопоследовательностями, снятыми в том числе в условиях, не представленных в оригинальном пакете данных, либо с моделированием других типов атак.

4. Поскольку большинство документов, удостоверяющих личность, выполняются либо с гладким пластиком в качестве носителя, либо их основные страницы защищаются полимерными пленками, при съемке документов камерой мобильных устройств на поверхности документов часто возникают блики. Для имитации такого эффекта распечатанные примеры документов перед съемкой необходимо ламинировать.
5. При видеосъемке и фотографировании документа следует классифицировать внешние условия съемки и вносить информацию об условиях съемки в структуру пакета данных, либо в его описание. Важными параметрами съемки является устройство, на которое производится съемка или фотографирование, условия освещения, особенности сценария съемки (к примеру, избегание бликов, допустимые диапазоны проективных искажений и т.п.), характеристики фона и др.
6. На подготовленных сканах или фотографиях, как правило, достаточно разметить точное геометрическое положение (координаты углов четырехугольника) документа и идентификатор типа шаблона. Вместе с разметкой содержимого шаблона этой информации достаточно для того, чтобы получить точное эталонное местоположение каждого объекта документа.
7. Подготовленные видеопоследовательности с целью разметки должны быть раскадрованы с известным и указанным в описании пакета данных количеством кадров в секунду. На каждом кадре, также как в случае фотографий, необходимо разметить точные координаты углов четырехугольника положения документа и идентификатор типа шаблона. Важно, что исходные видеофайлы следует также предоставлять как часть пакета данных, так, чтобы исследователи, которым требуется глубокое моделирование процесса съемки документа, могли использовать

не только размеченные видеокадры, но и полный набор всех полученных исходных данных.

8. При создании пакетов данных, предназначенных для оценки методов обнаружения атак на предъявление документа, либо проверки подлинности документа или его составных частей, необходимо описать варианты возможных атак, для каждой из них провести соответствующее моделирование, и представить в пакете данных фотографии или видеопоследовательности, соответствующие атакам, в таких же условиях съемки как и фотографии или видеопоследовательности, соответствующие «оригинальным» документам.

4.7 Выводы по главе

Объективное количественное оценивание работы является одной из наиболее важных задач при разработке системы распознавания идентификационных документов. В рамках диссертационной работы представлена методология оценки качества работы систем распознавания идентификационных документов, с подробным описанием критериев, показателей и методом определения соответствующих показателей.

В качестве критериев рассматривается оценка точности и скорости работы распознающей системы. При этом, при определении способа оценки точности распознавания детальный упор сделан на способе оценки локализации документа на изображении, точности определения типа документа, качестве распознавания текстовых полей и точности выделения графических полей.

Для возможности определения соответствующих показателей в настоящей главе представлена концепция определения таких показателей с использованием «верификационных» пакетов данных, изложены основные принципы построения таких пакетов данных. Кроме того, в настоящей главе представлены уникальные пакеты данных, созданные и опубликованные в открытом доступе в рамках настоящей диссертационной работы, полностью пригодные для оценки качества работы систем распознавания идентификационных документов.

Анализ их использования показывает общие тренды использования проблемно-ориентированных пакетов данных, включая различающиеся среди кол-

лективов требования к признакам, которыми должны обладать изображения и идеальная разметка, и были выявлены принципы, использование которых при дальнейшем создании и расширении открытых пакетов данных позволит использовать их более эффективным образом и повышать качество публикуемых исследований. На основе этих принципов была представлена методология создания открытых пакетов данных изображений документов, удостоверяющих личность, позволяющая расширить существующий корпус имеющихся в открытом доступе данных для исследований в области систем распознавания таких документов и для создания отдельных методов и алгоритмов.

Работа над открытыми пакетами данных важна не только для того, чтобы давать возможность исследователям проводить объективное сравнение и анализ методов решения тех или иных задач, но и для создания и развития новых научных связей между международными исследовательскими группами, объединенных интересом к общему кругу научных проблем.

Глава 5. Реализация системы распознавания

5.1 Введение

В настоящей главе приведено описание архитектуры промышленной системы распознавания удостоверяющих документов Smart ID Engine, реализующей высокоточное и безопасное программное обеспечение для распознавания данных более 2000 типов удостоверяющих документов 210 юрисдикций мира, и проведены замеры производительности на машинах семейства Эльбрус. В заключение приведены данные о внедрении этих программ в реальные государственные, промышленные и финансовые системы.

5.2 Общая архитектура системы

Архитектура мобильного распознавания в промышленной системе ввода идентификационных документов Smart ID Engine представлена на рисунке 1. Компоненты системы можно разделить на три категории:

- компоненты, которые обрабатывают входные изображения или видеокадры, выполняя поиск всех шаблонов и определение их координат (выделены зеленым на рисунке 5.1);
- компоненты, которые обрабатывают каждый отдельный шаблон документа (выделены желтым на рисунке 5.1);
- компоненты, которые комбинируют результаты распознавания шаблонов в логические представления документов, выполняют постобработку и выводят результат распознавания (выделены синим на рисунке 5.1).

Система Smart ID Engine предназначена для распознавания удостоверений личности на основе заранее определенного множества типов документов. Общие параметры системы могут быть разделены на три блока:

- база данных известных шаблонов документов с индексом, который используется при обнаружении и локализации шаблонов (блок F5 на рисунке 5.1);

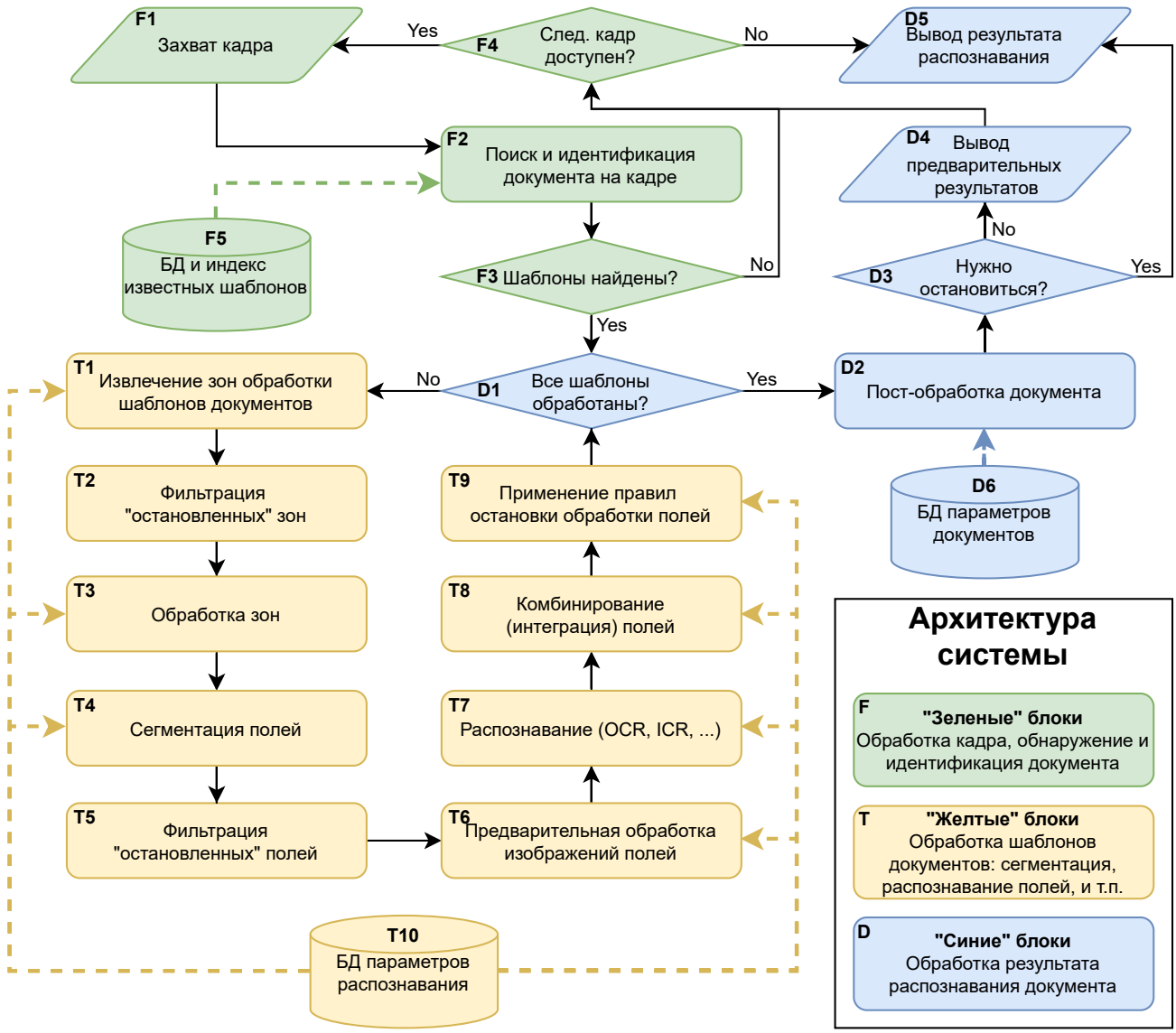


Рисунок 5.1 — Архитектура мобильного распознавания в промышленной системе ввода идентификационных документов Smart ID Engine.

– база данных параметров распознавания для каждого шаблона – информация о шаблоне, его полях и их свойствах, и других данных, необходимых для извлечения и распознавания компонентов шаблона (блок T10 на рисунке 5.1);
– база данных документов, которая содержит информацию о том, как результаты распознавания отдельных шаблонов комбинируются и обрабатываются для получения окончательного результата распознавания документа (блок D6 на рисунке 5.1).

5.3 Подсистема локализации и идентификации документа

На вход системы Smart ID Engine подается либо одно изображение (фотография или скан-копия), либо серия изображений (видеокадры). Получение каждого входного изображения осуществляется в блоке F1 на рисунке 5.1. Первым шагом обработки полученного изображения в рамках предлагаемой схемы является обнаружение и локализация шаблонов документов (блок F2) с учетом индекса известных шаблонов (блок F5). В системе Smart ID Engine использовались методы, локализирующие шаблоны по общему визуальному представлению, поскольку основной задачей Smart ID Engine является распознавание документов с фиксированной структурой. В частности, методы, основанные на алгоритме Виолы и Джонса, обобщенном в виде дерева решений сильных классификаторов, который может быть применен для обнаружения страниц документа и определения области, устойчивой к умеренным перспективным искажениям, а также методы, основанные на обнаружении границ документа, или на сегментации для выделения документа на фоне с помощью глубокого обучения. Также применяется универсальный метод, который использует идентификацию шаблона и сопоставление с помощью обнаружения особых точек и дескрипторов с помощью RANSAC.

Результатом блока F2 является совокупность обнаруженных на изображении шаблонов, каждый из которых имеет собственный индекс (что позволяет определить параметры для дальнейшей обработки) и геометрические параметры, то есть координаты границ шаблонов. Если нет соответствия ни одному из известных шаблонов (блок F3), то изображение дальше не обрабатывается системой Smart ID Engine, то есть либо процесс завершается без результата (если системе доступно единственное изображение), либо система начинает работу над следующим кадром (блок F4).

Если последовательность видеокадров подается на вход системы, и каждый кадр содержит шаблоны одного и того же документа, процесс определения местоположения и идентификации документа F2 может быть более успешным благодаря результатам обработки предыдущих кадров. Поэтому процесс F2 имеет доступ к временному хранилищу данных, которое накапливает результаты обработки нескольких кадров одного цикла распознавания документа.

5.4 Подсистема обработки шаблона документа

После того как все шаблоны, различимые на входном изображении, найдены и идентифицированы, каждый из них обрабатывается в соответствии с заранее определенным процессом, параметры которого хранятся в настройках распознавания шаблонов (блок T10).

Поскольку шаблон идентифицирован и его геометрическое положение на изображении определено, положение объектов может быть вычислено, а образы большинства отдельных элементов могут быть извлечены с поправкой на угловой поворот и проективные искажения. Более того, если известен физический размер страницы документа, соответствующей обрабатываемому шаблону, образ каждого отдельного объекта может быть сгенерирован с заданным пространственным разрешением (например, с фиксированным числом пикселей на дюйм). Однако фактическое разрешение с точки зрения количества информации, хранящейся в каждом пикселе, зависит от разрешения исходного изображения. Тем не менее, точные координаты шаблона (или, по крайней мере, гипотеза) не всегда означают, что известны точные координаты каждого объекта для распознавания и анализа. Хотя положение статических элементов шаблона фиксировано, отдельные объекты, такие как текстовые поля, могут иметь разную длину и даже менять положение. Например, рассмотрим третью (главную) страницу паспорта гражданина РФ (см. рисунок 5.2). Первые три поля сверху - Фамилия, Имя, Отчество. В шаблоне есть соответствующие статические метки и подчеркивания, которые указывают, где должны располагаться эти поля. Однако горизонтальное положение каждого поля может отличаться в разных экземплярах одного типа документа. Более того, из-за дефектов печати эти поля могут быть смещены вдоль вертикальных осей (возможно в том числе пересечение со статическими линиями на фоне), а также поля могут иметь небольшой наклон.

Это приводит к необходимости введения промежуточной субструктуры, представляющей локализованную область шаблона, которая должна использоваться для поиска отдельных объектов. В системе Smart ID Engine эти субструктуры называются "зонами"(показаны синими прямоугольниками на рисунке 5.3).

Рисунок 5.2 — Пример третьей (главной) страницы паспорта гражданина РФ.

Рисунок 5.3 — Пример паспорта Уругвая и иллюстрация зон распознавания.

Зона – это область шаблона с заранее определенными координатами, которая может быть обработана единичным применением некоторого заранее определенного алгоритма, который сегментирует зону и извлекает отдельные объекты, такие как текстовые поля или какие-либо другие объекты, представ-

ляющие интерес. Зона шаблона может включать в себя один объект (например, одно текстовое поле), несколько полей или объектов, и даже соответствовать всему шаблону документа – в зависимости от сложности шаблона, характеристик извлекаемых объектов и специфики алгоритмов, используемых для извлечения объектов.

Таким образом, первым шагом обработки шаблона в системе Smart ID Engine является извлечение образов отдельных зон (блок T1), в соответствии с информацией о зонах и их положениях, закодированной в параметрах обработки шаблона (блок T10). Каждая зона, указанная в параметрах, определяет набор отдельных объектов (например, текстовых полей), которые могут быть извлечены из данной зоны. При использовании видеопотока результаты распознавания или извлечения отдельных объектов обновляются после каждого обработанного кадра. Важно автоматически определить, когда этот процесс должен быть остановлен для каждого объекта. Если процесс распознавания объекта завершен, то нет необходимости искать его на следующем кадре, так как это сэкономит время обработки. Аналогично, если все отдельные объекты, соответствующие определенной зоне, уже обнаружены, сама зона может быть пропущена, что сэкономит время, затрачиваемое на ее анализ. Таким образом, после определения множества зон рассматриваемого шаблона документа и извлечения зон в процессе T1, зоны, поля (объекты) которых уже извлечены, отфильтровываются (блок T2).

Следующим шагом является обработка образа зоны (блок T3), например обнаружение и коррекция углового отклонения, обнаружение специфических элементов защиты, подавление текстуры фона и многое другое [19]. Хотя такая обработка может быть выполнена на уровне отдельных объектов, часто надежность результатов операций повышается, если доступен контекст всей зоны. Хорошим примером может служить изображение, текстовые поля которого имеют одинаковый угловой наклон (из-за дефекта печати) – хотя угол может быть определен на уровне каждого поля, для получения более надежного и согласованного результата полезно проанализировать зону целиком. После обработки зона сегментируется на отдельные объекты (блок T4). Метод сегментации может варьироваться от зоны к зоне, в зависимости от ее структуры: некоторые зоны могут иметь фиксированные локальные координаты каждого отдельного объекта или поля, заданные в соответствующих параметрах, и, таким образом, дополнительный поиск не требуется. В других зонах может потребоваться поиск

точных координат полей относительно заранее определенной структуры, или даже применение распознавания текста в свободной форме. Система Smart ID Engine поддерживает разные способы сегментации зоны на отдельные объекты.

Результатом процесса T4 является множество отдельных идентифицированных объектов (таких как текстовые поля, штампы, локализованные оптически переменные элементы, графические зоны документа, например, подпись или фотография, и т. д.) с известными координатами внутри зоны (а также в шаблоне документа и в исходном изображении). Аналогично фильтрации зон в процессе T2, некоторые из извлеченных полей в текущем цикле распознавания видеопотока могут быть уже детектированы. Таким образом, объекты, которые удовлетворяют соответствующим условиям, отфильтровываются (блок T5). Оба процесса (обработка изображений зон T3 и сегментация объектов T4) могут иметь доступ к временному хранилищу текущего цикла распознавания, чтобы использовать информацию, полученную на предыдущих этапах обработки, для повышения надежности анализа зон.

Последующие шаги обработки шаблона связаны с анализом отдельных объектов. Сначала изображения отдельных объектов подвергаются предварительной обработке (блок T6) – исходя из той же мотивации, что и при предобработке изображения зоны (блок T3). Блок T6 может включать ректификацию специфических особенностей объектов, например сдвиг текстовой строки. Затем каждый отдельный объект подлежит распознаванию (блок T7), если этого требует специфика данного объекта. Отметим, что свойства поля, подлежащего распознаванию, в системе Smart ID Engine известны заранее, и сохраненная информация о языке, особенностях используемого шрифта или других характеристиках визуального представления могут быть использованы для повышения надежности и эффективности распознавания.

Если в качестве входных данных для системы используется видеопоток, результаты распознавания комбинируются в общий результат распознавания (блок T8), а также применяются условия для остановки процесса распознавания или для определения необходимости дополнительных наблюдений того же объекта (блок T9). Процесс интеграции T8 и условия остановки T9 используют временное хранилище цикла распознавания, поскольку данным блокам необходимо считывать и обновлять текущее состояние анализа видеопотока.

Результаты распознавания объектов, не являющихся текстом, например, личных подписей или фотографий, также могут быть комбинированы. Напри-

мер, в процессе T8 можно выбрать одно лучшее изображение путем анализа фокусировки, освещения или наличия бликов. Процесс T8 может быть использован для анализа оптически изменяемых элементов, например голографических элементов безопасности. Такой анализ позволяет проверить изменение этих элементов между кадрами и сравнить с тем, как должен вести себя действительный (неподдельный) объект.

5.5 Подсистема формирования результата распознавания документа

После обработки всех отдельных шаблонов (в конце цикла при условии D1), следующий этап в системе Smart ID Engine – это общий вывод о документе в целом. В параметрах документа (блок D6) хранится заранее известная информация о документе, например, тип, составляющие шаблоны и объекты распознавания, которые ожидаются в выводе.

Заключительным этапом анализа документов является этап постобработки (блок D2). Текстовые поля удостоверений личности обычно имеют определенную синтаксическую и семантическую структуру, которая известна заранее. Такая структура, как правило, включает в себя следующие компоненты:

1. Синтаксис: правила для структуры текстовых полей. Например, поле «дата рождения» машиночитаемой зоны загранпаспорта состоит из шести символов, каждый из которых может принимать одно из 11 возможных значений (десятичные цифры и символ заполнения).
2. Семантика поля: правила для семантической интерпретации текстового поля или его составляющих. Например, поле «дата рождения» машиночитаемой зоны загранпаспорта записывается в фиксированном формате «YYMMDD», где «YY» - две последние цифры года, «MM» - месяц с 01 по 12, «DD» – день (в соответствии с номером месяца), а неизвестные компоненты даты заменяются парами символов заполнения «<<».
3. Семантические отношения: правила для структурных или семантических отношений между различными полями одного и того же

документа. Например, в достоверных документах, дата в значении поля «дата выдачи» быть более ранней, чем дата в значении поля «дата рождения».

Если на вход системы подается видеопоток, то после этапа постобработки документа D2 должно быть принято решение о том, можно ли считать результат конечным (блок D3). На это решение влияют не только условия остановки распознавания для отдельных объектов (которые проверяются в блоке T9), но и наличие в сформированном результате для документа в целом ожидаемых шаблонов и соответствующих объектов. Если результат можно считать конечным, то он становится результатом процесса (блок D5), а процесс распознавания завершается. Если результат не считается конечным, то происходит получение следующего кадра (блок F1), и процесс продолжается. Промежуточный результат внутри цикла процесса также может быть возвращен пользователю для визуализации и внешнего контроля (блок D4).

5.6 Оценка быстродействия в задаче мобильного распознавания

Время, необходимое системе распознавания для полной обработки одного кадра видеопоследовательности на мобильном устройстве, включает в себя полный набор этапов, т.е. поиск документа на изображении, определение координат его страниц (шаблонов), определение типа документа, поиск текстовых полей и других значимых объектов, распознавание текстовых полей. Общее время настолько сложной последовательности операций неизбежно зависит от большого количества факторов, включая разрешение исходного изображения, количество значимых объектов и текстовых полей на документе, и особенности самого вычислительного устройства.

Для количественной оценки производительности системы на мобильных устройствах было замерено время обработки реализованной системы одного кадра видеопоследовательности, с использованием различных мобильных устройств и распечатанных примеров документов из пакета данных MIDV-500. Результаты замеров представлены в таблице 24.

Как можно заметить из таблицы 24, время обработки кадра значительно варьируется в зависимости от используемого устройства (в частности, можно

Таблица 24 — Среднее покадровое время (в секундах) обработки документа, удостоверяющего личность, на мобильном устройстве.

Мобильное устройство	Паспорт РФ (стр. 3)	Вод. удост. Германии	Паспорт Сербии	Паспорт Азербайджана	Вод. удост. Японии
Huawei Honor 8 (2016)	0.21	0.52	0.74	0.76	0.92
Xiaomi Pocophone F1 (2018)	0.12	0.28	0.49	0.42	0.61
Apple iPhone XS (2018)	0.08	0.22	0.31	0.37	0.39
Samsung Galaxy S10 (2019)	0.10	0.21	0.30	0.31	0.42
Apple iPhone SE 2 (2020)	0.07	0.18	0.20	0.34	0.41

заметить явную связь производительности системы с годом выпуска модели мобильного устройства, ввиду увеличения производительности мобильных центральных процессоров), так и в зависимости от распознаваемого документа. Минимальное время обработки достигается на 3-й странице паспорта гражданина РФ без машиночитаемой зоны, содержащей лишь 5 текстовых полей, выполненных кириллическим алфавитом и 2 поля из цифр и знаков препинания, максимальное время обработки требуется для распознавания водительского удостоверения Японии, на котором присутствует стандартная коллекция полей на японском языке (порядка 11 тысяч иероглифов).

5.7 Оценка быстродействия в задаче массового ввода документов

Как было отмечено ранее, повышение быстродействие систем распознавания документов имеет важное значение не только для исполнения на мобильных устройствах, но и для повышения эффективности систем массового ввода документов. Такие системы, как правило, строятся на основе высокопроизводительных серверов, также обладающих процессорами с различной архитектурой, как общего назначения, так и специального. Вопросы повышения производительности алгоритмов обработки изображения для мобильных вычислительных устройств были рассмотрены в предыдущих разделах, в этом же разделе также кратко рассмотрим вопросы повышения производительности систем рас-

познавания для серверных платформ, в частности, для платформ с широким командным словом.

Архитектуры с принципом широкого командного слова (Very Long Instruction Word, VLIW) является одной из специальных архитектур для построения высокопроизводительных вычислительных систем. Примером такой архитектуры является архитектура Эльбрус [346]. При генерации исполняемого кода для процессоров с VLIW-архитектурой компилятор формирует последовательности групп команд (широкие командные слова), в которых отсутствуют зависимости между командами внутри каждой группы и сведены к минимуму зависимости между командами в разных группах. При исполнении команды каждой группы запускаются параллельно, что обеспечивает высокий уровень параллелизма на уровне команд [347]. Такое распараллеливание на уровне команд целиком обеспечивается оптимизирующим компилятором, разработанном с учетом особенностей конкретной архитектуры. В случае с архитектурой Эльбрус, эту функцию выполняет оптимизирующий компилятор `lcc`.

Другой особенностью процессоров архитектуры Эльбрус являются методы работы с памятью. Помимо наличия кэша, позволяющего оптимизировать время доступа в память, ими поддерживаются методы предварительной подкачки данных, которые позволяют прогнозировать обращения в память и производить подкачку данных в кэш или другое специальное устройство за некоторое время до их использования. Процессоры архитектуры Эльбрус поддерживают программно-аппаратный метод подкачки. Это означает, что аппаратная часть микропроцессора включает в себя специальное устройство для обращения к массивам (Array Access Unit, AAU), в то время как необходимость подкачки определяется компилятором, генерирующим специальные инструкции для AAU. Использование устройства подкачки эффективнее помещения элементов массива в кэш, поскольку элементы массивов чаще всего обрабатываются последовательно и редко используются более одного раза [346]. Однако необходимо отметить, что использование буфера предварительной подкачки на Эльбрусе возможно только при работе с выровненными данными. За счет этого чтение/запись выровненных данных происходят заметно быстрее, чем соответствующие операции для невыровненных данных.Также микропроцессоры Эльбрус поддерживают несколько видов параллелизма помимо параллелизма на уровне команд: векторный параллелизм [348], параллелизм потоков управления, и параллелизм задач в многомашинном комплексе.

В целях оптимизации работы систем распознавания для архитектуры Эльбрус можно использовать две технологии: распараллеливание вычислений и реализация функций низкоуровневой обработки изображений с помощью специальных встроенных функций (по принципам, аналогичным изложенным в разделе 3.5).

Распараллеливание вычислений можно выполнять на максимально доступное число потоков при помощи открытых программных библиотек для реализации параллельных вычислений, таких как `tbb` (Intel Threading Building Blocks) [349]. Библиотека `tbb` предоставляет набор универсальных функций, таких как `parallel_for`, позволяющая распараллелить цикл с независимыми по данным итерациями, и `task_group`, позволяющая создать набор из нескольких независимых задач, которые затем могут исполняться параллельно.

В рамках системы распознавания документов, удостоверяющих личность, существует несколько этапов обработки документов, которые могут быть эффективно распараллелены: этап поиска документов, в случае если используется подход из семейства Виолы и Джонса [97], типизация и поиск документа при помощи сравнения документа с несколькими шаблонами с помощью расширенного алгоритма RANSAC [209], поиск текстовых полей и других графических элементов (что может выполняться независимо в различных частях найденного документа), а также сегментация и распознавание отдельных текстовых полей.

Для архитектуры Эльбрус также доступен специальный класс встроенных функций, вызовы которых заменяются компилятором на высокоэффективный код для данной платформы. С помощью встроенных функций разработчики могут использовать векторный параллелизм: выполнять одну и ту же операцию над одним регистром, содержащим сразу несколько элементов данных. Микропроцессоры Эльбрус-4С и Эльбрус-8С поддерживают набор встроенных функций, для которого размер регистра составляет 64 бита. Он включает в себя операции для преобразования данных, инициализации элементов вектора, арифметических операций, побитовых логических операций, перестановки элементов вектора и др. Кроме того, для процессоров Эльбрус доступна библиотека `EML`, содержащая эффективные реализации основных функций обработки данных.

При использовании встроенных функций на платформе Эльбрус следует уделять особое внимание доступу в память, поскольку в задачах обработки изображений часто требуется невыровненное чтение данных в 64-битном реги-

стре. Такое чтение само по себе неэффективно, так как требует пары команд чтения и последующей команды формирования блока данных, но, что еще важнее, при этом не может использоваться буфер подкачки массивов, повышающий скорость доступа к данным на платформе Эльбрус. Таким образом, низкоуровневую обработку данных, например числовых массивов, следует выполнять в несколько этапов: обработка начальной части (до границы выравнивания на 64-бита), обработка основной части, использующая выровненный доступ к памяти, и обработка оставшихся элементов массива. Поскольку анализ указателей во время компиляции является нетривиальной задачей, можно использовать флаг компилятора `-faligned`, с которым все операции доступа к памяти выполняются выровненным образом.

Следующая особенность использования встроенных функций на платформе Эльбрус связана непосредственно с VLIW-архитектурой. Благодаря наличию нескольких арифметико-логических устройств (АЛУ), которые работают параллельно и загружаются при формировании широких командных слов, несколько команд могут исполняться одновременно. Процессоры Эльбрус-4С и Эльбрус-8С содержат шесть АЛУ, которые можно задействовать в рамках одной широкой команды, однако каждое АЛУ поддерживает свой набор встроенных функций. Простые операции, например сложение или умножение элементов в 64-битных регистрах, как правило, поддерживаются двумя АЛУ. Это означает, что процессор Эльбрус может исполнить по две таких инструкции за один такт. Для этого в исполняемом коде следует использовать развертывание циклов. Компилятор `lcc` поддерживает программу `#pragma unroll(n)`, которая позволяет выполнить развертывание n итераций цикла.

Таким образом, для эффективного использования встроенных функций на платформе Эльбрус необходимо:

– обеспечить выровненный доступ в память;
– обеспечить использование всех доступных АЛУ путем развертывания циклов.

Для анализа потенциальной эффективности использования параллелизма и встроенных функций в рамках системы распознавания документов, оперирующей на вычислительном устройстве с архитектурой Эльбрус, с помощью функций библиотеки `EML` был реализован ряд операций, применяемых при низкоуровневой обработке изображений и распознавании образов:

– арифметические операции над отдельными пикселями изображения;

– масштабирование изображения;
– транспонирование изображения (см. раздел 3.5.1);
– поворот изображения;
– фильтрация изображения (в т.ч. морфологическая фильтрация, см. раздел 3.5.2);
– умножение и сложение матриц вещественных чисел.

После этого система распознавания документов была скомпилирована из исходного кода с помощью оптимизирующего компилятора `lcc` [350] версии `1.21.19`. Распараллеливание выполнялось на максимальное доступное число потоков при помощи библиотеки `tbb` [349]. Библиотека `tbb` является кроссплатформенной и в свою очередь может быть собрана компилятором `lcc` для процессоров семейства Эльбрус из исходного кода.

Далее распознающая система была запущена на пяти различных машинах с процессорами Эльбрус:

1. Эльбрус 101-PC;
2. Эльбрус 401-PC;
3. Эльбрус-4.4;
4. Эльбрус 801-PC;
5. Эльбрус-8.4.

Основные характеристики тестируемых устройств приведены в таблице 25.

Была проведена экспериментальная оценка рассмотренных методов повышения быстродействия, а именно, проведены замеры среднего времени распознавания одного изображения, содержащего произвольным образом повернутый документ «Паспорт РФ», до и после использования данных методов. Это время не включало время загрузки изображения и файлов конфигурации из памяти. Усреднение выполнялось по набору из 800 изображений. Результаты приведены в таблице 26. Можно видеть, что оптимизация позволила повысить быстродействие системы в 5,3 раза, причем распараллеливание дало ускорение в 2,4 раза, а использование встроенных функций – еще в 2,2 раза. Несмотря на то, что Эльбрус-401PC содержит 4 вычислительных ядра, не все части вычислительной системы были распараллелены, поэтому результирующее ускорение не достигло 4. Встроенные функции также существенно помогли повысить быстродействие, однако далеко не все операции были ускорены с их помощью. Тем не менее, было получено ускорение в 2,2 раза.

Таблица 25 — Характеристики тестируемых машин с архитектурой Эльбрус.

Машина	Эльбрус 101-PC	Эльбрус 401-PC	Эльбрус 4.4	Эльбрус 801-PC	Эльбрус 8.4
Процессор	Эльбрус-1C+	Эльбрус-4C	Эльбрус-4C	Эльбрус-8C	Эльбрус-8C
Ядра общего назначения	1	4	16	8	32
Тактовая частота, МГц	985	800	750	1200	1200
Операции за такт (на ядро)	до 25	до 23	до 23	до 25	до 25
Объем ОЗУ	16 ГБ	24 ГБ	96 ГБ	32 ГБ	128 ГБ
Ширина SIMD-регистра	64 бита	64 бита	64 бита	64 бита	64 бита
L1 кэш (на ядро)	64 КБ данные + 128 КБ команды				
L2 кэш (на ядро)	2 МБ	2 МБ	2 МБ	512 КБ	512 КБ
L3 кэш (общая)	—	—	—	512 КБ	512 КБ

Таблица 26 — Среднее время распознавания документа «Паспорт РФ» на Эльбрус 401-PC с различными методами повышения быстродействия.

Вид оптимизации	Среднее время распознавания, сек
Без оптимизации	10,01
С распараллеливанием без использования встроенных функций	4,22
С распараллеливанием и встроенными функциями	1,90

Для оценки производительности всей распознающей системы использовалось среднее время распознавания одного изображения, содержащего произвольным образом повернутый документ заранее известного типа. Это время не включало время загрузки изображения и файлов конфигурации из памяти. Для документа каждого типа усреднение выполнялось по набору данных из 800 изображений.

Были рассмотрены 6 различных типов документов:

– Паспорт РФ;

– Биометрический паспорт РФ;

– Водительское удостоверение (ВУ) РФ;
– Водительское удостоверение (ВУ) Великобритании;
– Идентификационная карта Германии;
– Листок нетрудоспособности.

В таблице 27 приведено среднее время распознавания документов каждого типа в режимах «клиентского» (распознавание одиночного изображения) и «серверного» (распознавание в рамках потока большого объема) распознавания. Время распознавания на одноядерной машине Эльбрус 101-PC составило 2,3 – 7,6 с для разных документов, поскольку они значительно различаются по объему распознаваемой информации. На остальных тестируемых устройствах время распознавания всех документов, кроме листка нетрудоспособности, не превышает 2 с в «клиентском» режиме и 1,8 с в «серверном» режиме. Можно видеть, что в «клиентском» режиме распознавание больше, чем в 4 потока, не дает значительного прироста производительности на всех документах, кроме листка нетрудоспособности. Этот результат связан со свойствами документов: в паспорте распознается 12 текстовых полей, а в водительском удостоверении и идентификационных картах – 7. В таких условиях распараллеленной оказывается не такая большая часть алгоритма распознавания. Листок нетрудоспособности содержит значительно больше полей, поэтому ускорение между 401-PC и Эльбрус 4.4 и 801-PC и Эльбрус 8.4 заметнее.

В «серверном» режиме запускается несколько независимых процессов распознавания документов. Каждый вызов распознавания распараллелен так же, как и в предыдущем эксперименте, однако здесь время обработки включало время загрузки изображения из файла. Такой режим обеспечивает полную загрузку ядер процессора и позволяет более реалистично оценить производительность соответствующих устройств.

Результаты экспериментов показали, что серверные модули на основе процессоров Эльбрус демонстрируют ускорение в 3–4 раза за счет параллелизма на уровне задач. При этом сервер Эльбрус-4.4 на 20–30% мощнее рабочей станции Эльбрус 801-PC. В свою очередь 801-PC практически в 3 раза быстрее своего предшественника 401-PC за счет повышения тактовой частоты и значительного усовершенствования архитектуры. Для Эльбрус-4.4 и Эльбрус 8.4 это соотношение сохранилось.

Таким образом, результаты проведенных экспериментальных исследований подтверждают, что использование высокой степени параллелизма за счет

Таблица 27 — Среднее время распознавания различных типов документов на аппаратных платформах с вычислительной архитектурой Эльбрус.

Документ	Эльбрус 101-PC	Эльбрус 401-PC	Эльбрус 4.4	Эльбрус 801-PC	Эльбрус 8.4
«Клиентский» режим – распознавание одиночного изображения					
Паспорт РФ	3,87 с	1,90 с	1,80 с	1,21 с	1,09 с
Биом. паспорт РФ	3,33 с	1,85 с	1,80 с	1,10 с	1,05 с
ВУ РФ	4,24 с	2,12 с	1,81 с	1,24 с	1,09 с
ВУ Великобритании	2,26 с	1,08 с	1,03 с	0,69 с	0,66 с
Идент. карта Германии	2,32 с	1,22 с	1,13 с	0,77 с	0,72 с
Листок нетрудоспособности	7,59 с	3,40 с	2,65 с	1,97 с	1,49 с
«Серверный» режим – пакетная обработка с максимальной нагрузкой					
Паспорт РФ	—	1,27 с	0,36 с	0,43 с	0,11 с
Биом. паспорт РФ	—	1,13 с	0,36 с	0,42 с	0,11 с
ВУ РФ	—	1,79 с	0,47 с	0,64 с	0,16 с
ВУ Великобритании	—	0,93 с	0,26 с	0,32 с	0,08 с
Идент. карта Германии	—	0,99 с	0,26 с	0,37 с	0,10 с
Листок нетрудоспособности	—	2,22 с	0,66 с	0,86 с	0,22 с

широких командных слов на устройствах с архитектурой Эльбрус, а также наличие SIMD-расширений, позволяет добиться снижения времени распознавания документов до 5 раз, что позволяет строить более эффективные системы массового ввода документов.

5.8 Опыт внедрения системы

На основе предложенной системы были разработаны ряд прикладных систем, внедренных в различных областях экономики и управления. Так, распознавание российского паспорта используется для процессов регистрации пользователей ведущих банков РФ – Тинькофф, Альфа-банк, Газпромбанк, Открытие, Райффайзен, Росбанк, ДомРФ, Точка банк, Совкомбанк, МТС банк, Хоум Кредит Банк, МКБ, и ряд региональных банков. Распознавание водительских удостоверений, свидетельств о регистрации транспортных средств,

паспортов транспортных средств и паспорта граждан РФ внедрены в ряде страховых компаний – Ингострах, Альфа-страхование, РЕСО-Гарантия, Согласие.

Комплекс распознавания документов всего мира, включающий паспорта 210 стран и организаций, внутренних идентификационных карт, видов на жительство и водительских удостоверений всех стран мира, внедрен в банке ЕАБР, сервисе iDenfy, travizory, Oman Arab Bank, Emirates NDB, Dukascopy Swiss Banking Group, Kaspi.kz и ряде других крупных организаций.

В области мобильной связи, операторы МТС, Билайн и Мегафон используют созданный комплекс для идентификации абонентов при продаже SIM-карт.

Транспортная отрасль активно использует программный и программно-аппаратный комплексы идентификации, созданный на основе разработанной системы. Так, РЖД использует его для продажи железнодорожных билетов в 850 кассах, международный конгломерат SITA, авиакомпании Turkish Airlines и Croatia Airlines используют мобильную систему распознавания для регистрации на рейс, а в аэропорту Шереметьево программно-аппаратный комплекс используется для автоматического пересечения границы. Крупнейший оператор круизных линий в мире RCCL использует его для продажи билетов и прохода на лайнеры.

Созданный комплекс используется системой изготовления и выдачи паспортно-визовых документов ГС МИР. Для нужд ФНС России программный комплекс используется в мобильном приложении регистрации самозанятых и ИП. Также программно-аппаратный комплекс используется для выдачи ЭЦП руководителям организаций.

Созданный совместно с китайской компанией Pixsur программно-аппаратный комплекс использовался для регистрации посетителей стадионов во время чемпионата мира по футболу в Катаре.

Общее число организаций, использующих решения, построенные на основе разработанного подхода, составляет более 200 по всему миру. Таким образом можно утверждать, что решения, предложенные в данной работе, прошли широкую апробацию в прикладных системах.

Заключение

Основные результаты работы заключаются в следующем.

1. Предложен подход к созданию систем распознавания документов, удостоверяющих личность, учитывающий и объединяющий особенности формирования изображений, получаемых с мобильных устройств, специальных сканеров документов и классических сканеров. Рассмотрено место и возможность использования особенностей видеопотока для повышения качественных и функциональных характеристик в неконтролируемых условиях съемки мобильным устройством (неизвестны: освещение, геометрия сцены, камера и оптическое устройство) и ограниченности вычислительных возможностей самого мобильного устройства. Предложена новая схема построения такого типа систем, на основе алгоритмов компьютерного зрения.
2. Созданы первые в своем роде пакеты данных, которые позволяют объективно оценивать различные аспекты систем распознавания документов, удостоверяющих личность – семейство пакетов данных MIDV (Mobile Identity Document Video): MIDV-500, MIDV-2019, MIDV-LAIT, MIDV-2020, DLC-2021, MIDV-Holo. Данное семейство состоит из видео и фото синтезированных и распечатанных документов, удостоверяющих личность, снятых в различных условиях. При этом впервые удалось не только создать репрезентативные пакеты данных, но и соблюсти все законодательные ограничения, что позволяет сделать результаты в данной области исследования публичными и проверяемыми. В созданных пакетах данных впервые представлены изображения, видеопоследовательности и соответствующая разметка, предназначенные для оценки алгоритмов проверки подлинности документов и обнаружения атак на предъявление документов. Для обучения распознаванию символов в задаче распознавания документов на фотографиях и видеопотоке предложена оригинальная система сбора и разметки обучающих выборок. Для задачи поиска печати предложена оригинальная система аугментации и показано, что предложенный подход позволяет получать высококачественные детекторы печатей.

3. Предложено и обосновано использование видеопотока как формы входных данных для системы распознавания документов, позволяющее повысить качество распознавания. Предложены и проанализированы модели с различными стратегиями комбинирования результатов распознавания. Исследованы методы на основе выбора лучшего изображения, выбор лучшего результата распознавания и ансамблирования результатов распознавания. Экспериментально показано, что в условиях съемки близких к идеальным наилучшим результатам комбинирования дает выбор лучшего кадра с помощью оценки фокуса, а метод ROVER дает лучший результат в условиях помех.
4. Построены новые вероятностные модели, описывающие результаты распознавания символов текстовых полей документов в видеопоследовательности. Введено понятие потока результатов распознавания. Рассмотренные модели предполагают, что результат распознавания знакоместа в поле документа можно представить в виде композиции случайных величин и случайных векторов. Проведены проверки, которые подтвердили адекватность вероятностных моделей. Полученные при моделировании потока результатов распознавания параметры можно использовать для комбинирования результатов классификации одиночных символов, решая тем самым задачу распознавания объекта в видеопотоке.
5. Рассмотрена проблема останова распознавания документа, возникающая при использовании видеопотока. Для ее решения предложены и проанализированы методы на основе анализа популяций и на основе моделирования следующего комбинированного результата распознавания. Экспериментально показано, что метод, основанный на моделировании следующего комбинированного результата, превосходит остальные, но требует дополнительных вычислений, связанных с расчетом следующего результата.
6. Рассмотрены проблемы производительности систем распознавания видеопотока и предложены новые методы эффективной реализации алгоритмов, активно применяющихся для современных процессоров, используемых в мобильных устройствах. Предложены алгоритмы, использующие инструкции ARM NEON для эффективного транспонирования матриц, повышающие скорость в 5,7–12 раз в зависимости от

размера матриц, и алгоритм эффективной реализации морфологической фильтрации изображений.

7. На основе предложенных подходов были разработаны прикладные программные средства для распознавания документов, удостоверяющих личность, Smart IDReader и Smart ID Engine. Программные средства позволяют проводить распознавание документов в видеопотоке, фотографии, изображении, полученным с использованием обычного или специального сканера, на мобильном устройстве и обычном компьютере в реальном времени. При распознавании видеопотока используются предложенные в алгоритмы комбинирования результатов распознавания и принятия решения об остановке. Проведена оценка производительности системы распознавания на мобильных вычислительных устройствах и процессорах с широким командным словом семейства Эльбрус.
8. Smart IDReader и Smart ID Engine активно применяются в мобильных и серверных приложениях ряда российских государственных и коммерческих организаций: ФНС, МВД, РЖД, Банк Тинькофф, Альфабанк, Банк Открытие, Газпромбанк, МТС, Beeline, МегаФон, что подтверждено актами о внедрении. Разработанные программные системы также являются частью паспортно-визовой системы ГС МИР. В рамках работы по диссертации было получено 2 патента США, 5 патентов на изобретение РФ, 16 патентов на полезные модели и 4 свидетельства о государственной регистрации программы для ЭВМ.

Благодарность. Эта диссертация вряд ли могла бы состояться вне работы над большим комплексом программ распознавания документов, удостоверяющих личность. В работе над ним принимало и принимает участие много моих коллег. Всем им я хочу выразить свою благодарность.

Основные публикации автора по теме диссертации

Основные результаты по теме диссертации изложены в 40 печатных работах, 20 из которых изданы в журналах, рекомендованных ВАК, 30 работ индексируется Web of Science и Scopus (включая 10 работ, опубликованных в журналах Q1 и Q2).

Список публикаций:

1. Arlazarov V. L., Arlazarov V. V., Bulatov K. B., Chernov T. S., Nikolaev D. P., Polevoy D. V., Sheshkus A. V., Skoryukina N. S., Slavin O. A., Usilin S. A. Mobile ID Document Recognition-Coarse-to-Fine Approach // Pattern Recognit. Image Anal.. — 2022. — Vol. 32. — No 1. — P. 89-108. — DOI: 10.1134/S1054661822010023. (ВАК, WoS, Scopus Q3)
2. Bulatov K. B., Bezmaternykh P. V., Nikolaev D. P., Arlazarov V. V. Towards a unified framework for identity documents analysis and recognition // Компьютерная оптика. — 2022. — Т. 46. — № 3. — С. 436-454. — DOI: 10.18287/2412-6179-CO-1024. (ВАК, WoS, Scopus Q1)
3. Bulatov K. B., Emelyanova E. V., Tropin D. V., Skoryukina N. S., Chernyshova Y. S., Sheshkus A. V., Usilin S. A., Ming Z., Burie J., Luqman M., Arlazarov V. V. MIDV-2020: A Comprehensive Benchmark Dataset for Identity Document Analysis // Компьютерная оптика. — 2022. — Т. 46. — № 2. — С. 252-270. — DOI: 10.18287/2412-6179-CO-1006. (ВАК, WoS, Scopus Q1)
4. Arlazarov V. V., Andreeva E. I., Bulatov K. B., Nikolaev D. P., Petrova O. O., Savelev B. I., Slavin O. A. Document image analysis and recognition: A survey // Компьютерная оптика. — 2022. — Т. 46. — № 4. — С. 567-589. — DOI: 10.18287/2412-6179-CO-1020. (ВАК, WoS, Scopus Q1)
5. Polevoy D. V., Sigareva I. V., Ershova D. M., Arlazarov V. V., Nikolaev D. P., Zuheng M., Muhammad M. L., Burie J. Document Liveness Challenge dataset (DLC-2021) // J. Imaging. — 2022. — Vol. 8. — No 7. — P. 181-1-181-12. — DOI: 10.3390/jimaging8070181. (WoS, Scopus Q2)
6. Gayer A. V., Ershova D. M., Arlazarov V. V. Fast and accurate deep learning model for stamps detection for embedded devices // Pattern Recognit. Image Anal.. — 2022. — Vol. 32. — No 4. — P. 772-779. — DOI: 10.1134/S1054661822040046. (ВАК, WoS, Scopus Q3)

7. Matalov D. P., Limonova E. E., Skoryukina N. S., Arlazarov V. V. Memory effcient local features descriptor for identity document detection on mobile and embedded devices // IEEE Access. — 2022. — Vol. 11. — P. 1104-1114. — DOI: 10.1109/ACCESS.2022.3233463. (WoS Q2, Scopus Q1)
8. Арлазаров В. В. Проблемы и особенности 2D, 3D и 4D-систем распознавания документов, удостоверяющих личность // Труды ИСА РАН. — 2022. — Т. 72. — № 3. — С. 3-9. — DOI: 10.14357/20790279220301. (ВАК)
9. Арлазаров В. В. Ключевые этапы обработки шаблона документа современных систем распознавания ID-карт // Труды ИСА РАН. — 2022. — Т. 72. — № 3. — С. 19-25. — DOI: 10.14357/20790279220303. (ВАК)
10. Арлазаров В. В. Анализ использования проблемно-ориентированных пакетов данных в научных исследованиях // ИТиВС. — 2022. — № 3. — С. 10-23. — DOI: 10.14357/20718632220302. (ВАК)
11. Арлазаров В. В. Методы комбинирования множественных результатов распознавания текста // Искусственный интеллект и принятие решений. — 2022. — № 3. — С. 106-116. — DOI: 10.14357/20718594220309. (ВАК)
12. Arlazarov V. V., Chuiko A. V., Slavin O. A. A Model for Assessing the Reliability of Document Text Field Recognition // ИТиВС. — 2022. — № 4. — С. 3-12. — DOI: 10.14357/20718632220401. (ВАК)
13. Bulatov K. B., Fedotova N. V., Arlazarov V. V. Fast Approximate Modelling of the Next Combination Result for Stopping the Text Field Recognition in a Video Stream // ICPR 2020 / Manhattan, New York, U.S: Institute of Electrical and Electronics Engineers (IEEE). — 2021. — ISSN 1051-4651. — ISBN 978-17-28188-09-6. — P. 239-246. — DOI: 10.1109/ICPR48806.2021.9412574. (WoS, Scopus)
14. Petrova O. O., Bulatov K. B., Arlazarov V. V., Arlazarov V. L. Weighted combination of per-frame recognition results for text recognition in a video stream // Компьютерная оптика. — 2021. — Т. 45. — № 1. — С. 77-89. — DOI: 10.18287/2412-6179-CO-795. (ВАК, WoS, Scopus Q1)
15. Kondrashev I. V., Sheshkus A. V., Arlazarov V. V. Distance-based online pairs generation method for metric networks training // ICMV 2020 / Bellingham, Washington 98227-0010 USA: Society of Photo-Optical Instrumentation Engineers (SPIE). — 2021. — Vol. 11605. — ISSN

0277-786X. — ISBN 978-15-10640-40-5. — 2021. — Vol. 11605. — P. 1160508-1-1160508-6. — DOI: 10.1117/12.2587175. (WoS, Scopus)

16. Skoryukina N. S., Arlazarov V. V., Milovzorov A. N. Memory Consumption Reduction for Identity Document Classification with Local and Global Features Combination // ICMV 2020 / Bellingham, Washington 98227-0010 USA: Society of Photo-Optical Instrumentation Engineers (SPIE). — 2021. — Vol. 11605. — ISSN 0277-786X. — ISBN 978-15-10640-40-5. — 2021. — Vol. 11605. — 116051G. — C. 116051G1-116051G8. — DOI: 10.1117/12.2587033. (WoS, Scopus)
17. Y Shemyakina Y. A., Limonova E. E., Skoryukina N. S., Arlazarov V. V., Nikolaev D. P. A method of image quality assessment for text recognition on camera-captured and projectively distorted documents // Mathematics. — 2021. — Vol. 9. — No 17. — P. 1-22. — DOI: 10.3390/math9172155. (WoS Q1, Scopus Q1)
18. Matalov D. P., Limonova E. E., Skoryukina N. S., Arlazarov V. V. RFDoc: memory efficient local descriptors for ID documents localization and classification // ICDAR 2021. — 2 изд. / Josep Lladós, Daniel Lopresti, Seiichi Uchida. — London, UK (main office): Springer Nature Group. — (Lecture Notes in Computer Science (LNCS)). — 2021. — Vol. 12822. — ISSN 0302-9743. — ISBN 978-3-03086-330-2. — 2021. — Vol. 12822. — P. 209-224. — DOI: 10.1007/978-3-030-86331-9_14. (Scopus)
19. Bulatov K. B., Arlazarov V. V. Determining optimal frame processing strategies for real-time document recognition systems // ICDAR 2021. — 2nd ed. / Josep Lladós, Daniel Lopresti, Seiichi Uchida. — London, UK (main office): Springer Nature Group. — (Lecture Notes in Computer Science (LNCS)). — 2021. — Vol. 12822. — ISSN 0302-9743. — ISBN 978-3-03086-330-2. — 2021. — Vol. 12822. — P. 273-288. — DOI: 10.1007/978-3-030-86331-9_18. (Scopus)
20. Chernyshova Y. S., Emelianova E. V., Sheshkus A. V., Arlazarov V. V. MIDV-LAIT: a challenging dataset for recognition of IDs with Perso-Arabic, Thai, and Indian scripts // ICDAR 2021. — 2nd ed. / Josep Lladós, Daniel Lopresti, Seiichi Uchida. — London, UK (main office): Springer Nature Group. — (Lecture Notes in Computer Science (LNCS)). —2021. — Vol. 12822. — ISSN 0302-9743. — ISBN 978-3-03086-330-2. — 2021. — Vol. 12822. — P. 258-272. — DOI: 10.1007/978-3-030-86331-9_17. (Scopus)

21. Arlazarov V. V., Voysyat J. S., Matalov D. P., Nikolaev D. P., Usilin S. A. Evolution of the Viola-Jones object detection method: a survey // Вестник ЮУрГУ ММП. — 2021. — Т. 14. — № 4. — С. 5-23. — DOI: 10.14529/mmp210401. (ВАК, WoS, Scopus Q3)
22. Skoryukina N., Arlazarov V. V., Nikolaev D. P. Fast method of ID documents location and type identification for mobile and server application // ICDAR 2019 / Manhattan, New York, U.S.: The Institute of Electrical and Electronics Engineers (IEEE). — 2020. — ISSN 2379-2140. — ISBN 978-17-28130-14-9. — P. 850-857. — DOI: 10.1109/ICDAR.2019.00141. (WoS, Scopus)
23. Matalov D. P., Usilin S. A., Arlazarov V. V. Single-sample augmentation framework for training Viola-Jones classifiers // ICMV 2019 / Wolfgang Osten, Dmitry Nikolaev, Jianhong Zhou. — Bellingham, Washington 98227-0010 USA: Society of Photo-Optical Instrumentation Engineers (SPIE). — 2020. — Vol. 11433. — ISSN 0277-786X. — ISBN 978-15-10636-44-6. — 2020. — Vol. 11433. — С. 114330I1-114330I8. — DOI: 10.1117/12.2559435. (WoS, Scopus)
24. Bulatov K., Matalov D., Arlazarov V. V. MIDV-2019: Challenges of the Modern Mobile-Based Document OCR // ICMV 2019 / Wolfgang Osten, Dmitry Nikolaev, Jianhong Zhou. — Bellingham, Washington 98227-0010 USA: Society of Photo-Optical Instrumentation Engineers (SPIE). — 2020. — Vol. 11433. — ISSN 0277-786X. — ISBN 978-15-10636-44-6. — 2020. — Vol. 11433. — P. 114332N1-114332N6. — DOI: 10.1117/12.2558438. (WoS, Scopus)
25. Chernyshova Y. S., Sheshkus A. V., Arlazarov V. V. Two-step CNN framework for text line recognition in camera-captured images // IEEE Access. — 2020. — Vol. 8. — P. 32587-32600. — DOI: 10.1109/ACCESS.2020.2974051. (WoS Q1, Scopus Q1)
26. Bulatov K. B., Savelyev B. I., Arlazarov V. V., Fedotova N. V. Analysis of a stopping method for text recognition in video stream using an extended result model with per-character alternatives // Сенсорные системы. — 2020. — Т. 34. — № 3. — С. 217-225. — DOI: 10.31857/S0235009220030026. (ВАК)
27. Лимонова Е. Е., Бочаров Н. А., Парамонов Н. Б., Богданов Д. С., Арлазаров В. В., Славин О. А., Николаев Д. П. Оценка быстродействия

системы распознавания на VLIW архитектуре на примере платформы Эльбрус // Программирование. — 2019. — № 1. — С. 15-21. — DOI: 10.1134/S0132347419010047. [E. E. Limonova, N. A. Bocharov, N. B. Paramonov, D. S. Bogdanov, V. V. Arlazarov, O. A. Slavin and D. P. Nikolaev, "Performance Evaluation of a Recognition System on the VLIW Architecture by the Example of the Elbrus Platform," PACS, Vol. 45, No 1, P. 12-17, 2019, DOI: 10.1134/S0361768819010055.] (BAK, WoS Q4, Scopus Q4)

28. Matalov D. P., Usilin S. A., Arlazarov V. V. Modification of the Viola-Jones approach for the detection of the government seal stamp of the Russian Federation // ICMV 2018 / Bellingham, Washington 98227-0010 USA: Society of Photo-Optical Instrumentation Engineers (SPIE). —2019. — Vol. 11041. — ISSN 0277-786X. — ISBN 978-15-10627-48-2. — 2019. — Vol. 11041. — P. 110411Y1-110411Y7. — DOI: 10.1117/12.2522793. (WoS, Scopus)
29. Andreeva E., Arlazarov V. V., Slavin O., Janiszewski I. Experimental modeling the flow of character recognition results in video stream for document recognition // ICMV 2018 / Bellingham, Washington 98227-0010 USA: Society of Photo-Optical Instrumentation Engineers (SPIE). — 2019. — Vol. 11041. — ISSN 0277-786X. — ISBN 978-15-10627-48-2. — 2019. — Vol. 11041. — P. 110411L1-110411L6. — DOI: 10.1117/12.2522970. (WoS, Scopus)
30. Arlazarov V. V., Bulatov K., Chernov T., Arlazarov V. L. MIDV-500: A Dataset for Identity Document Analysis and Recognition on Mobile Devices in Video Stream // Компьютерная оптика. — 2019. — Т. 43. — № 5. — С. 818-824. — DOI: 10.18287/2412-6179-2019-43-5-818-824. (BAK, WoS, Scopus Q1)
31. Bulatov K., Razumnyi N., Arlazarov V. V. On optimal stopping strategies for text recognition in a video stream as an application of a monotone sequential decision model // IJDAR. — 2019. — Vol. 22. — No 3. — P. 303-314. — DOI: 10.1007/s10032-019-00333-0. (WoS Q2, Scopus Q2)
32. Arlazarov V. V., Bulatov K., Manzhikov T., Slavin O. A., Janiszewski I. Method of determining the necessary number of observations for video stream documents recognition // ICMV 2017 / Antanas Verikas, Petia Radeva, Dmitry Nikolaev, Jianhong Zhou. — Bellingham, Washington

98227-0010 USA: Society of Photo-Optical Instrumentation Engineers (SPIE). — 2018. — Vol. 10696. — 758 p. — ISBN 978-15-10619-41-8. — 2018. — Vol. 10696. — P. 106961X1-106961X6. — DOI: 10.1117/12.2310132. (WoS, Scopus)

33. Chernov T. S., Razumnuy N. P., Kozharinov A. S., Nikolaev D. P., Arlazarov V. V. Image quality assessment for video stream recognition systems // ICMV 2017 / Antanas Verikas, Petia Radeva, Dmitry Nikolaev, Jianhong Zhou. — Bellingham, Washington 98227-0010 USA: Society of Photo-Optical Instrumentation Engineers (SPIE). — 2018. — Vol. 10696. — 758 p. — ISBN 978-15-10619-41-8. — 2018. — Vol. 10696. — P. 106961U1-106961U8. — DOI: 10.1117/12.2309628. (WoS, Scopus)
34. Arlazarov V. V., Slavin O. A., Uskov A. V., Janiszewski I. M. Modelling the flow of character recognition results in video stream // Вестник ЮУрГУ ММП. — 2018. — Т. 11. — № 2. — С. 14-28. — DOI: 10.14529/mmp180202. (ВАК, WoS, Scopus Q3)
35. Арлазаров В. В., Булатов К. Б., Усков А. В. Модель системы распознавания объектов в видеопотоке мобильного устройства // Труды ИСА РАН. — 2018. — Т. 68. — Спецвыпуск № S1. — С. 73-82. — DOI: 10.14357/20790279180508. (ВАК)
36. Арлазаров В. В., Маталов Д. П., Усилин С. А. Локализация образа печати на документе, удостоверяющем личность, методом машинного обучения // Труды ИСА РАН. — 2018. — Т. 68. — Спецвыпуск № S1. — С. 158-166. — DOI: 10.14357/20790279180518. (ВАК)
37. Слугин Д. Г., Арлазаров В. В. Поиск текстовых полей документа с помощью методов обработки изображений // Труды ИСА РАН. — 2017. — Т. 67. — № 4. — С. 65-73. (ВАК)
38. Чернов Т. С., Разумный Н. П., Кожаринов А. С., Николаев Д. П., Арлазаров В. В. Оценка качества входных изображений в системах распознавания видеопотока // ИТиВС. — 2017. — № 4. — С. 71-82. (ВАК)
39. Bulatov K., Arlazarov V. V., Chernov T., Slavin O., Nikolaev D. Smart IDReader: Document Recognition in Video Stream // ICDAR 2017 / Manhattan, New York, U.S.: Institute of Electrical and Electronics Engineers Inc. (IEEE). — ISSN 2379-2140. — ISBN 978-15-38635-86-5. — 2017. — Vol. 6. — P. 39-44. — DOI: 10.1109/ICDAR.2017.347. (WoS, Scopus)

40. Limonova E., Terekhin A., Nikolaev D., Arlazarov V. Fast Implementation of Morphological Filtering Using ARM NEON Extension // IJAER. — 2016. — Vol. 11. — No 24. — P. 11675-11680. (Scopus)

В рамках работы по диссертации было получено 2 патента США, 5 патентов на изобретение РФ, 16 патентов на полезные модели и 4 свидетельства о государственной регистрации программы для ЭВМ.

1. Patent No.: US20220067363A1. Efficient Location and Identification of Documents in Images / Skoryukina N.S., Arlazarov V.V., Nikolaev D. P., Faradjev I.A.
2. Patent No.: US10354142B2. Method for holographic elements detection in video stream / Arlazarov V. V., Chernov T. S., Nikolaev D. P., Skoryukina N. S., Slavin O. A.
3. Пат. № 2771005 РФ. Способ детектирования голографической защиты на документах в видеопотоке / В.В. Арлазаров, Л. И. Коляскина, Д.П. Николаев, Д.В. Полевой, Д.В. Тропин, С.А. Усилин.
4. Пат. 2774058 № РФ. Способ определения (распознавания) факта предъявления цифровой копии документа в виде пересъемки экрана / В.В. Арлазаров, Д.П. Николаев, Д.В. Полевой, Д.Г. Слугин, И. А. Кунина, И. В. Сигарева. Заявка № 2021128626. Дата регистрации 14.06.2022 , Бюл. № 17
5. Пат. № 2750395 РФ. Способ оценки действительности документа при помощи оптического распознавания текста на изображении круглого оттиска печати/штампа на цифровом изображении документа / М. А. Алиев , В. В. Арлазаров , Д. П. Маталов, Д. П. Николаев , Д. В. Полевой , С. А. Усилин; № 2020127688 заявл. 19.08.2020; опубл. 28.06.2021, Бюл. № 19 - 10 с.
6. Пат. № 2724967 РФ. Система дистанционного приобретения билетов на культурно-массовые мероприятия с использованием распознавания на мобильном устройстве / В. В. Арлазаров,Н. В. Арлазаров, Д. П. Николаев, О. А. Славин, С. А. Усилин, А. В. Шешкус; № 220110146 заявл. 11.03.2020; опубл. 29.06.2020, Бюл. № 19 - 10 с.
7. Пат. № 2643130 РФ. Автоматизированное рабочее место контроля паспортных документов / В.В. Арлазаров, А.П. Гладков, Д.П. Николаев,

С.А. Усилин; № 2017105336; заявл. 20.02.2017; опубл. 30.01.2018, Бюл. № 4 - 13 с.

8. Пат. на ПМ № 210539 РФ. Система дистанционной регистрации граждан на избирательном участке с распознаванием паспорта РФ / Арлазаров В. В., Арлазаров Н. В., Булатов К. Б., Славин О. А., Усилин С. А., 2021136611; заявлено 10.12.2021; - 2 с.
9. Пат. на ПМ № 210845. Система контроля соблюдения санитарно-эпидемиологических правил при дистанционной продаже билетов на транспорт при помощи мобильных устройств / Арлазаров В. В., Арлазаров Н. В., Безматерных П. В., Булатов К. Б., Полевой Д. В., Славин О. А., 2022100687; заявлено 12.01.2022; - 2 с.
10. Пат. на ПМ №210846 РФ. Система предотвращения атаки на предъявление дистанционного скоринга клиентов при выдаче кредитов с помощью мобильного устройства / Арлазаров В. В., Арлазаров Н. В., Безматерных П. В., Булатов К. Б., Кунина И. А,, Маталов Д. П., Тропин Д. В., Усилин С. А., 2022102438 заявлено; 01.02.2022 - 2 с.
11. Пат. на ПМ № 210919 РФ. Система контроля соблюдения правил дистанционной торговли рецептурными препаратами (лекарствами) при помощи мобильного устройства / Арлазаров В. В., Арлазаров Н. В., Лимонова Е. Е., Маталов Д. П., Полевой Д. В., Тропин Д. В., 2022101508; заявлено 24.01.2022; - 2 с.
12. Пат. на ПМ № 211342 РФ. Система дистанционного опроса граждан при переписи населения с распознаванием электронного паспорта РФ на пластиковой карте / Арлазаров В. В., Арлазаров Н. В., Маталов Д. П., Славин О. А., Шешкус А. В., 2021135025; заявлено 29.11.2021; - 2 с.
13. Пат. на ПМ № 204787 РФ. Система удаленной регистрации абонентов сети связи с использованием мобильного устройства / Безматерных П. В., Арлазаров В. В., Арлазаров Н. В., Скорюкина Н. С., Славин О. А., № 2021100924; заявлено 18.01.2021; опублик. 10.06.2021, Бюл. № 16 - 2 с.
14. Пат. на ПМ № 204371 РФ. Система автоматического подтверждения отсутствия признаков инфекционного заболевания с помощью документального свидетельства / Арлазаров В. В., Арлазаров Н. В., Булатов К. Б., Славин О. А., № 2021103918; заявлено 24.03.2021; опублик. 21.05.2021, Бюл. № 15 - 2 с.

15. Пат. на ПМ № 207759 РФ. Система проверки подлинности паспортно-визовых документов с использованием мобильного устройства / Арлазаров В. В., Арлазаров Н. В., Усилин С. А., 2021125026; заявлено 23.08.2021; опублик. 15.11.2021, Бюл. № 32 - 2 с.
16. Пат. на ПМ № 196455 РФ. Система удаленной регистрации данных граждан на избирательном участке с использованием мобильного устройства / Арлазаров В. В., Пат. № 196455 РФ. Арлазаров Н. В., Булатов К. Б., Шешкус А. В., № 2019141280; заявл. 13.12.2019; опубл. 02.03.2020; Бюл. № 7 - 2с.
17. Пат. на ПМ № 191682. Система покупки цифрового контента с использованием мобильного устройства / Арлазаров В. В., Арлазаров Н. В., Булатов К. Б., Скорюкина Н. С., Николаев Д. П.; № 2019116550; заявл. 29.05.2019; опубл. 15.08.2019; Бюл. № 23 - 2 с.
18. Пат. на ПМ № 180207 РФ. Система автоматического контроля личности избирателя / В. В. Арлазаров, Н.В. Арлазаров, К. Б. Булатов, Н. С. Скорюкина, О.А. Славин, Т.С. Чернов; № 2017139215; заявл. 13.11.2017; опубл. 06.06.2018, Бюл. № 16 - 2 с.
19. Пат. на ПМ № 161478 РФ. Система доступа к дистанционному получению банковских услуг / В.В. Арлазаров, А.Р. Арлазарова, Н.В. Арлазаров, Д.П. Николаев, О.А. Славин, С.А. Усилин, А.В. Шешкус; № 2015156508/08; заявл. 29.12.2015; опубл. 20.04.2016, Бюл. № 11 - 3 с.
20. Пат. на ПМ № 159733 РФ. Система распознавания документов в видеопоследовательности / В.В. Арлазаров, В.Л. Арлазаров, К.Б. Булатов, Д.П. Николаев, Д.В.Полевой, О.А. Славин; № 2015145155; заявл. 21.10.2015; опубл. 20.02.2016, Бюл. № 5 - 3 с.
21. Пат. на ПМ № 166152 РФ. Автономное автоматизированное рабочее место контроля паспортных документов / В.В. Арлазаров, А.П. Гладков, Д.П. Николаев, С.А. Усилин; 2016122432/08; заявл. 07.07.2016; опубл. 20.11.2016, Бюл. № 32 - 2 с.
22. Пат. на ПМ № 163168 РФ. Технологическая платформа электронного документооборота осмотра автомобиля для оформления страховки / В.В. Арлазаров, А.Р. Арлазарова, О.А. Славин, С.А. Усилин; № 2015148236; заявл. 10.11.2015; опубл. 10.07.2016, Бюл. № 12 - 4 с.
23. Пат. на ПМ № 166038 РФ. Автоматизированное рабочее место контроля паспортных документов / В.В. Арлазаров, А.П. Гладков, Д.П. Нико-

лаев, С.А. Усилин; 2016106183/08; заявл. 25.02.2016; опубл. 10.11.2016, Бюл. № 31 - 2 с.

24. Программа точной локализации и типизации объекта страницы документа в видеопотоке: свидетельство о государственной регистрации программы для ЭВМ № 2019665480 / Арлазаров В.В., Янишевский И.М., № 2019664400; заявл. 13.11.2019; зарегистрировано в реестре программ для ЭВМ 22.11.2019. - [1] с. ЭВМ № RU2019665480
25. Программа распознавания признаков подлинности “Smart Document Forensics”: свидетельство о государственной регистрации программы для ЭВМ № 2018615343 / Усилин С.А., Арлазаров В.В., Алиев М.В., Маталов Д.П. , № 2018612851; заявл. 23.03.2018; зарегистрировано в реестре программ для ЭВМ 07.05.2018. - [1] с.
26. Программа поиска плоских ригидных объектов “Smart ARTour”: свидетельство о государственной регистрации программы для ЭВМ № 2018615952 / Арлазаров В.В., Булатов К.Б., Николаев Д.П., Скорюкина Н.С., № 2018612805; зарегистрировано в реестре программ для ЭВМ 18.05.2018. - [1] с.
27. Программа для распознавания идентификационных карт личности “Smart IDReader”: свидетельство о государственной регистрации программы для ЭВМ № 2016616961 / Арлазаров В.В., Николаев Д.П., Усилин С.А., Булатов К.Б., Чернов Т.С., Слугин Д.Г., Ильин Д.А., Безматерных П.В., Муковозов А.А., Лимонова Е. Е., № 2016612014; заявл. 10.03.2016; зарегистрировано в реестре программ для ЭВМ 22.06.2016. - [1] с.

Список литературы

1. SJR. Scimago Journal & Country Rank. Proc Int Conf on Document Analysis and Recognition (ICDAR) [Электронный ресурс]. — 2022. — URL: https://www.scimagojr.com/journalsearch.php?q=75898&tip=sid (дата обр. 02.11.2022).
2. Document image classification: Progress over two decades [Текст] / L. Liu [et al.] // Neurocomputing. — 2021. — Vol. 453. — P. 223—240.
3. Efficient Automated Processing of the Unstructured Documents Using Artificial Intelligence: A Systematic Literature Review and Future Directions [Текст] / D. Baviskar [et al.] // IEEE Access. — 2021. — Vol. 9. — P. 72894—72936.
4. *Rehman, A.* Document skew estimation and correction: analysis of techniques, common problems and possible solutions [Текст] / A. Rehman, T. Saba // Applied Artificial Intelligence. — 2011. — Vol. 25, no. 9. — P. 769—787.
5. *Chen, D.* A survey of text detection and recognition in images and videos [Текст] / D. Chen, J. Luettin, K. Shearer // IDIAP Research Report. — 2000. — Vol. 00—38. — P. 1—21.
6. *Nagy, G.* Twenty years of document image analysis in PAMI [Текст] / G. Nagy // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2000. — Vol. 22, no. 1. — P. 38—62.
7. *Mao, S.* Document structure analysis algorithms: a literature survey [Текст] / S. Mao, A. Rosenfeld, T. Kanungo // Document Recognition and Retrieval X. Vol. 5010 / ed. by T. Kanungo [et al.]. — International Society for Optics, Photonics. SPIE, 2003. — P. 197—207.
8. *Doermann, D.* Progress in camera-based document image analysis [Текст] / D. Doermann, J. Liang, H. Li // Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. Vol. 1. — 2003. — P. 606—616.
9. *Zanibbi, R.* A survey of table recognition [Текст] / R. Zanibbi, D. Blostein, J. Cordy // International Journal of Document Analysis and Recognition. — 2004. — Vol. 7. — P. 1—16.

10. *Jung, K.* Text information extraction in images and video: a survey [Текст] / K. Jung, K. In Kim, A. K. Jain // Pattern Recognition. — 2004. — Vol. 37, no. 5. — P. 977—997.

11. *Liang, J.* Camera-based analysis of text and documents: a survey [Текст] / J. Liang, D. Doermann, H. Li // International Journal of Document Analysis and Recognition. — 2005. — Vol. 7. — P. 84—104.

12. *Marinai, S.* Artificial neural networks for document analysis and recognition [Текст] / S. Marinai, M. Gori, G. Soda // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2005. — Vol. 27, no. 1. — P. 23—35.

13. *Chen, N.* A survey of document image classification: problem statement, classifier architecture and performance evaluation [Текст] / N. Chen, D. Blostein // International Journal of Document Analysis and Recognition. — 2007. — Vol. 10. — P. 1—16.

14. A review of machine learning algorithms for text-documents classification [Текст] / A. Khan [et al.] // Journal of Advances in Information Technology. — 2010. — Vol. 1, no. 1. — P. 4—20.

15. *Dixit, U.* A survey on document image analysis and retrieval system [Текст] / U. Dixit, M. Shirdhonkar // International Journal on Cybernetics and Informatics. — 2015. — Vol. 4, no. 2. — P. 259—270.

16. *Eskenazi, S.* A comprehensive survey of mostly textual document segmentation algorithms since 2008 [Текст] / S. Eskenazi, P. Gomez-Krämer, J.-M. Ogier // Pattern Recognition. — 2017. — Vol. 64. — P. 1—14.

17. *Binmakhashen, G. M.* Document Layout Analysis: A Comprehensive Survey [Текст] / G. M. Binmakhashen, S. A. Mahmoud // ACM Comput. Surv. — New York, NY, USA, 2019. — Vol. 52, no. 6. — Art. No. 109.

18. *Lombardi, F.* Deep Learning for Historical Document Analysis and Recognition—A Survey [Текст] / F. Lombardi, S. Marinai // Journal of Imaging. — 2020. — Vol. 6, no. 10. — Art. No. 110.

19. A Survey of Graphical Page Object Detection with Deep Neural Networks [Текст] / J. Bhatt [et al.] // Applied Sciences. — 2021. — Vol. 11, no. 12. — Art. No. 5344.

20. *Doermann, D.* Handbook of document image processing and recognition [Текст] / D. Doermann, K. Tombre. — Springer-Verlag London, 2014. — 1055 p.

21. *Liu, C.-L.* Advances in Chinese Document and Text Processing [Текст] / C.-L. Liu, Y. Lu. — WORLD SCIENTIFIC, 2017. — 292 p.

22. *Fischer, A.* Handwritten Historical Document Analysis, Recognition, and Retrieval — State of the Art and Future Trends [Текст] / A. Fischer, M. Liwicki, R. Ingold. — WORLD SCIENTIFIC, 2020. — 268 p.

23. *Dey, S.* Light-Weight Document Image Cleanup Using Perceptual Loss [Текст] / S. Dey, P. Jawanpuria // Document Analysis and Recognition – ICDAR 2021 / ed. by J. Lladós, D. Lopresti, S. Uchida. — Springer International Publishing, 2021. — P. 238—253.

24. *Bloomberg, D. S.* Measuring document image skew and orientation [Текст] / D. S. Bloomberg, G. E. Kopec, L. Dasari // Document Recognition II. Vol. 2422 / ed. by L. M. Vincent, H. S. Baird. — International Society for Optics, Photonics. SPIE, 1995. — P. 302—316.

25. *Hull, J. J.* Document image skew detection: survey and annotated bibliography [Текст] / J. J. Hull // Document Analysis Systems II. — P. 40—64.

26. *Bezmaternykh, P. V.* A document skew detection method using fast Hough transform [Текст] / P. V. Bezmaternykh, D. P. Nikolaev // Twelfth International Conference on Machine Vision (ICMV 2019). Vol. 11433 / ed. by W. Osten, D. P. Nikolaev. — International Society for Optics, Photonics. SPIE, 2020. — 114330J.

27. *Akhter, S. S. M. N.* Improving Skew Detection and Correction in Different Document Images Using a Deep Learning Approach [Текст] / S. S. M. N. Akhter, P. P. Rege // 2020 11th International Conference on Computing, Communication and Networking Technologies (ICCCNT). — 2020. — P. 1—6.

28. ICDAR 2013 Document Image Skew Estimation Contest (DISEC 2013) [Текст] / A. Papandreou [et al.] // 2013 12th International Conference on Document Analysis and Recognition. — 2013. — P. 1444—1448.

29. *Fabrizio, J.* A precise skew estimation algorithm for document images using KNN clustering and fourier transform [Текст] / J. Fabrizio // 2014 IEEE International Conference on Image Processing (ICIP). — 2014. — P. 2585—2588.

30. *Uchida, S.* Nonuniform slant correction using dynamic programming [Текст] / S. Uchida, E. Taira, H. Sakoe // Proceedings of Sixth International Conference on Document Analysis and Recognition. — 2001. — P. 434—438.

31. Distort-and-recover: Color enhancement using deep reinforcement learning [Текст] / J. Park [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2018. — P. 5928—5936.

32. *Simonyan, K.* Very Deep Convolutional Networks for Large-Scale Image Recognition [Текст] / K. Simonyan, A. Zisserman // CoRR. — 2014. — Vol. abs/1409.1556.

33. Exposure: A white-box photo post-processing framework [Текст] / Y. Hu [et al.] // ACM Transactions on Graphics (TOG). — 2018. — Vol. 37, no. 2. — P. 26.

34. Progressive Image Enhancement under Aesthetic Guidance [Текст] / X. Du [et al.] // Proceedings of the 2019 on International Conference on Multimedia Retrieval. — Ottawa ON, Canada : Association for Computing Machinery, 2019. — P. 349—353. — (ICMR '19).

35. Image-to-image translation with conditional adversarial networks [Текст] / P. Isola [et al.] // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2017. — P. 1125—1134.

36. *Shan, C.* A Coarse-to-Fine Framework for Learned Color Enhancement with Non-Local Attention [Текст] / C. Shan, Z. Zhang, Z. Chen // 2019 IEEE International Conference on Image Processing (ICIP). — 2019. — P. 949—953.

37. Deep residual learning for image recognition [Текст] / K. He [et al.] // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2016. — P. 770—778.

38. Non-local Neural Networks [Текст] / X. Wang [et al.] // 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. — 2018. — P. 7794—7803.

39. *Chai, Y.* Supervised and Unsupervised Learning of Parameterized Color Enhancement [Текст] / Y. Chai, R. Giryes, L. Wolf // 2020 IEEE Winter Conference on Applications of Computer Vision (WACV). — 2020. — P. 981—989.

40. *Tatanov, O.* LFIEM: Lightweight Filter-based Image Enhancement Model [Текст] / O. Tatanov, A. Samarin // 2020 25th International Conference on Pattern Recognition (ICPR). — 2021. — P. 873—878.

41. Generative adversarial nets [Текст] / I. Goodfellow [et al.] // Advances in neural information processing systems. — 2014. — P. 2672—2680.

42. Deep photo enhancer: Unpaired learning for image enhancement from photographs with gans [Текст] / Y.-S. Chen [et al.] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2018. — P. 6306—6314.

43. *Deng, Y.* Aesthetic-Driven Image Enhancement by Adversarial Learning [Текст] / Y. Deng, C. C. Loy, X. Tang // Proceedings of the 26th ACM International Conference on Multimedia. — Seoul, Republic of Korea : Association for Computing Machinery, 2018. — P. 870—878. — (MM '18).

44. Removing Shadows from Images of Documents [Текст] / S. Bako [et al.] // Computer Vision – ACCV 2016 / ed. by S.-H. Lai [et al.]. — Springer International Publishing, 2017. — P. 173—183.

45. *Wang, B.* An Effective Background Estimation Method for Shadows Removal of Document Images [Текст] / B. Wang, C. L. P. Chen // 2019 IEEE Int. Conf. on Image Processing (ICIP). — 2019. — P. 3611—3615.

46. *Wang, J.* Shadow Removal of Text Document Images by Estimating Local and Global Background Colors [Текст] / J. Wang, Y. Chuang // ICASSP 2020 - 2020 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP). — 2020. — P. 1534—1538.

47. *S. Jung Md. A. Hasan, C. K.* Water-Filling: An Efficient Algorithm for Digitized Document Shadow Removal [Текст] / C. K. S. Jung Md. A. Hasan // 2018, 14th Asian Conf. on Computer Vision (ACCV). — 2018. — P. 398—414.

48. *Kligler, N.* Document Enhancement Using Visibility Detection [Текст] / N. Kligler, S. Katz, A. Tal // 2018 IEEE/CVF Conf. on Computer Vision and Pattern Recognition. — 2018. — P. 2374—2382.

49. *Lin, Y. .-.* BEDSR-Net: A Deep Shadow Removal Network From a Single Document Image [Текст] / Y. .-. Lin, W. .-. Chen, Y. .-. Chuang // 2020 IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR). — 2020. — P. 12902—12911.

50. Document Rectification and Illumination Correction Using a Patch-Based CNN [Текст] / X. Li [et al.] // ACM Trans. Graph. — New York, NY, USA, 2019. — Vol. 38, no. 6.

51. Skip-Connected Deep Convolutional Autoencoder for Restoration of Document Images [Текст] / G. Zhao [et al.] // 2018 24th Int. Conf. on Pattern Recognition (ICPR). — 2018. — P. 2935—2940.

52. *Souibgui, M. A.* DE-GAN: A Conditional Generative Adversarial Network for Document Enhancement [Текст] / M. A. Souibgui, Y. Kessentini // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2022. — Vol. 44, no. 3. — P. 1180—1191.

53. DeepLPF: Deep Local Parametric Filters for Image Enhancement [Текст] / S. Moran [et al.] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2020. — P. 12823—12832.

54. Underexposed Photo Enhancement Using Deep Illumination Estimation [Текст] / R. Wang [et al.] // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2019. — P. 6842—6850.

55. *K. Kise.* Page Segmentation Techniques in Document Analysis [Текст] / K. Kise // Handbook of Document Image Processing and Recognition / ed. by D. Doermann, K. Tombre. — Springer London, 2014. — P. 135—175.

56. *Otsu, N.* A Threshold Selection Method from Gray-Level Histograms [Текст] / N. Otsu // IEEE Trans. on Systems, Man, and Cybernatics. — 1979. — Vol. 9, no. 1. — P. 62—66.

57. *Lu, S.* Document image binarization using background estimateion and stroke edges [Текст] / S. Lu, B. Su, C. L. Tan // International Journal of Document Analysis and Recognition. — 2010. — Vol. 13, no. 4. — P. 303—314.

58. *Moghaddam, R. F.* AdOtsu: An adaptive and parameterless generalization of Otsu's method for document image binarization [Текст] / R. F. Moghaddam, M. Cheriet // Pattern Recognition. — 2012. — Vol. 45, no. 6. — P. 2419—2431.

59. *Gatos, B.* Adaptive degraded document image binarization [Текст] / B. Gatos, I. Pratikakis, S. Perantonis // Pattern Recognition. — 2006. — Vol. 39, no. 3. — P. 317—327.

60. A generalization of Otsu method for linear separation of two unbalanced classes in document image binarization [Текст] / E. I. Ershov [et al.] // Computer Optics. — 2021. — Vol. 45, no. 1. — P. 66—76.

61. *Sauvola, J.* Adaptive document image binarization [Текст] / J. Sauvola, M. Pietikäinen // Pattern Recognition. — 2000. — Vol. 33. — P. 225—236.

62. *Lazzara, G.* Efficient Multiscale Sauvola's Binarization [Текст] / G. Lazzara, T. Géraud // Int. Journal of Document Analysis and Recognition. — Berlin, Heidelberg, 2014. — Vol. 17, no. 2. — P. 105—123.

63. *Niblack, W.* An Introduction to Digital Image Processing [Текст] / W. Niblack. — Prentice-Hall, 1986. — 215 p.

64. *Vo, G. D.* Robust regression for image binarization under heavy noise and nonuniform background [Текст] / G. D. Vo, C. Park // Pattern Recognition. — 2018. — Vol. 81. — P. 224—239.

65. *Calvo-Zaragoza, J.* A selectional auto-encoder approach for document image binarization [Текст] / J. Calvo-Zaragoza, A.-J. Gallego // Pattern Recognition. — 2019. — Vol. 86. — P. 37—47.

66. *Bezmaternykh, P. V.* U-Net-bin: hacking the document image binarization contest [Текст] / P. V. Bezmaternykh, D. A. Ilin, D. P. Nikolaev // Computer Optics. — 2019. — Vol. 43, no. 5. — P. 825—832.

67. Exploiting Stroke Orientation for CRF Based Binarization of Historical Documents [Текст] / X. Peng [et al.] // 2013 12th Int. Conf. on Document Analysis and Recognition. — 2013. — P. 1034—1038.

68. Learning Free Document Image Binarization Based on Fast Fuzzy C-Means Clustering [Текст] / T. Mondal [et al.] // 2019 Int. Conf. on Document Analysis and Recognition (ICDAR). — 2019. — P. 1384—1389.

69. An Iterative Refinement Framework for Image Document Binarization with Bhattacharyya Similarity Measure [Текст] / N. Liu [et al.] // 14th Int. Conf. on Document Analysis and Recognition. — IEEE Computer Society, 2017. — P. 93—98. — (ICDAR '17).

70. Document image binarization [Электронный ресурс]. — 2022. — URL: https://dib.cin.ufpe.br (дата обр. 02.11.2022).

71. Assessing the relationship between binarization and OCR in the context of deep learning-based ID document analysis [Текст] / R. Sánchez-Rivero [et al.] // IWAIPR 2021. Vol. 13055. — London, UK (main office) : Springer Nature Group, 2021. — P. 134—144. — (Lecture Notes in Computer Science (LNCS)).

72. Challenge 1: Smartphone document capture competition [Электронный ресурс]. — 2022. — URL: https://sites.google.com/site/icdar15smartdoc/challenge-1 (дата обр. 02.11.2022).

73. *Schmid, C.* Local grayvalue invariants for image retrieval [Текст] / C. Schmid, R. Mohr // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1997. — Vol. 19, no. 5. — P. 530—535.

74. *Harris, C.* A Combined Corner and Edge Detector [Текст] / C. Harris, M. Stephens // Proceedings of the Alvey Vision Conference. — Alvety Vision Club, 1988. — P. 23.1—23.6.

75. *Rosten, E.* Machine Learning for High-Speed Corner Detection [Текст] / E. Rosten, T. Drummond // Computer Vision – ECCV 2006 / ed. by A. Leonardis, H. Bischof, A. Pinz. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2006. — P. 430—443.

76. *Lowe, D. G.* Distinctive image features from scale-invariant keypoints [Текст] / D. G. Lowe // International journal of computer vision. — 2004. — Vol. 60, no. 2. — P. 91—110.

77. *Lepetit, V.* Towards recognizing feature points using classification trees [Текст] / V. Lepetit, P. Fua // Swiss Federal Institute of Technology (EPFL) Technical report. — 2004. — URL: https://infoscience.epfl.ch/record/52666.

78. Speeded-Up Robust Features (SURF) [Текст] / H. Bay [et al.] // Computer Vision and Image Understanding. — 2008. — Vol. 110, no. 3. — P. 346—359. — Similarity Matching in Computer Vision and Multimedia.

79. *Rosin, P. L.* Measuring Corner Properties [Текст] / P. L. Rosin // Computer Vision and Image Understanding. — 1999. — Vol. 73, no. 2. — P. 291—307.

80. *Leutenegger, S.* BRISK: Binary Robust invariant scalable keypoints [Текст] / S. Leutenegger, M. Chli, R. Y. Siegwart // 2011 International Conference on Computer Vision. — 2011. — P. 2548—2555.

81. *Zhang, H.* Extension and evaluation of the AGAST feature detector [Текст] / H. Zhang, J. Wohlfeil, D. Grießbach // ISPRS Ann. Photogramm. Remote Sens. Spatial Inf. — 2016. — Vol. 3, no. 4. — P. 133—137.

82. *Verma, R.* Enhanced character recognition using SURF feature and neural network technique [Текст] / R. Verma, R. Kaur // Int J Comput Sci Inf Technol Res. — 2014. — Vol. 5, no. 4. — P. 5565—5570.

83. A comparison of local features for camera-based document image retrieval and spotting [Текст] / O. B. Dang [et al.] // International Journal of Document Analysis and Recognition. — 2019. — Vol. 22. — P. 247—263.

84. Building a Test Collection for Complex Document Information Processing [Текст] / D. Lewis [et al.] // Proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval. — New York, NY, USA : Association for Computing Machinery, 2006. — P. 665—666.

85. The Legacy Tobacco Document Library (LTDL) [Электронный ресурс]. — 2007. — URL: http://legacy.library.ucsf.edu (дата обр. 02.11.2022).

86. *Zhang, Z.* Whiteboard scanning and image enhancement [Текст] / Z. Zhang, L.-W. He // Digital Signal Processing. — 2007. — Vol. 17, no. 2. — P. 414—432.

87. *Liu, N.* Dynamic detection of an object framework in a mobile device captured image [Текст] / N. Liu, L. Wang. — 2018. — US Patent US10134163B2.

88. *Hartl, A.* Rectangular target extraction for mobile augmented reality applications [Текст] / A. Hartl, G. Reitmayr // Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). — 2012. — P. 81—84.

89. Real time rectangular document detection on mobile devices [Текст] / N. Skoryukina [et al.] // Seventh International Conference on Machine Vision (ICMV 2014). Vol. 9445 / ed. by A. Verikas [et al.]. — International Society for Optics, Photonics. SPIE, 2015. — 94452A.

90. Automatic Business Card Scanning with a Camera [Текст] / G. Hua [et al.] // 2006 International Conference on Image Processing. — 2006. — P. 373—376.

91. Hierarchical Segmentation Using Tree-Based Shape Spaces [Текст] / Y. Xu [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2017. — Vol. 39, no. 3. — P. 457—469.

92. An Automatic Reader of Identity Documents [Текст] / F. Attivissimo [et al.] // 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC). — 2019. — P. 3525—3530.

93. Machine Learning Techniques for Identity Document Verification in Uncontrolled Environments: A Case Study [Текст] / A. Castelblanco [et al.] // Pattern Recognition / ed. by K. M. Figueroa Mora [et al.]. — Cham : Springer International Publishing, 2020. — P. 271—281.

94. *Sheshkus, A.* Houghencoder: Neural Network Architecture for Document Image Semantic Segmentation [Текст] / A. Sheshkus, D. Nikolaev, V. L. Arlazarov // 2020 IEEE International Conference on Image Processing (ICIP). — 2020. — P. 1946—1950.

95. *Javed, K.* Real-Time Document Localization in Natural Images by Recursive Application of a CNN [Текст] / K. Javed, F. Shafait // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 01. — 2017. — P. 105—110.

96. A Fast Fully Octave Convolutional Neural Network for Document Image Segmentation [Текст] / R. B. das Neves [et al.] // 2020 International Joint Conference on Neural Networks (IJCNN). — 2020. — P. 1—6.

97. *Viola, P.* Robust Real-Time Face Detection [Текст] / P. Viola, M. J. Jones // Int. J. Comput. Vision. — Hingham, MA, USA, 2004. — Vol. 57, no. 2. — P. 137—154.

98. Visual appearance based document image classification [Текст] / S. Usilin [et al.] // 2010 IEEE International Conference on Image Processing. — 2010. — P. 2133—2136.

99. *Roy, P. P.* Seal Detection and Recognition: An Approach for Document Indexing [Текст] / P. P. Roy, U. Pal, J. Lladós // 2009 10th International Conference on Document Analysis and Recognition. — 2009. — P. 101—105.

100. You Only Look Once: Unified, Real-Time Object Detection [Текст] / J. Redmon [et al.] // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2016. — P. 779—788.

101. *Bochkovskiy, A.* Yolov4: Optimal speed and accuracy of object detection [Текст] / A. Bochkovskiy, C.-Y. Wang, H.-Y. M. Liao // CoRR. — 2020. — T. arXiv/2004.10934.

102. *Wang, Y.* Comic frame extraction via line segments combination [Текст] / Y. Wang, Y. Zhou, Z. Tang // 2015 13th International Conference on Document Analysis and Recognition (ICDAR). — 2015. — P. 856—860.

103. *Slavin, O. A.* Using Special Text Points in the Recognition of Documents [Текст] / O. A. Slavin // Cyber-Physical Systems: Advances in Design & Modelling / ed. by A. G. Kravets, A. A. Bolshakov, M. V. Shcherbakov. — Cham : Springer International Publishing, 2020. — P. 43—53.

104. *Shafait, F.* The Effect of Border Noise on the Performance of Projection-Based Page Segmentation Methods [Текст] / F. Shafait, T. M. Breuel // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2011. — Vol. 33, no. 4. — P. 846—851.

105. *Melinda, L.* Document Layout Analysis Using Multigaussian Fitting [Текст] / L. Melinda, R. Ghanapuram, C. Bhagvati // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 01. — 2017. — P. 747—752.

106. CNN Based Page Object Detection in Document Images [Текст] / X. Yi [et al.] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 01. — 2017. — P. 230—235.

107. DoT-Net: Document Layout Classification Using Texture-Based CNN [Текст] / S. C. Kosaraju [et al.] // 2019 International Conference on Document Analysis and Recognition (ICDAR). — 2019. — P. 1029—1034.

108. Multi-Scale Multi-Task FCN for Semantic Page Segmentation and Table Detection [Текст] / D. He [et al.] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 01. — 2017. — P. 254—261.

109. A Robust Symmetry-Based Method for Scene/Video Text Detection through Neural Network [Текст] / Y. Wu [et al.] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 01. — 2017. — P. 1249—1254.

110. A Realistic Dataset for Performance Evaluation of Document Layout Analysis [Текст] / A. Antonacopoulos [et al.] // 2009 10th International Conference on Document Analysis and Recognition. — 2009. — P. 296—300.

111. COCO-Text: Dataset and Benchmark for Text Detection and Recognition in Natural Images [Текст] / A. Veit [et al.] // CoRR. — 2016. — Vol. arXiv/1601.07140.

112. The Maurdor Project: Improving Automatic Processing of Digital Documents [Текст] / S. Brunessaux [et al.] // 2014 11th IAPR International Workshop on Document Analysis Systems. — 2014. — P. 349—354.

113. *Soares, Á.* BID Dataset: a challenge dataset for document processing tasks [Текст] / Á. Soares, R. das Neves Junior, B. Bezerra // Anais Estendidos do XXXIII Conference on Graphics, Patterns and Images. — Evento Online : SBC, 2020. — P. 143—146.

114. Grayscale-Projection Based Optimal Character Segmentation for Camera-Captured Faint Text Recognition [Текст] / F. Jia [et al.] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 01. — 2017. — P. 1301—1306.

115. Multi-oriented touching text character segmentation in graphical documents using dynamic programming [Текст] / P. Pratim Roy [et al.] // Pattern Recognition. — 2012. — Vol. 45, no. 5. — P. 1972—1983.

116. *Saba, T.* Effects of artificially intelligent tools on pattern recognition [Текст] / T. Saba, A. Rehman // International Journal of Machine Learning and Cybernetics. — 2013. — Vol. 4. — P. 155—162.

117. *Alvear-Sandoval, R. F.* On improving CNNs performance: The case of MNIST [Текст] / R. F. Alvear-Sandoval, J. L. Sancho-Gómez, A. R. Figueiras-Vidal // Information Fusion. — 2019. — Vol. 52. — P. 106—109.

118. Understanding Deep Learning (Still) Requires Rethinking Generalization [Текст] / C. Zhang [et al.] // Commun. ACM. — New York, NY, USA, 2021. — Vol. 64, no. 3. — P. 107—115.

119. *Smith, R.* An Overview of the Tesseract OCR Engine [Текст] / R. Smith // Ninth International Conference on Document Analysis and Recognition (ICDAR 2007). Vol. 2. — 2007. — P. 629—633.

120. *Bahi, H. E.* Text recognition in document images obtained by a smartphone based on deep convolutional and recurrent neural network [Текст] / H. E. Bahi, A. Zatni // Multimedia Tools and Applications. — 2019. — Vol. 78. — P. 26453—26481.

121. *Rubner, Y.* The Earth Mover's Distance as a Metric for Image Retrieval [Текст] / Y. Rubner, C. Tomasi, L. J. Guibas // International Journal of Computer Vision. — 2000. — Vol. 40. — P. 99—121.

122. Language-invariant novel feature descriptors for handwritten numeral recognition [Текст] / S. Ghosh [et al.] // The Visual Computer. — 2021. — Vol. 37, no. 7. — P. 1781—1803.

123. Devnet: an efficient cnn architecture for handwritten devanagari character recognition [Текст] / R. Guha [et al.] // International Journal of Pattern Recognition and Artificial Intelligence. — 2020. — Vol. 34, no. 12. — P. 2052009.

124. *Melnyk, P.* A high-performance CNN method for offline handwritten Chinese character recognition and visualization [Текст] / P. Melnyk, Z. You, K. Li // Soft computing. — 2020. — Т. 24, № 11. — С. 7977—7987.

125. *Lincy, R. B.* Optimally configured convolutional neural network for Tamil Handwritten Character Recognition by improved lion optimization model [Текст] / R. B. Lincy, R. Gayathri // Multimedia Tools and Applications. — 2021. — Vol. 80, no. 4. — P. 5917—5943.

126. Arabic ligatures: Analysis and application in text recognition [Текст] / Y. Elarian [et al.] // 2015 13th International Conference on Document Analysis and Recognition (ICDAR). — 2015. — P. 896—900.

127. *Ilyuhin, S. A.* Recognition of images of Korean characters using embedded networks [Текст] / S. A. Ilyuhin, A. V. Sheshkus, V. L. Arlazarov // Twelfth International Conference on Machine Vision (ICMV 2019). Vol. 11433 / ed. by W. Osten, D. P. Nikolaev. — International Society for Optics, Photonics. SPIE, 2020. — P. 1143311.

128. *Pramanik, R.* A study on the effect of CNN-based transfer learning on handwritten Indic and mixed numeral recognition [Текст] / R. Pramanik, P. Dansena, S. Bag // Workshop on Document Analysis and Recognition. — Springer. 2018. — P. 41—51.

129. *Deore, S. P.* Devanagari handwritten character recognition using fine-tuned deep convolutional neural network on trivial dataset [Текст] / S. P. Deore, A. Pravin // Sādhanā. — 2020. — Vol. 45, no. 1. — P. 1—13.

130. Pioneer dataset and automatic recognition of Urdu handwritten characters using a deep autoencoder and convolutional neural network [Текст] / H. Ali [et al.] // SN Applied Sciences. — 2020. — Vol. 2, no. 2. — P. 1—12.

131. *Kaur, S.* Handwritten devanagari character generation using deep convolutional generative adversarial network [Текст] / S. Kaur, K. Verma // Soft Computing: Theories and Applications. — Springer, 2020. — P. 1243—1253.

132. *Manjusha, K.* On developing handwritten character image database for Malayalam language script [Текст] / K. Manjusha, M. A. Kumar, K. Soman // Engineering Science and Technology, an International Journal. — 2019. — Vol. 22, no. 2. — P. 637—645.

133. *Chauhan, V. K.* HCR-Net: A deep learning based script independent handwritten character recognition network [Текст] / V. K. Chauhan, S. Singh, A. Sharma // CoRR. — 2021. — Vol. abs/2108.06663.

134. *Kišš, M.* Brno Mobile OCR Dataset [Текст] / M. Kišš, M. Hradiš, O. Kodym // 2019 International Conference on Document Analysis and Recognition (ICDAR). — 2019. — P. 1352—1357.

135. *Doush, I. A.* Yarmouk Arabic OCR Dataset [Текст] / I. A. Doush, F. AIKhateeb, A. H. Gharibeh // 2018 8th International Conference on Computer Science and Information Technology (CSIT). — 2018. — P. 150—154.

136. *Mathew, M.* Multilingual OCR for Indic Scripts [Текст] / M. Mathew, A. K. Singh, C. V. Jawahar // 2016 12th IAPR Workshop on Document Analysis Systems (DAS). — 2016. — P. 186—191.

137. A Japanese OCR post-processing approach based on dictionary matching [Текст] / C.-y. Guo [et al.] // 2013 International Conference on Wavelet Analysis and Pattern Recognition. — 2013. — P. 22—26.

138. *Kissos, I.* OCR Error Correction Using Character Correction and Feature-Based Word Classification [Текст] / I. Kissos, N. Dershowitz // 2016 12th IAPR Workshop on Document Analysis Systems (DAS). — 2016. — P. 198—203.

139. Statistical learning for OCR error correction [Текст] / J. Mei [et al.] // Information Processing & Management. — 2018. — Vol. 54, no. 6. — P. 874—887.

140. *Bassil, Y.* OCR Post-Processing Error Correction Algorithm using Google Online Spelling Suggestion [Текст] / Y. Bassil, M. Alwani // CoRR. — 2012. — Vol. arXiv/1204.0191.

141. *Eutamene, A.* Ontologies and Bigram-based approach for Isolated Non-word Errors Correction in OCR System [Текст] / A. Eutamene, M. K. Kholladi, H. Belhadef // International Journal of Electrical and Computer Engineering (IJECE). — 2015. — Vol. 5, no. 6. — P. 1458—1467.

142. Lexicographical-Based Order for Post-OCR Correction of Named Entities [Текст] / A. Jean-Caurant [et al.] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 01. — 2017. — P. 1192—1197.

143. Trigram-based algorithms for OCR result correction [Текст] / K. Bulatov [et al.] // Ninth International Conference on Machine Vision (ICMV 2016). Vol. 10341 / ed. by A. Verikas [et al.]. — International Society for Optics, Photonics. SPIE, 2017. — 103410O.

144. *Fonseca Cacho, J. R.* OCR Post Processing Using Support Vector Machines [Текст] / J. R. Fonseca Cacho, K. Taghva // Intelligent Computing / ed. by K. Arai, S. Kapoor, R. Bhatia. — Cham : Springer International Publishing, 2020. — P. 694—713.

145. *Bouchaffra, D.* Postprocessing of recognized strings using nonstationary Markovian models [Текст] / D. Bouchaffra, V. Govindaraju, S. Srihari // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1999. — Vol. 21, no. 10. — P. 990—999.

146. Sub-Word Embeddings for OCR Corrections in Highly Fusional Indic Languages [Текст] / R. Saluja [et al.] // 2019 International Conference on Document Analysis and Recognition (ICDAR). — 2019. — P. 160—165.

147. OCR Post-processing Using Weighted Finite-State Transducers [Текст] / R. Llobet [et al.] // 2010 20th International Conference on Pattern Recognition. — 2010. — P. 2021—2024.

148. *Булатов, К. Б.* Универсальный алгоритм пост-обработки результатов распознавания на основе проверяющих грамматик [Текст] / К. Б. Булатов, Д. П. Николаев, В. В. Постников // Труды ИСА РАН. — 2015. — Т. 65, № 4. — С. 68—73.

149. Approach to recognition of flexible form for credit card expiration date recognition as example [Текст] / A. Sheshkus [et al.] // Eighth International Conference on Machine Vision (ICMV 2015). Vol. 9875 / ed. by A. Verikas, P. Radeva, D. Nikolaev. — International Society for Optics, Photonics. SPIE, 2015. — 98750R.

150. *Wang, K.* Word Spotting in the Wild [Текст] / K. Wang, S. Belongie // Computer Vision – ECCV 2010 / ed. by K. Daniilidis, P. Maragos, N. Paragios. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2010. — P. 591—604.

151. *Epshtein, B.* Detecting text in natural scenes with stroke width transform [Текст] / B. Epshtein, E. Ofek, Y. Wexler // 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. — 2010. — P. 2963—2970.

152. *Felzenszwalb, P. F.* Dynamic Programming and Graph Algorithms in Computer Vision [Текст] / P. F. Felzenszwalb, R. Zabih // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2011. — Vol. 33, no. 4. — P. 721—740.

153. *Povolotskiy, M. A.* Dynamic programming approach to template-based OCR [Текст] / M. A. Povolotskiy, D. V. Tropin // Eleventh International Conference on Machine Vision (ICMV 2018). Vol. 11041 / ed. by A. Verikas [et al.]. — International Society for Optics, Photonics. SPIE, 2019. — 110411T.

154. *Jain, R.* Localized document image change detection [Текст] / R. Jain, D. Doermann // 2015 13th International Conference on Document Analysis and Recognition (ICDAR). — 2015. — P. 786—790.

155. *Lopresti, D. P.* A Comparison of Text-Based Methods for Detecting Duplication in Scanned Document Databases [Текст] / D. P. Lopresti // Information Retrieval. — 2001. — Vol. 4. — P. 153—173.

156. Fast document image comparison in multilingual corpus without OCR [Текст] / Y. Lin [et al.] // Multimedia Systems. — 2017. — Vol. 23. — P. 315—324.

157. *Eglin, V.* Document page similarity based on layout visual saliency: application to query by example and document classification [Текст] / V. Eglin, S. Bres // Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. — 2003. — P. 1208—1212.

158. *Liu, L.* Near-duplicate document image matching: A graphical perspective [Текст] / L. Liu, Y. Lu, C. Y. Suen // Pattern Recognition. — 2014. — Vol. 47, no. 4. — P. 1653—1663.

159. Detecting near-duplicate document images using interest point matching [Текст] / S. Vitaladevuni [et al.] // Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012). — 2012. — P. 347—350.

160. *Ahmed, A. G. H.* Forgery Detection Based on Intrinsic Document Contents [Текст] / A. G. H. Ahmed, F. Shafait // 2014 11th IAPR International Workshop on Document Analysis Systems. — 2014. — P. 252—256.

161. *Beusekom, J. van.* Document Signature Using Intrinsic Features for Counterfeit Detection [Текст] / J. van Beusekom, F. Shafait, T. M. Breuel // Computational Forensics / ed. by S. N. Srihari, K. Franke. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2008. — P. 47—57.

162. A dataset for forgery detection and spotting in document images [Текст] / N. Sidere [et al.] // 2017 Seventh International Conference on Emerging Security Technologies (EST). — 2017. — P. 26—31.

163. ICAR: Identity Card Automatic Reader [Текст] / J. Llados [et al.] // Proceedings of Sixth International Conference on Document Analysis and Recognition. — 2001. — P. 470—474.

164. Design of an Optical Character Recognition System for Camera-based Handheld Devices [Текст] / A. F. Mollah [et al.] // Int'l J. of Computer Science Issues. — 2011. — Vol. 8, no. 4. — P. 283—289.

165. *Ryan, M.* An Examination of Character Recognition on ID card using Template Matching Approach [Текст] / M. Ryan, N. Hanafiah // Procedia Computer Science. — 2015. — Vol. 59. — P. 520—529.

166. Indonesian ID Card Recognition using Convolutional Neural Networks [Текст] / M. O. Pratama [et al.] // 2018 5th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). — 2018. — P. 178—181.

167. Citizen Id Card Detection using Image Processing and Optical Character Recognition [Текст] / W. Satyawan [et al.] // Journal of Physics: Conference Series. — 2019. — Vol. 1235. — P. 012049.

168. *Viet, H. T.* A Robust End-To-End Information Extraction System for Vietnamese Identity Cards [Текст] / H. T. Viet, Q. Hieu Dang, T. A. Vu // 2019 6th NAFOSTED Conference on Information and Computer Science (NICS). — 2019. — P. 483—488.

169. *Thanh, T. N. T.* A Method for Segmentation of Vietnamese Identification Card Text Fields [Текст] / T. N. T. Thanh, K. N. Trong // International Journal of Advanced Computer Science and Applications. — 2019. — Vol. 10, no. 10.

170. Mobilenetv2: Inverted residuals and linear bottlenecks [Текст] / M. Sandler [et al.] // Proceedings of the IEEE conference on computer vision and pattern recognition. — 2018. — P. 4510—4520.

171. *Guo, Q.* Attention OCR [Текст] / Q. Guo, Y. Deng. — 2017. — URL: https://github.com/da03/Attention-OCR (дата обр. 06.11.2022).

172. *Xu, J.* A System to Localize and Recognize Texts in Oriented ID Card Images [Текст] / J. Xu, X. Wu // 2018 IEEE International Conference on Progress in Informatics and Computing (PIC). — 2018. — P. 149—153.

173. Identity authentication on mobile devices using face verification and ID image recognition [Текст] / X. Wu [et al.] // Procedia Computer Science. — 2019. — Vol. 162. — P. 932—939.

174. *Fang, X.* ID card identification system based on image recognition [Текст] / X. Fang, X. Fu, X. Xu // 2017 12th IEEE Conference on Industrial Electronics and Applications (ICIEA). — 2017. — P. 1488—1492.

175. *Ngoc, M. O. V.* Saliency-Based Detection of Identy Documents Captured by Smartphones [Текст] / M. O. V. Ngoc, J. Fabrizio, T. Géraud // 2018 13th IAPR International Workshop on Document Analysis Systems (DAS). — 2018. — P. 387—392.

176. SmartDoc 2017 Video Capture: Mobile Document Acquisition in Video Mode [Текст] / J. Chazalon [et al.] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 04. — 2017. — P. 11—16.

177. *Ôn Vũ Ngoc, M.* Document Detection in Videos Captured by Smartphones using a Saliency-Based Method [Текст] / M. Ôn Vũ Ngoc, J. Fabrizio, T. Géraud // 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW). Vol. 4. — 2019. — P. 19—24.

178. You Only Recognize Once: Towards Fast Video Text Spotting [Текст] / Z. Cheng [et al.] // Proceedings of the 27th ACM International Conference on Multimedia. — New York, NY, USA : Association for Computing Machinery, 2019. — P. 855—863.

179. HighRes-net: Multi-Frame Super-Resolution by Recursive Fusion [Электронный ресурс] / M. Deudon [и др.]. — 2020. — URL: https://openreview.net/forum?id=HJxJ2h4tPr (дата обр. 06.11.2022).

180. A Video Deblurring Algorithm Based on Motion Vector and An Encorder-Decoder Network [Текст] / S. Zhang [et al.] // IEEE Access. — 2019. — Vol. 7. — P. 86778—86788.

181. AdderNet and its Minimalist Hardware Design for Energy-Efficient Artificial Intelligence [Текст] / Y. Wang [et al.] // CoRR. — 2021. — Vol. arXiv/2101.10015v2.

182. ShiftAddNet: A Hardware-Inspired Deep Network [Текст] / H. You [et al.] // CoRR. — 2020. — Vol. arXiv/2010.12785.

183. Kernel Based Progressive Distillation for Adder Neural Networks [Текст] / Y. Xu [et al.] // Proceedings of the 34th International Conference on Neural Information Processing Systems. — Vancouver, BC, Canada : Curran Associates Inc., 2020. — (NIPS'20). — Art. No. 1033.

184. DeepShift: Towards Multiplication-Less Neural Networks [Текст] / M. Elhoushi [et al.] // 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). — 2021. — P. 2359—2368.

185. Training Quantized Neural Networks With a Full-Precision Auxiliary Module [Текст] / B. Zhuang [et al.] // 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2020. — P. 1485—1494.

186. Simulate-the-Hardware: Training Accurate Binarized Neural Networks for Low-Precision Neural Accelerators [Текст] / J. Li [et al.] // Proceedings of the 24th Asia and South Pacific Design Automation Conference. — Tokyo, Japan : Association for Computing Machinery, 2019. — P. 323—328. — (ASPDAC '19).

187. Hybrid 8-Bit Floating Point (HFP8) Training and Inference for Deep Neural Networks [Текст] / X. Sun [et al.] // Proceedings of the 33rd International Conference on Neural Information Processing Systems. — Red Hook, NY, USA : Curran Associates Inc., 2019.

188. Stable Low-Rank Tensor Decomposition for Compression of Convolutional Neural Network [Текст] / A.-H. Phan [et al.] // Computer Vision – ECCV 2020 / ed. by A. Vedaldi [et al.]. — Cham : Springer International Publishing, 2020. — P. 522—539.

189. *Avoine, G.* ePassport: Securing International Contacts with Contactless Chips [Текст] / G. Avoine, K. Kalach, J.-J. Quisquater // Financial Cryptography and Data Security / ed. by G. Tsudik. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2008. — P. 141—155.

190. Национальный стандарт РФ ГОСТ Р ИСО/МЭК 7810-2015 «Карты идентификационные. Физические характеристики» [Текст]. — М. : Стандартинформ, 2018. — 16 с.

191. ISO/IEC 7810:2003 Identification cards – Physical characteristics [Электронный ресурс]. — 2003. — URL: https://www.iso.org/standard/31432.html (дата обр. 06.11.2022).

192. *European Union, C. of the.* PRADO – Public Register of Authentic identity and travel Documents Online [Электронный ресурс] / C. of the European Union. — URL: https://www.consilium.europa.eu/prado (дата обр. 06.11.2022).

193. AAMVA DL/ID Card Design Standard (2020) [Электронный ресурс]. — URL: https://www.aamva.org/assets/aamva-dl-id-card-design-standard-(2020) (дата обр. 06.11.2022).

194. *Конушин, А.* Геометрия камеры и структура движения [Электронный ресурс] / А. Конушин. — 2012. — URL: https://www.graphicon.ru/ru/courses/vision2-2011/lecture06 (дата обр. 06.11.2022).

195. Image noise [Электронный ресурс]. — URL: http://en.wikipedia.org/wiki/Image_noise (дата обр. 06.11.2022).

196. *Тропченко, А. Ю.* Методы сжатия изображений, аудиосигналов и видео. Учебное пособие по дисциплине «Теоретическая информатика» [Текст] / А. Ю. Тропченко, А. А. Тропченко. — Санкт-Петербург, 2009. — 108 с.

197. Compression artifact [Электронный ресурс]. — URL: https://en.wikipedia.org/wiki/Compression_artifact (дата обр. 06.11.2022).

198. *Hartley, R.* Multiple View Geometry in Computer Vision [Текст] / R. Hartley, A. Zisserman. — 2-е изд. — Cambridge University Press, 2004.

199. Алгоритмы поиска границ печатных символов, используемые при оптическом распознавании символов [Текст] / В. Л. Арлазаров [и др.] // ИТиВС / под ред. Ю. С. Попков. — Адрес: 119333, г. Москва, ул. Вавилова, д.44, кор.2, 2004. — № 4. — С. 59—70.

200. Способы защиты документов [Электронный ресурс]. — 2011. — URL: http://www.bnti.ru/showart.asp?aid=940&lvl=01.03.05 (дата обр. 06.11.2022).

201. *Skoryukina, N.* Fast Method of ID Documents Location and Type Identification for Mobile and Server Application [Текст] / N. Skoryukina, V. Arlazarov, D. Nikolaev // 2019 International Conference on Document Analysis and Recognition (ICDAR). — 2019. — P. 850—857.

202. Smart ID Engine – распознавание удостоверений личности [Электронный ресурс]. — 2023. — URL: https://smartengines.ru/smart-idreader (дата обр. 16.02.2023).

203. *Lowe, D.* Object recognition from local scale-invariant features [Текст] / D. Lowe // Proceedings of the Seventh IEEE International Conference on Computer Vision. Vol. 2. — 1999. — P. 1150—1157.

204. Receptive Fields Selection for Binary Feature Description [Текст] / B. Fan [et al.] // IEEE Transactions on Image Processing. — 2014. — Vol. 23, no. 6. — P. 2583—2595.

205. BEBLID: Boosted efficient binary local image descriptor [Текст] / I. Suárez [et al.] // Pattern Recognition Letters. — 2020. — Vol. 133. — P. 366—372.

206. *Trzcinski, T.* Learning Image Descriptors with Boosting [Текст] / T. Trzcinski, M. Christoudias, V. Lepetit // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2015. — Vol. 37, no. 3. — P. 597—610.

207. MIDV-500: A Dataset for Identity Document Analysis and Recognition on Mobile Devices in Video Stream [Текст] / V. V. Arlazarov [et al.] // Computer Optics / ed. by V. A. Soyfer. — 151, Molodogvardeyskaya street, Samara, 443001, 2019. — Vol. 43, no. 5. — P. 818—824.

208. *Bulatov, K.* MIDV-2019: challenges of the modern mobile-based document OCR [Текст] / K. Bulatov, D. Matalov, V. V. Arlazarov // Twelfth International Conference on Machine Vision (ICMV 2019). Vol. 11433 / ed. by W. Osten, D. P. Nikolaev. — International Society for Optics, Photonics. SPIE, 2020. — 114332N.

209. *Fischler, M. A.* Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography [Текст] / M. A. Fischler, R. C. Bolles // Commun. ACM. — New York, NY, USA, 1981. — Vol. 24, no. 6. — P. 381—395.

210. Viability of Viola-Jones method for the problem of image classification [Текст] / A. Sheshkus [et al.] // Eleventh International Conference on Machine Vision (ICMV 2018). Vol. 11041 / ed. by A. Verikas [et al.]. — International Society for Optics, Photonics. SPIE, 2019. — 110410E.

211. *Kuhn, H. W.* Nonlinear programming [Текст] / H. W. Kuhn, A. W. Tucker // Proceedings of 2nd Berkeley Symposium. Berkeley: University of California Press. — 1951. — P. 481—492.

212. *Сараев, А. А.* Выделение графических примитивов для анализа структуры документа на примере локализации печатей [Текст] / А. А. Сараев, Д. П. Николаев // ИТиС 2012. — ИППИ РАН, 2012. — С. 371—376.

213. *Арлазаров, В. В.* Локализация образа печати на документе, удостоверяющем личность, методом машинного обучения [Текст] / В. В. Арлазаров, Д. П. Маталов, С. А. Усилин // Труды ИСА РАН / под ред. Ю. С. Попков. — 119312, г. Москва, проспект 60-летия Октября, д.9, к.501, 2018. — Т. 68, Спецвыпуск № S1. — С. 158—166.

214. Smart IDReader: Document Recognition in Video Stream [Текст] / K. Bulatov [et al.] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 06. — 2017. — P. 39—44.

215. Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видео камер для распознавания документов [Текст] / В. В. Арлазаров [и др.] // ИТиВС / под ред. Ю. С. Попков. — Адрес: 119333, г. Москва, ул. Вавилова, д.44, кор.2, 2014. — № 3. — С. 71—81.

216. Gradient-based learning applied to document recognition [Текст] / Y. Lecun [et al.] // Proceedings of the IEEE. — 1998. — Vol. 86, no. 11. — P. 2278—2324.

217. *Krizhevsky, A.* ImageNet Classification with Deep Convolutional Neural Networks [Текст] / A. Krizhevsky, I. Sutskever, G. E. Hinton // Commun. ACM. — New York, NY, USA, 2017. — Vol. 60, no. 6. — P. 84—90.

218. DeepFace: Closing the Gap to Human-Level Performance in Face Verification [Текст] / Y. Taigman [et al.] // 2014 IEEE Conference on Computer Vision and Pattern Recognition. — 2014. — P. 1701—1708.

219. *Moosavi-Dezfooli, S.-M.* DeepFool: a simple and accurate method to fool deep neural networks [Текст] / S.-M. Moosavi-Dezfooli, A. Fawzi, P. Frossard // CoRR. — 2016. — Vol. arXiv/1511.04599.

220. The Limitations of Deep Learning in Adversarial Settings [Текст] / N. Papernot [et al.] // 2016 IEEE European Symposium on Security and Privacy (EuroS&P). — 2016. — P. 372—387.

221. *Su, J.* One Pixel Attack for Fooling Deep Neural Networks [Текст] / J. Su, D. V. Vargas, K. Sakurai // IEEE Transactions on Evolutionary Computation. — 2019. — Vol. 23, no. 5. — P. 828—841.

222. *Park, S. C.* Super-resolution image reconstruction: a technical overview [Текст] / S. C. Park, M. K. Park, M. G. Kang // IEEE Signal Processing Magazine. — 2003. — Vol. 20, no. 3. — P. 21—36.

223. *Арлазаров, В. В.* Определение достоверности результатов распознавания символа в системе Cognitive Forms [Текст] / В. В. Арлазаров, В. М. Кляцкин // Труды ИСА РАН / под ред. Ю. С. Попков. — 119312, г. Москва, проспект 60-летия Октября, д.9, к.501, 2004. — Т. 5. — С. 16—30.

224. A Man-Machine Cooperating System Based on the Generalized Reject Model [Текст] / S. Kimura [et al.] // 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR). Vol. 01. — 2017. — P. 1324—1329.

225. Quality based frame selection for video face recognition [Текст] / K. Anantharajah [et al.] // 2012 6th International Conference on Signal Processing and Communication Systems. — 2012. — P. 1—5.

226. *Haris, M.* Recurrent Back-Projection Network for Video Super-Resolution [Текст] / M. Haris, G. Shakhnarovich, N. Ukita // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2019. — P. 3892—3901.

227. *Mehregan, K.* Super-resolution of license-plates using frames of low-resolution video [Текст] / K. Mehregan, A. Ahmadyfard, H. Khosravi // 2019 5th Iranian Conference on Signal Processing and Intelligent Systems (ICSPIS). — 2019. — P. 1—6.

228. *Merino-Gracia, C.* Real-time text tracking in natural scenes [Текст] / C. Merino-Gracia, M. Mirmehdi // IET Computer Vision. — 2014. — Vol. 8, no. 6. — P. 670—681.

229. *Мясников, В. В.* Исследование зависимости точности одновременной реконструкции сцены и позиционирования камеры от погрешностей, вносимых датчиками мобильного устройства [Текст] / В. В. Мясников, Е. А. Дмитриев // Компьютерная оптика. — 2019. — Т. 43, № 3. — С. 492—503.

230. *Polikar, R.* Ensemble based systems in decision making [Текст] / R. Polikar // IEEE Circuits and Systems Magazine. — 2006. — Vol. 6, no. 3. — P. 21—45.

231. On combining classifiers [Текст] / J. Kittler [et al.] // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 1998. — Vol. 20, no. 3. — P. 226—239.

232. Оценка качества входных изображений в системах распознавания видеопотока [Текст] / Т. С. Чернов [и др.] // ИТиВС / под ред. П. Ю. Соломонович. — Адрес: 119333, г. Москва, ул. Вавилова, д.44, кор.2, 2017. — № 4. — С. 71—82.

233. *Bulatov, K.* Reducing overconfidence in neural networks by dynamic variation of recognizer relevance [Текст] / K. Bulatov, D. Polevoy // ECMS 2015. — 2015. — P. 488—491.

234. *Fiscus, J.* A post-processing system to yield reduced word error rates: Recognizer Output Voting Error Reduction (ROVER) [Текст] / J. Fiscus // 1997 IEEE Workshop on Automatic Speech Recognition and Understanding Proceedings. — 1997. — P. 347—354.

235. *Yujian, L.* A Normalized Levenshtein Distance Metric [Текст] / L. Yujian, L. Bo // IEEE Transactions on Pattern Analysis and Machine Intelligence. — 2007. — Vol. 29, no. 6. — P. 1091—1095.

236. *Chernyshova, Y. S.* Two-Step CNN Framework for Text Line Recognition in Camera-Captured Images [Текст] / Y. S. Chernyshova, A. V. Sheshkus, V. V. Arlazarov // IEEE Access. — 2020. — Vol. 8. — P. 32587—32600.

237. *Hartl, A.* Real-time Detection and Recognition of Machine-Readable Zones with Mobile Devices [Текст] / A. Hartl, C. Arth, D. Schmalstieg // Proceedings of the 10th International Conference on Computer Vision Theory and Applications - Volume 1: VISAPP, (VISIGRAPP 2015). — INSTICC. SciTePress, 2015. — P. 79—87.

238. *Арлазаров, В. Л.* Накопительные контексты в задаче распознавания [Текст] / В. Л. Арлазаров, А. Е. Марченко, Д. Л. Шоломов // Труды ИСА РАН / под ред. Ю. С. Попков. — 119312, г. Москва, проспект 60-летия Октября, д.9, к.501, 2014. — Т. 64, № 4. — С. 64—72.

239. *Булатов, К.* Выбор оптимальной стратегии комбинирования покадровых результатов распознавания символа в видеопотоке [Текст] / К. Булатов // ИТиВС / под ред. Ю. С. Попков. — Адрес: 119333, г. Москва, ул. Вавилова, д.44, кор.2, 2017. — № 3. — С. 45—55.

240. *Ricci, V.* Fitting ditributions with R [Электронный ресурс] / V. Ricci. — 2005. — URL: https://cran.r-project.org/doc/contrib/Ricci-distributions-en.pdf (дата обр. 06.11.2022).

241. *Ongaro, A.* A generalization of the Dirichlet distribution [Текст] / A. Ongaro, S. Migliorati // Journal of Multivariate Analysis. — 2013. — Vol. 114. — P. 412—426.

242. *Connor, R. J.* Concepts of Independence for Proportions with a Generalization of the Dirichlet Distribution [Текст] / R. J. Connor, J. E. Mosimann // Journal of the American Statistical Association. — 1969. — Vol. 64, no. 325. — P. 194—206.

243. *Ng, K. W.* Dirichlet and Related Distributions: Theory, Methods and Applications [Текст] / K. W. Ng, G.-L. Tian, M.-L. Tang. — 2011.

244. *Elfadaly, F. G.* Eliciting Dirichlet and Connor–Mosimann prior distributions for multinomial models [Текст] / F. G. Elfadaly, P. H. Garthwaite // TEST. — 2013. — Vol. 22. — P. 628—646.

245. *Fang, K. W.* Symmetric Multivariate and Related Distributions [Текст] / K. W. Fang. — Chapman & Hall, 2017. — 230 p.

246. *Ronning, G.* Maximum likelihood estimation of dirichlet distributions [Текст] / G. Ronning // Journal of Statistical Computation and Simulation. — 1989. — Vol. 32, no. 4. — P. 215—221.

247. *Robitzsch, A.* sirt: Supplementary Item Response Theory Models [Электронный ресурс] / A. Robitzsch. — URL: https://cran.r-project.org/web/packages/sirt/index.html (дата обр. 06.11.2022).

248. *Migliorati, S.* A structured Dirichlet mixture model for compositional data: inferential and applicative issues [Текст] / S. Migliorati, A. Ongaro, G. S. Monti // Statistics and Computing. — 2017. — Vol. 27. — P. 963—983.

249. *Migliorati, S.* FlexDir: Tools to Work with the Flexible Dirichlet Distribution [Электронный ресурс] / S. Migliorati, A. M. D. Brisco, M. Vestrucci. — URL: https://cran.r-project.org/web/packages/FlexDir/index.html (дата обр. 06.11.2022).

250. *Li, Y.* Goodness-of-Fit tests for Dirichlet Distributions with Applications: PhD Thesis [Текст] / Y. Li. — 2015.

251. *Stephens, M. A.* Goodness of Fit, Anderson–Darling Test of [Текст] / M. A. Stephens // Encyclopedia of Statistical Sciences. — John Wiley & Sons, Ltd, 2006.

252. Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход [Текст] / Б. Ю. Лемешко [и др.]. — НИЦ ИНФРА-М, 2015. — 890 с.

253. *Большев, Л.* Таблицы математической статистики [Текст] / Л. Большев, Н. Смирнов. — М. : Наука, 1983. — 416 с.

254. *Bechhofer, R. E.* Truncation of the bechhofer-kiefer-sobel sequential procedure for selecting the multinomial event which has the largest probability (ii): extended tables and an improved procedure [Текст] / R. E. Bechhofer, D. M. Goldsman // Communications in Statistics - Simulation and Computation. — 1986. — Vol. 15, no. 3. — P. 829—851.

255. *Huang, W.-T.* Selecting the Best Population with Two Controls: An Empirical Bayes Approach [Текст] / W.-T. Huang, Y.-T. Lai // Advances in Ranking and Selection, Multiple Comparisons, and Reliability: Methodology and Applications / ed. by N. Balakrishnan, H. N. Nagaraja, N. Kannan. — Boston, MA : Birkhäuser Boston, 2005. — P. 133—142.

256. *Kareev, I.* Lower bounds for the expected sample size of sequential procedures for selecting and ranking of binomial and Poisson populations [Текст] / I. Kareev // Lobachevskii Journal of Mathematics. — 2016. — Vol. 37. — P. 455—465.

257. *He, J.* Posterior probability calculation procedure for recognition rate comparison [Текст] / J. He, Q. Fu // Journal of Systems Engineering and Electronics. — 2016. — Vol. 27, no. 3. — P. 700—711.

258. *Malloy, M. L.* The Sample Complexity of Search over Multiple Populations [Текст] / M. L. Malloy, G. Tang, R. D. Nowak // CoRR. — 2013. — Vol. arXiv/1209.1380.

259. *Limonova, E.* Improving neural network performance on SIMD architectures [Текст] / E. Limonova, D. Ilin, D. Nikolaev // Eighth International Conference on Machine Vision (ICMV 2015). Vol. 9875 / ed. by A. Verikas, P. Radeva, D. Nikolaev. — International Society for Optics, Photonics. SPIE, 2015. — P. 98750L.

260. *Mamalet, F.* Real-Time Video Convolutional Face Finder on Embedded Platforms [Текст] / F. Mamalet, S. Roux, C. Garcia // EURASIP J. Embedded Syst. — London, GBR, 2007. — Vol. 2007, no. 1. — P. 22.

261. *Kuznetsova, E.* Viola-Jones based hybrid framework for real-time object detection in multispectral images [Текст] / E. Kuznetsova, E. Shvets, D. Nikolaev // Eighth International Conference on Machine Vision (ICMV 2015). Vol. 9875 / ed. by A. Verikas, P. Radeva, D. Nikolaev. — International Society for Optics, Photonics. SPIE, 2015. — 98750N.

262. *Mastov, A.* Application of Random Ferns for non-planar object detection [Текст] / A. Mastov, I. Konovalenko, A. Grigoryev // Eighth International Conference on Machine Vision (ICMV 2015). Vol. 9875 / ed. by A. Verikas, P. Radeva, D. Nikolaev. — International Society for Optics, Photonics. SPIE, 2015. — P. 98750M.

263. *Kopenkov, V. N.* Detection and tracking of vehicles based on the videoregistration information [Текст] / V. N. Kopenkov, V. V. Myasnikov // WSCG 2015: poster papers proceedings: 23rd International Conference in Central Europe on Computer Graphics, Visualization and Computer Visionin co-operation with EUROGRAPHICS Association. — Václav Skala - UNION Agency, 2015. — P. 65—68.

264. *Patterson, D.* Computer organization and design (4th edition) [Текст] / D. Patterson, J. Hennessy. — Elsevier, 2009. — 457 p.

265. Dynamic Trace-Based Analysis of Vectorization Potential of Applications [Текст] / J. Holewinski [et al.] // SIGPLAN Not. — New York, NY, USA, 2012. — Vol. 47, no. 6. — P. 371—382.

266. *Gonzalez, R. C.* Digital Image Processing, 3rd edition [Текст] / R. C. Gonzalez, R. E. Woods. — Prentice-Hall, 2008. — 954 p.

267. Coding for Neon – Part 5: Rearranging Vectors [Электронный ресурс]. — URL: https://community.arm.com/arm-community-blogs/b/architectures-and-processors-blog/posts/coding-for-neon---part-5-rearranging-vectors (дата обр. 08.11.2022).

268. *Zekri, A. S.* Enhancing the matrix transpose operation using Intel AVX instruction set extension [Текст] / A. S. Zekri // International Journal of Computer Science and Information Technology (IJCSIT). — 2014. — Vol. 6, no. 3. — P. 67—78.

269. ARM non-confidential technical publications: Documentation [Электронный ресурс]. — URL: https://developer.arm.com/documentation/ (дата обр. 08.11.2022).

270. *Gevorkian, D. Z.* Improving Gil-Werman Algorithm for Running Min and Max Filters [Текст] / D. Z. Gevorkian, J. T. Astola, S. M. Atourian // IEEE Trans. Pattern Anal. Mach. Intell. — USA, 1997. — Vol. 19, no. 5. — P. 526—529.

271. *van Herk, M.* A fast algorithm for local minimum and maximum filters on rectangular and octagonal kernels [Текст] / M. van Herk // Pattern Recognition Letters. — 1992. — Vol. 13, no. 7. — P. 517—521.

272. *Галушка, В. В.* Формирование обучающей выборки при использовании искусственных нейронных сетей в задачах поиска ошибок баз данных [Электронный ресурс] / В. В. Галушка, В. А. Фатхи. — 2013. — URL: http://www.ivdon.ru/magazine/archive/n2y2013/1597 (дата обр. 08.11.2022) ; Инженерный вестник Дона, №2.

273. *Яглом, А. М.* Вероятность и информация [Текст] / А. М. Яглом, И. М. Яглом. — М. : Наука, 1973. — 513 с.

274. A Novel Comprehensive Database for Arabic Off-Line Handwriting Recognition [Текст] / H. Alamri [et al.] // Proc. of the 11th Int. Conference on Frontiers in Handwriting Recognition (ICFHR'2008). — 2008. — P. 664—669.

275. *Su, T.* HIT-MW dataset for offline Chinese handwritten text recognition [Текст] / T. Su, T. Zhang, D. Guan // 10th IWFHR. — 2006.

276. *Савчинский, Б. Д.* Поиск размеров эталонов при распознавании текстовых изображений [Текст] / Б. Д. Савчинский, С. А. Олефиренко // Сборник трудов Международного научно-обучающего центра информационных технологий и систем НАН и МОН Украины «Перспективные технологии обучения и учебных центров». — Киев : МННЦИТИС, Вып. 2, 2009. — С. 24—45.

277. *Славин, О. А.* Средства управления базами графических образов символов и их место в системах распознавания [Текст] / О. А. Славин // Сборник трудов ИСА РАН «Развитие безбумажных технологий в организациях». — 1999. — С. 277—289.

278. *Арлазаров, В. В.* Cognitive forms – система массового ввода структурированных документов [Текст] / В. В. Арлазаров, В. В. Постников, Д. Л. Шоломов // Труды ИСА РАН / под ред. Ю. С. Попков. — 119312, г. Москва, проспект 60-летия Октября, д.9, к.501, 2002. — Т. 1. — С. 35—46.

279. *Левенштейн, В. И.* Двоичные коды с исправлением выпадений, вставок и замещений символов [Текст] / В. И. Левенштейн // Доклады Академий Наук СССР. — 1965. — Т. 163, № 4. — С. 845—848.

280. *Постников, В. В.* Автоматическая идентификация и распознавание структурированных документов [Текст] / В. В. Постников. — 2001. — Дисс. на соискание уч. ст. канд. техн. наук.

281. *Gupta, A.* Synthetic Data for Text Localisation in Natural Images [Текст] / A. Gupta, A. Vedaldi, A. Zisserman // 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). — 2016. — P. 2315—2324.

282. Acquiring Custom OCR System with Minimal Manual Annotation [Текст] / J. Hula [et al.] // 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP). — 2020. — P. 231—236.

283. *Ren, X.* A CNN Based Scene Chinese Text Recognition Algorithm With Synthetic Data Engine [Текст] / X. Ren, K. Chen, J. Sun // CoRR. — 2016. — Vol. arXiv/1604.01891.

284. *Chernyshova, Y. S.* Generation method of synthetic training data for mobile OCR system [Текст] / Y. S. Chernyshova, A. V. Gayer, A. V. Sheshkus // Tenth International Conference on Machine Vision (ICMV 2017). Vol. 10696 / ed. by A. Verikas [et al.]. — International Society for Optics, Photonics. SPIE, 2018. — 106962G.

285. *Alonso, E.* Adversarial Generation of Handwritten Text Images Conditioned on Sequences [Текст] / E. Alonso, B. Moysset, R. Messina // 2019 International Conference on Document Analysis and Recognition (ICDAR). — 2019. — P. 481—486.

286. *Fröhling, L.* Feature-based detection of automated language models: tackling GPT-2, GPT-3 and Grover [Текст] / L. Fröhling, A. Zubiaga // PeerJ Computer Science. — 2021. — Vol. 7. — 10.7717/peerj-cs.443.

287. Character sequence prediction method for training data creation in the task of text recognition [Текст] / P. K. Zlobin [et al.] // Fourteenth International Conference on Machine Vision (ICMV 2021). Vol. 12084 / ed. by W. Osten, D. Nikolaev, J. Zhou. — International Society for Optics, Photonics. SPIE, 2022. — 120840R.

288. *Krizhevsky, A.* Learning Multiple Layers of Features from Tiny Images [Электронный ресурс] / A. Krizhevsky. — 2009. — URL: http://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf (visited on 11/08/2022).

289. *Simard, P.* Best practices for convolutional neural networks applied to visual document analysis [Текст] / P. Simard, D. Steinkraus, J. Platt // Seventh International Conference on Document Analysis and Recognition, 2003. Proceedings. — 2003. — P. 958—963.

290. *Gayer, A. V.* Effective real-time augmentation of training dataset for the neural networks learning [Текст] / A. V. Gayer, Y. S. Chernyshova, A. V. Sheshkus // Eleventh International Conference on Machine Vision (ICMV 2018). Vol. 11041 / ed. by A. Verikas [et al.]. — International Society for Optics, Photonics. SPIE, 2019. — P. 110411I.

291. *Dosovitskiy, A.* Unsupervised feature learning by augmenting single images [Текст] / A. Dosovitskiy, J. T. Springenberg, T. Brox // CoRR. — 2014. — Vol. arXiv/1312.5242.

292. *Skoryukina, N. S.* 2D art recognition in uncontrolled conditions using one-shot learning [Текст] / N. S. Skoryukina, D. P. Nikolaev, V. V. Arlazarov // Eleventh International Conference on Machine Vision (ICMV 2018). Vol. 11041 / ed. by A. Verikas [et al.]. — International Society for Optics, Photonics. SPIE, 2019. — 110412E.

293. Методы аугментации обучающих выборок в задачах классификации изображений [Текст] / С. О. Емельянов [и др.] // Сенсорные системы. — 117485, Москва, Профсоюзная улица, дом 90, 2018. — Т. 32, № 3. — С. 236—245.

294. *Nandedkar, A. V.* SPODS: A Dataset of Color-Official Documents and Detection of Logo, Stamp, and Signature [Текст] / A. V. Nandedkar, J. Mukherjee, S. Sural // Computer Vision, Graphics, and Image Processing / ed. by

S. Mukherjee [et al.]. — Cham : Springer International Publishing, 2017. — P. 219—230.

295. *Matalov, D. P.* Modification of the Viola-Jones approach for the detection of the government seal stamp of the Russian Federation [Текст] / D. P. Matalov, S. A. Usilin, V. V. Arlazarov // Eleventh International Conference on Machine Vision (ICMV 2018). Vol. 11041 / ed. by A. Verikas [et al.]. — International Society for Optics, Photonics. SPIE, 2019. — 110411Y.

296. MIDV-2020: A Comprehensive Benchmark Dataset for Identity Document Analysis [Текст] / K. B. Bulatov [et al.] // Computer Optics. — 2022. — Vol. 46, no. 2. — P. 252—270.

297. MIDV-LAIT: A Challenging Dataset for Recognition of IDs with Perso-Arabic, Thai, and Indian Scripts [Текст] / Y. Chernyshova [et al.] // Document Analysis and Recognition – ICDAR 2021 / ed. by J. Lladós, D. Lopresti, S. Uchida. — Cham : Springer International Publishing, 2021. — P. 258—272.

298. Document Liveness Challenge Dataset (DLC-2021) [Текст] / D. V. Polevoy [et al.] // Journal of Imaging. — 2022. — Vol. 8, no. 7. — Art. No. 181.

299. Wikipedia. Category: Serbian masculine given names [Электронный ресурс]. — URL: https://en.wikipedia.org/wiki/Category:Serbian_masculine_given_names (дата обр. 08.11.2022).

300. Fantasy name generators: Azerbaijani names [Электронный ресурс]. — URL: https://www.fantasynamegenerators.com/azerbaijani-names.php (дата обр. 08.11.2022).

301. Generated Photos [Электронный ресурс]. — URL: https://generated.photos/ (дата обр. 08.11.2022).

302. *Karras, T.* A Style-Based Generator Architecture for Generative Adversarial Networks [Текст] / T. Karras, S. Laine, T. Aila // 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). — 2019. — P. 4396—4405.

303. *Dutta, A.* The VIA Annotation Software for Images, Audio and Video [Текст] / A. Dutta, A. Zisserman // Proceedings of the 27th ACM International Conference on Multimedia. — Nice, France : Association for Computing Machinery, 2019. — P. 2276—2279. — (MM '19).

304. Council Regulation (EC) No 2252/2004 of 13 December 2004 on standards for security features and biometrics in passports and travel documents issued by Member States [Электронный ресурс]. — 2004. — URL: https://eur-lex.europa.eu/eli/reg/2004/2252/oj (дата обр. 13.02.2023).

305. *Каминская, Т. П.* Исследование рельефа пленочных дифракционных оптических элементов [Текст] / Т. П. Каминская, В. В. Попов, А. М. Салецкий // Компьютерная оптика. — 2016. — Т. 40, № 2. — С. 215—224.

306. Stock Images, Royalty-Free Pictures, Illustrations and Videos [Электронный ресурс]. — 2022. — URL: https://www.istockphoto.com (дата обр. 13.02.2023).

307. Центр Тонких Оптических Технологий [Электронный ресурс]. — 2022. — URL: https://center-tot.ru (дата обр. 13.02.2023).

308. Advanced Hough-based method for on-device document localization [Текст] / D. V. Tropin [et al.] // Computer Optics / ed. by V. A. Soyfer. — 151, Molodogvardeyskaya street, Samara, 443001, 2021. — Vol. 45, no. 5. — P. 702—712.

309. Approach for Document Detection by Contours and Contrasts [Текст] / D. V. Tropin [et al.] // 2020 25th International Conference on Pattern Recognition (ICPR). — 2021. — P. 9689—9695.

310. Improved algorithm of ID card detection by a priori knowledge of the document aspect ratio [Текст] / D. V. Tropin [et al.] // Thirteenth International Conference on Machine Vision. Vol. 11605 / ed. by W. Osten, D. P. Nikolaev, J. Zhou. — International Society for Optics, Photonics. SPIE, 2021. — 116051F.

311. Impact of geometrical restrictions in RANSAC sampling on the ID document classification [Текст] / N. Skoryukina [et al.] // Twelfth International Conference on Machine Vision (ICMV 2019). Vol. 11433 / ed. by W. Osten, D. P. Nikolaev. — International Society for Optics, Photonics. SPIE, 2020. — P. 1143306.

312. *Chiron, G.* ID documents matching and localization with multi-hypothesis constraints [Текст] / G. Chiron, N. Ghanmi, A. M. Awal // 2020 25th International Conference on Pattern Recognition (ICPR). — 2021. — P. 3644—3651.

313. *Chiron, G.* Fast End-to-End Deep Learning Identity Document Detection, Classification and Cropping [Текст] / G. Chiron, F. Arrestier, A. M. Awal // Document Analysis and Recognition – ICDAR 2021 / ed. by J. Lladós, D. Lopresti, S. Uchida. — Cham : Springer International Publishing, 2021. — P. 333—347.

314. BusiNet – a Light and Fast Text Detection Network for Business Documents [Текст] / O. Naparstek [et al.] // CoRR. — 2022. — Vol. arXiv/2207.01220.

315. *Sheshkus, A. V.* Transfer of a high-level knowledge in HoughNet neural network [Текст] / A. V. Sheshkus, D. Nikolaev // Twelfth International Conference on Machine Vision (ICMV 2019). Vol. 11433 / ed. by W. Osten, D. P. Nikolaev. — International Society for Optics, Photonics. SPIE, 2020. — P. 1143322.

316. HoughNet: Neural Network Architecture for Vanishing Points Detection [Текст] / A. Sheshkus [et al.] // 2019 International Conference on Document Analysis and Recognition (ICDAR). — 2019. — P. 844—849.

317. Fast projective image rectification for planar objects with Manhattan structure [Текст] / J. Shemiakina [et al.] // Twelfth International Conference on Machine Vision (ICMV 2019). Vol. 11433 / ed. by W. Osten, D. P. Nikolaev. — International Society for Optics, Photonics. SPIE, 2020. — 114331N.

318. *Baniadamdizaj, S.* Localization Using DeepLab in Document Images Taken by Smartphones [Текст] / S. Baniadamdizaj // Digital Interaction and Machine Intelligence / ed. by C. Biele [et al.]. — Cham : Springer International Publishing, 2022. — P. 63—74.

319. *Dizaj, S. B.* A New Image Dataset for Document Corner Localization [Текст] / S. B. Dizaj, M. Soheili, A. Mansouri // 2020 International Conference on Machine Vision and Image Processing (MVIP). — 2020. — P. 1—4.

320. Line detection via a lightweight CNN with a Hough layer [Текст] / L. Teplyakov [et al.] // Thirteenth International Conference on Machine Vision. Vol. 11605 / ed. by W. Osten, D. P. Nikolaev, J. Zhou. — International Society for Optics, Photonics. SPIE, 2021. — 116051B.

321. *Sheshkus, A.* Tiny CNN for feature point description for document analysis: approach and dataset [Текст] / A. Sheshkus, A. Chirvonaya, V. L. Arlazarov // Computer Optics / ed. by V. A. Soyfer. — 151, Molodogvardeyskaya street, Samara, 443001, 2022. — Vol. 46, no. 3. — P. 429—435.

322. RFDoc: Memory Efficient Local Descriptors for ID Documents Localization and Classification [Текст] / D. Matalov [et al.] // Document Analysis and Recognition – ICDAR 2021 / ed. by J. Lladós, D. Lopresti, S. Uchida. — Cham : Springer International Publishing, 2021. — P. 209—224.

323. Bayesian Feature Fusion Using Factor Graph in Reduced Normal Form [Текст] / A. Buonanno [et al.] // Applied Sciences. — 2021. — Vol. 11, no. 4. — Art. No. 1934.

324. Face Detection in Camera Captured Images of Identity Documents Under Challenging Conditions [Текст] / S. Bakkali [et al.] // 2019 International Conference on Document Analysis and Recognition Workshops (ICDARW). Vol. 4. — 2019. — P. 55—60.

325. Towards a unified framework for identity documents analysis and recognition [Текст] / K. B. Bulatov [et al.] // Computer Optics / ed. by V. A. Soyfer. — 151, Molodogvardeyskaya street, Samara, 443001, 2022. — Vol. 46, no. 3. — P. 436—454.

326. Fast Implementation of 4-bit Convolutional Neural Networks for Mobile Devices [Текст] / A. Trusov [et al.] // 2020 25th International Conference on Pattern Recognition (ICPR). — 2021. — P. 9897—9903.

327. MRZ code extraction from visa and passport documents using convolutional neural networks [Текст] / Y. Liu [et al.] // International Journal of Document Analysis and Recognition. — 2022. — Vol. 25. — P. 29—39.

328. Weighted combination of per-frame recognition results for text recognition in a video stream [Текст] / O. O. Petrova [et al.] // Computer Optics / ed. by V. A. Soyfer. — 151, Molodogvardeyskaya street, Samara, 443001, 2021. — Vol. 45, no. 1. — P. 77—89.

329. *Bulatov, K. B.* A Method to Reduce Errors of String Recognition Based on Combination of Several Recognition Results with Per-Character Alternatives [Текст] / K. B. Bulatov // Bulletin of the South Ural State University, Series: Mathematical Modelling, Programming and Computer Software. — 2019. — Vol. 12, no. 3. — P. 74—88.

330. *Chernov, T. S.* Application of dynamic saliency maps to the video stream recognition systems with image quality assessment [Текст] / T. S. Chernov, S. A. Ilyuhin, V. V. Arlazarov // Eleventh International Conference on Machine Vision (ICMV 2018). Vol. 11041 / ed. by A. Verikas [et al.]. — International Society for Optics, Photonics. SPIE, 2019. — 110410T.

331. A Method of Image Quality Assessment for Text Recognition on Camera-Captured and Projectively Distorted Documents [Текст] / J. Shemiakina [et al.] // Mathematics. — 2021. — Vol. 9, no. 17. — Art. No. 2155.

332. *Bulatov, K.* On optimal stopping strategies for text recognition in a video stream as an application of a monotone sequential decision model [Текст] / K. Bulatov, N. Razumnyi, V. V. Arlazarov // International Journal of Document Analysis and Recognition. — 2019. — Vol. 22. — P. 303—314.

333. *Bulatov, K.* Next integrated result modelling for stopping the text field recognition process in a video using a result model with per-character alternatives [Текст] / K. Bulatov, B. Savelyev, V. V. Arlazarov // Twelfth International Conference on Machine Vision (ICMV 2019). Vol. 11433 / ed. by W. Osten, D. P. Nikolaev. — International Society for Optics, Photonics. SPIE, 2020. — P. 114332M.

334. *Bulatov, K.* Determining Optimal Frame Processing Strategies for Real-Time Document Recognition Systems [Текст] / K. Bulatov, V. V. Arlazarov // Document Analysis and Recognition – ICDAR 2021 / ed. by J. Lladós, D. Lopresti, S. Uchida. — Cham : Springer International Publishing, 2021. — P. 273—288.

335. *Bulatov, K.* Fast Approximate Modelling of the Next Combination Result for Stopping the Text Recognition in a Video [Текст] / K. Bulatov, N. Fedotova, V. V. Arlazarov // 2020 25th International Conference on Pattern Recognition (ICPR). — 2021. — P. 239—246.

336. *Polevoy, D. V.* Choosing the best image of the document owner's photograph in the video stream on the mobile device [Текст] / D. V. Polevoy, M. A. Aliev, D. P. Nikolaev // Thirteenth International Conference on Machine Vision. Vol. 11605 / ed. by W. Osten, D. P. Nikolaev, J. Zhou. — International Society for Optics, Photonics. SPIE, 2021. — 116050F.

337. *Al-Ghadi, M.* CheckScan: a reference hashing for identity document quality detection [Текст] / M. Al-Ghadi, P. Gomez-Krämer, J.-C. Burie // Fourteenth International Conference on Machine Vision (ICMV 2021). Vol. 12084 / ed. by W. Osten, D. Nikolaev, J. Zhou. — International Society for Optics, Photonics. SPIE, 2022. — 120840J.

338. *Myasnikov, E.* Detection of Sensitive Textual Information in User Photo Albums on Mobile Devices [Текст] / E. Myasnikov, A. Savchenko // 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON). — 2019. — P. 0384—0390.

339. *Kopeykina, L.* Automatic Privacy Detection in Scanned Document Images Based on Deep Neural Networks [Текст] / L. Kopeykina, A. V. Savchenko // 2019 International Russian Automation Conference (RusAutoCon). — 2019. — P. 1—6.

340. Automated detection of unstructured context-dependent sensitive information using deep learning [Текст] / H. Ahmed [et al.] // Internet of Things. — 2021. — Vol. 16. — P. 100444.

341. Analysis of Financial Payments Text Labels in the Dynamic Client Profile Construction [Текст] / A. Startseva [et al.] // 2020 International Conference on Information Technology and Nanotechnology (ITNT). — 2020. — P. 1—10.

342. Identity Documents Authentication based on Forgery Detection of Guilloche Pattern [Текст] / M. Al-Ghadi [et al.] // CoRR. — 2022. — Vol. arXiv/2206.10989.

343. Hologram Detection for Identity Document Authentication [Текст] / O. Kada [et al.] // Pattern Recognition and Artificial Intelligence / ed. by M. El Yacoubi [et al.]. — Cham : Springer International Publishing, 2022. — P. 346—357.

344. A distortion model-based pre-screening method for document image tampering localization under recapturing attack [Текст] / C. Chen [et al.] // Signal Processing. — 2022. — Vol. 200. — P. 108666.

345. *Арлазаров, В. В.* Анализ использования проблемно-ориентированных пакетов данных в научных исследованиях [Текст] / В. В. Арлазаров // ИТиВС / под ред. Ю. С. Попков. — Адрес: 119333, г. Москва, ул. Вавилова, д.44, кор.2, 2022. — № 3. — С. 10—23.

346. *Ким, А. К.* Микропроцессоры и вычислительные комплексы семейства «Эльбрус» [Текст] / А. К. Ким, В. И. Перекатов, С. Г. Ермаков. — СПб. : Питер, 2013. — 272 с.

347. Российские технологии «Эльбрус» для персональных компьютеров, серверов и суперкомпьютеров [Текст] / А. К. Ким [и др.] // Современные информационные технологии и ИТ-образование. Т. 10. — 2014. — С. 39—50.

348. *Ишин, П. А.* Ускорение вычислений с использованием высокопроизводительных математических и мультимедийных библиотек для архитектуры Эльбрус [Текст] / П. А. Ишин, В. Е. Логинов, П. П. Васильев // Вестник воз- душно-космической обороны. — М., 2015. — Т. 8, № 4. — С. 64—68.

349. Intel oneAPI Threading Building Blocks [Электронный ресурс]. — URL: https://www.intel.com/content/www/us/en/developer/tools/oneapi/onetbb.html (дата обр. 08.11.2022).

350. Система программирования для платформ Эльбрус и МЦСТ-R [Электронный ресурс]. — URL: http://mcst.ru/sdk (дата обр. 08.11.2022).

Приложение А

Акты о внедрении

ООО «НВИАЙ Солюшенс»
ИНН / КПП: 9731001888 / 772501001
ОГРН 1187746473320, 14.05.2018 г.
Юр./ факт. адрес: 115162, г. Москва,
ул. Шухова, 14с9, этаж 2, пом 1 (офис 201)
тел.: +7 (499) 397-87-58
e-mail: info@nvi-solutions.com

АКТ
об использовании (внедрении) результатов диссертационной работы Арлазарова Владимира Викторовича «Мобильное распознавание и его применение к системе ввода идентификационных документов» в ООО «НВИАЙ Солюшенс»

Результаты диссертационной работы «Мобильное распознавание и его применение к системе ввода идентификационных документов» обладают высокой актуальностью и представляют практический интерес для решения задачи ввода идентификационных документов для автоматизации предоставления банковских услуг.

Технология распознавания идентификационных документов для мобильных устройств, разработанная Арлазаровым В.В., позволяют улучшить качество и эффективность предоставления банковских услуг. Данные технологии в составе программного продукта Smart ID Engine от компании ООО «Смарт Энджинс Сервис» внедрены компанией ООО «НВИАЙ Солюшенс» используются в информационных системах и мобильных приложениях в АО «Банк ГПБ» (Газпромбанк) и АО «Банк ДОМ.РФ».

Генеральный директор  **Беляев Филипп Владимирович**

ОБЩЕСТВО С ОГРАНИЧЕННОЙ ОТВЕТСТВЕННОСТЬЮ
«Интек»
Металлистов проспект, д.96,
Санкт-Петербург, Россия, 195221
тел.: (812) 454-54-52, 540-10-04
факс: (812) 540-68-77
E-mail: mail@gk-intek.ru
www.intekspb.ru
ОКПО 59487641, ОГРН 1027810297844,
ИНН/КПП 7826156727 / 783901001
Исх.. № 01/А/23 от 15.02.2023 г.

В Диссертационный совет 24.1.224.01

Справка о внедрении

Настоящим подтверждаем, что результаты диссертационного исследований Арлазарова Владимира Викторовича на тему «Мобильное распознавание и его применение к системе ввода идентификационных документов» обладают высокой актуальностью и представляют практический интерес в системах автоматизированного считывания данных с документов.

Результаты исследований были внедрены в аппаратно-программный комплекс «СЧИТЫВАТЕЛЬ ДОКУМЕНТОВ ПС4-02 ПШНК.468469.009», предназначенный для автоматизированного считывания данных со страницы документа посредством оптического сканирования с высоким разрешением, распознавания текстовых полей и штрих-кодов, считывания данных с бесконтактной RFID микросхемы, проверки достоверности документа.

Аппаратно-программный комплекс «СЧИТЫВАТЕЛЬ ДОКУМЕНТОВ ПС4-02 ПШНК.468469.009» установлен и эксплуатируется во всех отделения УЦ ФНС России (обеспечивает автоматическую проверку подлинности паспортов и ввод персональных данных заявителя при выдаче ЭЦП), отделениях МВД РФ (для целей проверки удостоверяющих личность документов иностранных граждан), аэропорту Шереметьево (для обеспечения работы автоматических пунктов пропуска через государственную границу в терминале С), а также обеспечивает безопасность прохода граждан на территорию всех судебных участков мировых судей г. Москвы и Московской области.

Генеральный директор Г.Г. Головастиков

СКБ Контур

Акционерное общество
«Производственная фирма «СКБ Контур»
(АО «ПФ «СКБ Контур»)

Народной Воли ул., 19а, Екатеринбург, 620144
тел. (343) 228-14-40, 228-14-41, факс (343) 228-14-43
info@skbkontur.ru
www.kontur.ru
ОГРН 1026605606620
ИНН/КПП 6663003127/997750001
р/счет № 40702810138030000017
в филиале «Екатеринбургский» АО «АЛЬФА-БАНК»
кор/счет № 30101810100000000964
БИК 046577964

17.02.2023 № 15018/АУП

На № _____ от _____

По месту требования

Об использовании (внедрении) результатов диссертационной работы Арлазарова Владимира Викторовича «Мобильное распознавание и его применение к системе ввода идентификационных документов»

Результаты диссертационной работы «Мобильное распознавание и его применение к системе ввода идентификационных документов» актуальны и представляют практический интерес для решения задачи ввода удостоверяющих личность документов.

Разработанные Арлазаровым В.В. технологии позволяют повысить качество обслуживания клиентов в процессах, требующих предоставления паспортных данных. В составе программных продуктов ООО «Смарт Энджинс Сервис» эти технологии используются в информационных системах АО «ПФ «СКБ Контур» для оформления электронной подписи и регистрации гостей в организациях индустрии гостеприимства.

Генеральный директор М. Ю. Сродных

Малкиев Максим Витальевич
m.malkiev@skbkontur.ru
(343)228-14-40

Назначение: по месту требования

АКТ

об использовании (внедрении) результатов диссертационной работы Арлазарова Владимира Викторовича «Мобильное распознавание и его применение к системе ввода идентификационных документов» в АО «Альфа-Банк»

Результаты диссертационной работы «Мобильное распознавание и его применение к системе ввода идентификационных документов» обладают высокой актуальностью и представляют практический интерес для решения задачи ввода идентификационных документов на мобильных устройствах.

Технологии распознавания в видеопотоке на мобильных устройствах, разработанные Арлазаровым В.В., позволяют улучшить качество и эффективность обслуживания клиентов и обработки документов в банковской сфере. Данные технологии в составе программных продуктов ООО «Смарт Энджинс Сервис» внедрены и используются в информационных системах и мобильных приложениях АО «Альфа-Банк».

Альфа-Банк стал первым в рейтинге Топ-50 российских банков (Top 50 Russian banks 2022) по версии авторитетного международного финансового журнала The Banker. Альфа-Банк победил в главной номинации премии Банки.ру и назван «Банком года» по итогам 2021 года.

Директор департамента развития
Цифровых каналов физических лиц — Б.В. Гаврилов

alfabank.ru

АО «АЛЬФА-БАНК»
ул. Каланчёвская, 27
Москва, 107078

+7 495 620 91 91
+7 495 974 25 15
mail@alfabank.ru

АЛЬФА СТРАХОВАНИЕ
СЕРВИС БУДУЩЕГО В НАСТОЯЩЕМ

АКТ
об использовании (внедрении) результатов диссертационной работы Арлазарова Владимира Викторовича «Мобильное распознавание и его применение к системе ввода идентификационных документов» в АО «АльфаСтрахование»

Результаты диссертационной работы «Мобильное распознавание и его применение к системе ввода идентификационных документов» актуальны и представляют практический интерес для решения задачи ввода документов в информационных системах и приложениях.

В составе программных продуктов ООО «Смарт Энджинс Сервис», разработанные Арлазаровым В.В. технологии позволяют повысить скорость обработки документов в дистанционных каналах обслуживания и внутренних бизнес-процессах АльфаСтрахование.

Директор
операционного департамента Корецкая Н.Г.

ТИНЬКОФФ

АКЦИОНЕРНОЕ ОБЩЕСТВО «ТИНЬКОФФ БАНК»
РОССИЯ, 127287, МОСКВА, УЛ. 2-Я ХУТОРСКАЯ, Д. 38А, СТР. 26
ТЕЛ.: +7 495 648-10-00, TINKOFF.RU

Исх. № КБ-0217.45
От 17.02.2023 г.

АКТ
об использовании (внедрении) результатов диссертационной работы Арлазарова Владимира Викторовича «Мобильное распознавание и его применение к системе ввода идентификационных документов» в АО «Тинькофф Банк»

Результаты диссертационной работы «Мобильное распознавание и его применение к системе ввода идентификационных документов» обладают высокой актуальностью и представляют практический интерес для решения задачи ввода идентификационных документов на мобильных устройствах.

Технологии распознавания в видеопотоке на мобильных устройствах разработанные Арлазаровым В.В., позволяют улучшить качество и эффективность обслуживания клиентов в банковской сфере. Данные технологии в составе программных продуктов ООО «Смарт Энджинс Сервис» внедрены и используются в информационных системах и мобильных приложениях АО «Тинькофф Банк».

«Тинькофф Банк» был признан лучшим розничным онлайн-банком в мире в 2020 и 2018 гг. по версии Global Finance. В 2020 г. банк также стал победителем в категории «Лучший розничный европейский банк» международной банковской премии Retail Banker International Awards. Мобильное приложение банка регулярно признается лучшим на рынке российскими и международными независимыми экспертами (Deloitte в 2013, 2014, 2015 и 2016 гг., Global Finance в 2018 г.).

Директор по информационным технологиям,
Заместитель председателя правления
АО «Тинькофф Банк» **В.В. Цыганов**

smart engines™

ООО Смарт Энджинс Сервис
ОГРН: 1167746085297
ИНН: 7728328449

117312, город Москва,
пр-кт 60-Летия Октября, д. 9

T: +7 (495) 649-82-60
E: office@smartengines.ru
https://smartengines.ru

17.02.2023 № 005
На № ..... от .....

АКТ

об использовании (внедрении) результатов диссертационной работы Арлазарова Владимира Викторовича «Мобильное распознавание и его применение к системе ввода идентификационных документов» в программных продуктах ООО «Смарт Энджинс Сервис»

Настоящий акт выдан В.В. Арлазарову для предоставления в Диссертационный совет, свидетельствующий о том, что результаты диссертационной работы «Мобильное распознавание и его применение к системе ввода идентификационных документов» внедрены в семейство программных продуктов распознавания изображений документов, а именно:

– Программа для распознавания идентификационных карт личности "Smart IDReader"
– Программа поиска плоских ригидных объектов "Smart ARTour"
– Программа распознавания признаков подлинности "Smart Document Forensics"
– Smart ID Engine
– Smart Document Engine
– Smart Code Engine

Указанные программы внедрены в различных областях экономики и управления. Так, распознавание российского паспорта используется для процессов регистрации пользователей ведущих банков РФ – Тинькофф, Альфа-банк, Газпромбанк, Открытие, Райффайзен, Росбанк, ДомРФ, Точка банк, Совкомбанк, МТС банк, Хоум Кредит Банк, МКБ, и ряд региональных банков. Распознавание водительских удостоверений, свидетельств о регистрации транспортных средств, паспортов транспортных средств и паспорта граждан РФ внедрены в ряде страховых компаний – Ингострах, Альфа-страхование, РЕСО-Гарантия, Согласие.

Комплекс распознавания документов всего мира, включающий паспорта 210 стран и организаций, внутренних идентификационных карт, видов на жительство и водительских удостоверений всех стран мира, внедрен в банке ЕАБР, сервисе iDenfy, travizory, Oman Arab Bank, Emirates NDB, Dukascopy Swiss Banking Group, Kaspi.kz и ряде других крупных организаций.

В области мобильной связи, операторы МТС, Билайн и Мегафон используют созданный комплекс для идентификации абонентов при продаже SIM-карт.

Транспортная отрасль активно использует программный и программно-аппаратный комплексы идентификации, созданный на основе указанных программ. Так, РЖД использует его для продажи железнодорожных билетов в 850 кассах, международный конгломерат SITA,

smart engines™

ООО Смарт Энджинс Сервис
ОГРН: 1167746085297
ИНН: 7728328449

117312, город Москва,
пр-кт 60-Летия Октября, д. 9

T: +7 (495) 649-82-60
E: office@smartengines.ru
https://smartengines.ru

17.02.2023 № 005
На № ..... от .....

авиакомпании Turkish Airlines и Croatia Airlines используют мобильную систему распознавания для регистрации на рейс, а в аэропорту Шереметьево программно-аппаратный комплекс используется для автоматического пересечения границы. Крупнейший оператор круизных линий в мире RCCL использует его для продажи билетов и прохода на лайнеры.

Созданный комплекс используется системой изготовления и выдачи паспортно-визовых документов ГС МИР. Для нужд ФНС России программный комплекс используется в мобильном приложении регистрации самозанятых и ИП. Также программно-аппаратный комплекс используется для выдачи ЭЦП руководителям организаций.

Созданный совместно с китайской компанией Pixsur программно-аппаратный комплекс использовался для регистрации посетителей стадионов во время чемпионата мира по футболу в Катаре.

Общее число организаций, использующих решения, построенные на основе перечисленных программ, составляет более 200 по всему миру.

Исполнительный директор
ООО «Смарт Энджинс Сервис»
Усилин Сергей Александрович

ВТБ

АКТ
об использовании (внедрении) результатов диссертационной работы Арлазарова Владимира Викторовича «Мобильное распознавание и его применение к системе ввода идентификационных документов» в Банке ВТБ

Результаты диссертационной работы «Мобильное распознавание и его применение к системе ввода идентификационных документов» обладают высокой актуальностью и представляют практический интерес для решения задачи ввода документов на мобильных устройствах.

Технологии распознавания в видеопотоке на мобильных устройствах разработанные Арлазаровым В.В., позволяют улучшить качество и эффективность обслуживания клиентов в дистанционных каналах обслуживания. Данные технологии в составе программных продуктов ООО «Смарт Энджинс Сервис» внедрены и используются в мобильных и веб-приложениях Банка ВТБ (ПАО).

С уважением,

Начальник Управления Платежи

Департамента транзакционного розничного бизнеса Р.В. Новиков

Приложение Б

Результаты интеллектуальной деятельности

(12) **United States Patent**
Skoryukina et al.

(10) **Patent No.: US 11,574,492 B2**
(45) **Date of Patent: Feb. 7, 2023**

(54) **EFFICIENT LOCATION AND IDENTIFICATION OF DOCUMENTS IN IMAGES**

(71) Applicant: **Smart Engines Service, LLC**, Moscow (RU)

(72) Inventors: **Natalya Sergeevna Skoryukina**, Moscow (RU); **Vladimir Viktorovich Arlazarov**, Moscow (RU); **Dmitry Petrovich Nikolaev**, Moscow (RU); **Igor Aleksandrovich Faradjev**, Moscow (RU)

(73) Assignee: **SMART ENGINES SERVICE, LLC**, Moscow (RU)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 113 days.

(21) Appl. No.: **17/237,596**

(22) Filed: **Apr. 22, 2021**

(65) **Prior Publication Data**

US 2022/0067363 A1 Mar. 3, 2022

(30) **Foreign Application Priority Data**

Sep. 2, 2020 (RU) 2020129039

(51) **Int. Cl.**

G06K 9/46	(2006.01)
G06K 9/66	(2006.01)
G06V 30/413	(2022.01)
G06K 9/62	(2022.01)
G06V 10/40	(2022.01)

(Continued)

(52) **U.S. Cl.**

CPC ***G06V 30/413*** (2022.01); ***G06K 9/6232*** (2013.01); ***G06V 10/40*** (2022.01); ***G06V 10/44*** (2022.01); ***G06V 10/757*** (2022.01); ***G06V 10/758*** (2022.01); ***G06V 30/418*** (2022.01); *G06V 30/10* (2022.01)

(58) **Field of Classification Search**

None

See application file for complete search history.

(56) **References Cited**

U.S. PATENT DOCUMENTS

6,310,984 B2 * 10/2001 Sansom-Wai G06V 10/24 382/173

8,705,836 B2 * 4/2014 Gorski G06V 30/274 382/137

(Continued)

OTHER PUBLICATIONS

Attivissimo et al., "An automatic reader of identity documents." In 2019 IEEE International Conference on Systems, Man and Cybernetics (SMC), pp. 3525-3530. IEEE, 2019. (Year: 2019).*

(Continued)

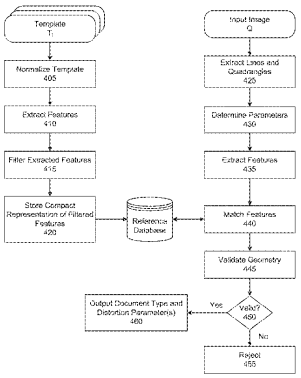
Primary Examiner — Feng Niu

(74) *Attorney, Agent, or Firm* — Procopio, Cory, Hargreaves & Savitch LLP

(57) **ABSTRACT**

Efficient location and identification of documents in images. In an embodiment, at least one quadrangle is extracted from an image based on line(s) extracted from the image. Parameter(s) are determined from the quadrangle(s), and keypoints are extracted from the image based on the parameter(s). Input descriptors are calculated for the keypoints and used to match the keypoints to reference keypoints, to identify classification candidate(s) that represent a template image of a type of document. The type of document and distortion parameter(s) are determined based on the classification candidate(s).

26 Claims, 6 Drawing Sheets



US010354142B2

(12) **United States Patent**
Arlazarov et al.

(10) **Patent No.: US 10,354,142 B2**
(45) **Date of Patent: Jul. 16, 2019**

(54) **METHOD FOR HOLOGRAPHIC ELEMENTS DETECTION IN VIDEO STREAM**

(71) Applicant: **Smart Engines Service LLC**, Moscow (RU)

(72) Inventors: **Vladimir Viktorovich Arlazarov**, Moscow (RU); **Timofey Sergeevich Chernov**, Dzerzhinsky (RU); **Dmitry Petrovich Nikolaev**, Moscow (RU); **Natalya Sergeevna Skoryukina**, Domodedovo (RU); **Oleg Anatolyevitch Slavin**, Moscow (RU)

(73) Assignee: **SMART ENGINES SERVICE LLC**, Moscow (RU)

(*) Notice: Subject to any disclaimer, the term of this patent is extended or adjusted under 35 U.S.C. 154(b) by 152 days.

(21) Appl. No.: **15/668,455**

(22) Filed: **Aug. 3, 2017**

(65) **Prior Publication Data**

US 2018/0247125 A1 Aug. 30, 2018

(30) **Foreign Application Priority Data**

Feb. 27, 2017 (RU) 2017106048

(51) **Int. Cl.**
G06K 9/00 (2006.01)
G06K 9/46 (2006.01)
(Continued)

(52) **U.S. Cl.**
CPC ***G06K 9/00711*** (2013.01); ***G06K 9/00442*** (2013.01); ***G06K 9/2054*** (2013.01);
(Continued)

(58) **Field of Classification Search**
CPC G06K 9/00; G06F 3/00
(Continued)

(56) **References Cited**

U.S. PATENT DOCUMENTS

9,442,459 B2 * 9/2016 Dluhos G03H 1/0005

FOREIGN PATENT DOCUMENTS

CN 101915617 B 8/2012
CN 103196560 A 7/2013
(Continued)

OTHER PUBLICATIONS

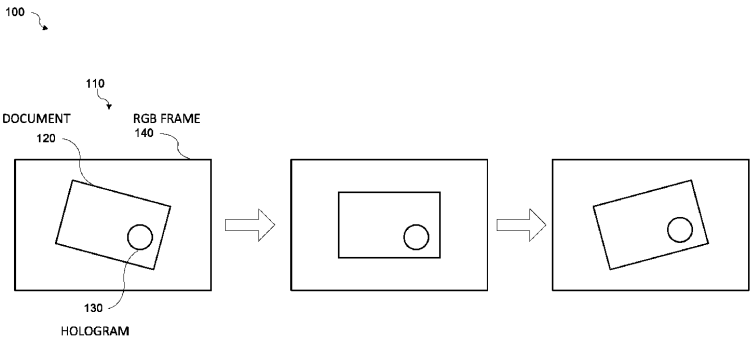
Hartl, A., et al., AR-Based Hologram Detection on Security Documents Using a Mobile Phone, Springer International Publishing, 2014, pp. 335-346.
(Continued)

Primary Examiner — Abolfazl Tabatabai
(74) *Attorney, Agent, or Firm* — Procopio; Mark W. Catanese; Noel C. Gillespie

(57) **ABSTRACT**

A method for detecting holographic elements in a video stream containing images in the form of documents includes: processing of a video stream in which the document image is stabilized; constructing saturation and color tone maps; analyzing color characteristics in image regions; constructing histograms of color characteristics; estimating a change in the color characteristics at least in part based on data obtained by calculating a difference between the histograms of a current and a previous frame; constructing an integrated map of hologram presence estimates by combining calculated estimates for all video stream frames based at least in part on the estimation of the change in color characteristics; and determining final regions of the holographic elements based at least in part on the integrated map of the hologram presence estimates.

15 Claims, 4 Drawing Sheets



РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2771005

Способ детектирования голографической защиты на документах в видеопотоке

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Коляскина Лейсан Ильдаровна (RU), Николаев Дмитрий Петрович (RU), Полевой Дмитрий Валерьевич (RU), Тропин Даниил Вячеславович (RU), Усилин Сергей Александрович (RU)***

Заявка № **2021121819**
Приоритет изобретения **22 июля 2021 г.**
Дата государственной регистрации в Государственном реестре изобретений Российской Федерации **25 апреля 2022 г.**
Срок действия исключительного права на изобретение истекает **22 июля 2041 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2774058

Способ определения (распознавания) факта предъявления цифровой копии документа в виде пересъемки экрана

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Николаев Дмитрий Петрович (RU), Полевой Дмитрий Валерьевич (RU), Слугин Дмитрий Геннадьевич (RU), Кунина Ирина Андреевна (RU), Сигарева Ирина Витальевна (RU)***

Заявка № **2021128626**
Приоритет изобретения **30 сентября 2021 г.**
Дата государственной регистрации
в Государственном реестре изобретений
Российской Федерации **14 июня 2022 г.**
Срок действия исключительного права
на изобретение истекает **30 сентября 2041 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2750395

Способ оценки действительности документа при помощи оптического распознавания текста на изображении круглого оттиска печати/штампа на цифровом изображении документа

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Алиев Михаил Александрович (RU), Арлазаров Владимир Викторович (RU), Маталов Даниил Павлович (RU), Николаев Дмитрий Петрович (RU), Полевой Дмитрий Валерьевич (RU), Усилин Сергей Александрович (RU)***

Заявка № **2020127688**
Приоритет изобретения **19 августа 2020 г.**
Дата государственной регистрации
в Государственном реестре изобретений
Российской Федерации **28 июня 2021 г.**
Срок действия исключительного права
на изобретение истекает **19 августа 2040 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2724967

Система дистанционного приобретения билетов на культурно-массовые мероприятия с использованием распознавания на мобильном устройстве

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Николаев Дмитрий Павлович (RU), Славин Олег Анатольевич (RU), Усилин Сергей Александрович (RU), Шешкус Александр Владимирович (RU)***

Заявка № **2020110146**

Приоритет изобретения **11 марта 2020 г.**

Дата государственной регистрации в Государственном реестре изобретений Российской Федерации **29 июня 2020 г.**

Срок действия исключительного права на изобретение истекает **11 марта 2040 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ИЗОБРЕТЕНИЕ

№ 2643130

Автоматизированное рабочее место контроля паспортных документов

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Гладков Андрей Павлович (RU), Николаев Дмитрий Петрович (RU), Усилин Сергей Александрович (RU)***

Заявка № **2017105336**
Приоритет изобретения **20 февраля 2017 г.**
Дата государственной регистрации в Государственном реестре изобретений Российской Федерации **30 января 2018 г.**
Срок действия исключительного права на изобретение истекает **20 февраля 2037 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 210539

Система дистанционной регистрации граждан на избирательном участке с распознаванием паспорта РФ

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

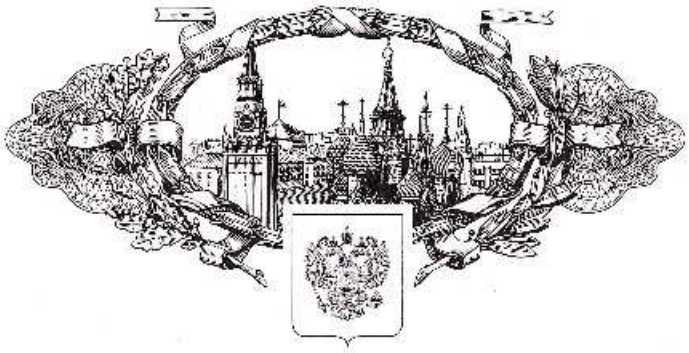
Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Булатов Константин Булатович (RU), Славин Олег Анатольевич (RU), Усилин Сергей Александрович (RU)***

Заявка № **2021136611**
Приоритет полезной модели **10 декабря 2021 г.**
Дата государственной регистрации
в Государственном реестре полезных
моделей Российской Федерации **19 апреля 2022 г.**
Срок действия исключительного права
на полезную модель истекает **10 декабря 2031 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 210845

Система контроля соблюдения санитарно-эпидемиологических правил при дистанционной продаже билетов на транспорт при помощи мобильных устройств

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Арлазоров Владимир Викторович (RU), Арлазоров Никита Викторович (RU), Безматерных Павел Владимирович (RU), Булатов Константин Булатович (RU), Полевой Дмитрий Валентинович (RU), Славин Олег Анатольевич (RU)***

Заявка № **2022100687**
Приоритет полезной модели **12 января 2022 г.**
Дата государственной регистрации в Государственном реестре полезных моделей Российской Федерации **11 мая 2022 г.**
Срок действия исключительного права на полезную модель истекает **12 января 2032 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 210846

Система предотвращения атаки на предъявление дистанционного скоринга клиентов при выдаче кредитов с помощью мобильного устройства

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

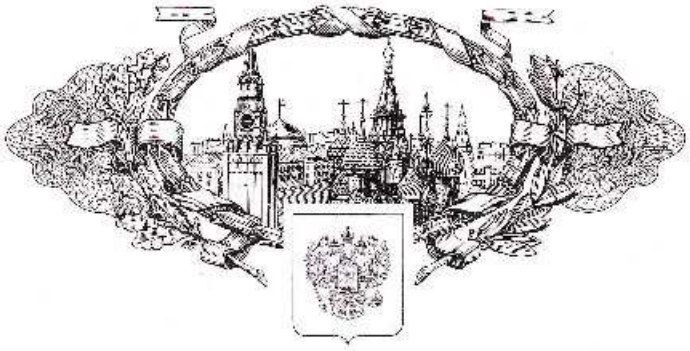
Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Безматерных Павел Владимирович (RU), Булатов Константин Булатович (RU), Кунина Ирина Андреевна (RU), Маталов Даниил Павлович (RU), Тропин Даниил Вячеславович (RU), Усилин Сергей Александрович (RU)***

Заявка № **2022102438**
Приоритет полезной модели **01 февраля 2022 г.**
Дата государственной регистрации в Государственном реестре полезных моделей Российской Федерации **11 мая 2022 г.**
Срок действия исключительного права на полезную модель истекает **01 февраля 2032 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 210919

Система контроля соблюдения правил дистанционной торговли рецептурными препаратами (лекарствами) при помощи мобильного устройства

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

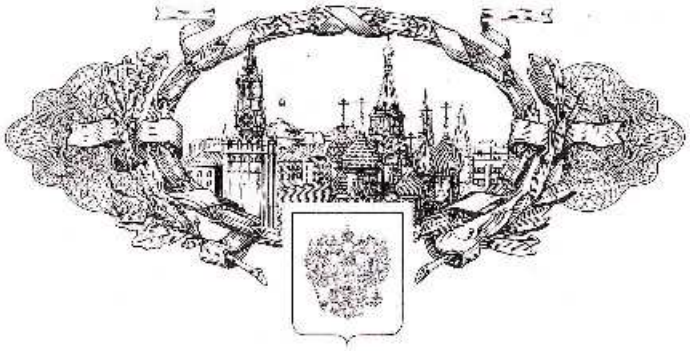
Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Лимонова Елена Евгеньевна (RU), Маталов Даниил Павлович (RU), Полевой Дмитрий Валерьевич (RU), Тропин Даниил Вячеславович (RU)***

Заявка № **2022101508**
Приоритет полезной модели **24 января 2022 г.**
Дата государственной регистрации
в Государственном реестре полезных
моделей Российской Федерации **13 мая 2022 г.**
Срок действия исключительного права
на полезную модель истекает **24 января 2032 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 211342

Система для дистанционного опроса граждан при переписи паселения

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Маталов Даниил Павлович (RU), Славин Олег Анатольевич (RU), Шешкус Александр Владимирович (RU)***

Заявка № **2021135025**
Приоритет полезной модели **29 ноября 2021 г.**
Дата государственной регистрации
в Государственном реестре полезных
моделей Российской Федерации **01 июня 2022 г.**
Срок действия исключительного права
на полезную модель истекает **29 ноября 2031 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Ю.С. Зубов

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 204787

Система удаленной регистрации абонентов сети связи с использованием мобильного устройства

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

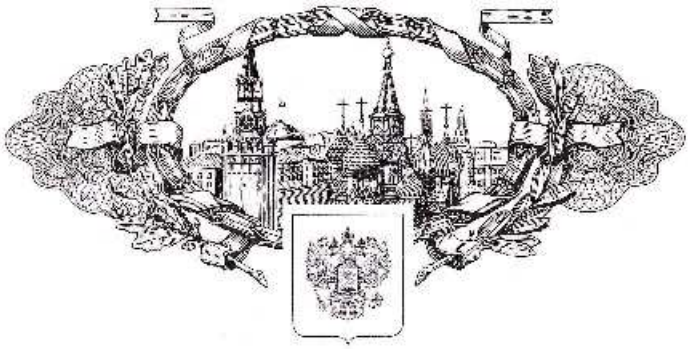
Авторы: ***Безматерных Павел Владимирович (RU), Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Скорюкина Наталья Сергеевна (RU), Славин Олег Анатольевич (RU)***

Заявка № **2021100924**
Приоритет полезной модели **18 января 2021 г.**
Дата государственной регистрации
в Государственном реестре полезных
моделей Российской Федерации **10 июня 2021 г.**
Срок действия исключительного права
на полезную модель истекает **18 января 2031 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 204371

Система автоматического подтверждения отсутствия признаков инфекционного заболевания с помощью документального свидетельства

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Булатов Константин Булатович (RU), Славин Олег Анатольевич (RU)***

Заявка № **2021103918**
Приоритет полезной модели **16 февраля 2021 г.**
Дата государственной регистрации в Государственном реестре полезных моделей Российской Федерации **21 мая 2021 г.**
Срок действия исключительного права на полезную модель истекает **16 февраля 2031 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 207759

Система проверки подлинности паспортно-визовых документов с использованием мобильного устройства

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

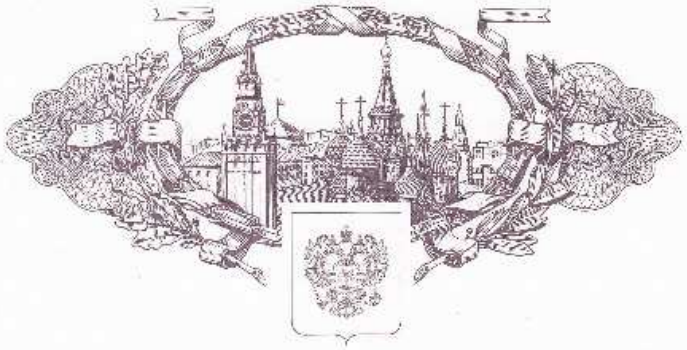
Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Усилин Сергей Александрович (RU)***

Заявка № **2021125026**
Приоритет полезной модели **23 августа 2021 г.**
Дата государственной регистрации
в Государственном реестре полезных
моделей Российской Федерации **15 ноября 2021 г.**
Срок действия неключительного права
на полезную модель истекает **23 августа 2031 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 196455

Система удаленной регистрации данных граждан на избирательном участке с использованием мобильного устройства

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Булатов Константин Булатович (RU), Шешкус Александр Владимирович (RU)***

Заявка № **2019141280**
Приоритет полезной модели **13 декабря 2019 г.**
Дата государственной регистрации в Государственном реестре полезных моделей Российской Федерации **02 марта 2020 г.**
Срок действия исключительного права на полезную модель истекает **13 декабря 2029 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 191682

Система покупки цифрового контента с использованием мобильного устройства

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Булатов Константин Булатович (RU), Скорюкина Наталья Сергеевна (RU), Николаев Дмитрий Петрович (RU)***

Заявка № **2019116550**
Приоритет полезной модели **29 мая 2019 г.**
Дата государственной регистрации в Государственном реестре полезных моделей Российской Федерации **15 августа 2019 г.**
Срок действия исключительного права на полезную модель истекает **29 мая 2029 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 180207

Система автоматического контроля личности избирателя

Патентообладатель: ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

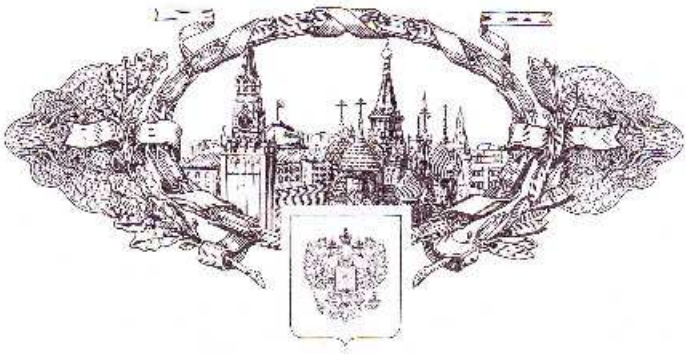
Авторы: ***Арлазаров Владимир Викторович (RU), Арлазаров Никита Викторович (RU), Булатов Константин Булатович (RU), Скорюкина Наталья Сергеевна (RU), Славин Олег Анатольевич (RU), Чернов Тимофей Сергеевич (RU)***

Заявка № **2017139215**
Приоритет полезной модели **13 ноября 2017 г.**
Дата государственной регистрации в Государственном реестре полезных моделей Российской Федерации **06 июня 2018 г.**
Срок действия исключительного права на полезную модель истекает **13 ноября 2027 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 161478

СИСТЕМА ДОСТУПА К ДИСТАНЦИОННОМУ ПОЛУЧЕНИЮ БАНКОВСКИХ УСЛУГ

Патентообладатель(ли): ***Арлазаров Владимир Викторович (RU), Арлазарова Анна Рудольфовна (RU), Арлазаров Никита Викторович (RU), Николаев Дмитрий Петрович (RU), Славин Олег Анатольевич (RU), Усилин Сергей Александрович (RU), Шешкус Александр Владимирович (RU)***

Автор(ы): ***см. на обороте***

Заявка № **2015156508**

Приоритет полезной модели **29 декабря 2015 г.**

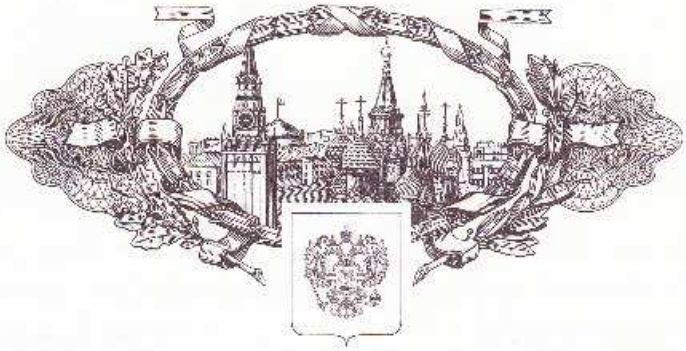
Зарегистрировано в Государственном реестре полезных моделей Российской Федерации ***04 апреля 2016 г.***

Срок действия патента истекает **29 декабря 2025 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ



ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 159733

СИСТЕМА РАСПОЗНАВАНИЯ ДОКУМЕНТОВ В ВИДЕОПОСЛЕДОВАТЕЛЬНОСТИ

Патентообладатель(ли): ***Арлазаров Владимир Викторович (RU), Арлазаров Владимир Львович (RU), Булатов Константин Булатович (RU), Николаев Дмитрий Петрович (RU), Полевой Дмитрий Валерьевич (RU), Славин Олег Анатольевич (RU)***

Автор(ы): ***см. на обороте***

Заявка № **2015145155**

Приоритет полезной модели **21 октября 2015 г.**

Зарегистрировано в Государственном реестре полезных моделей Российской Федерации ***26 января 2016 г.***

Срок действия патента истекает **21 октября 2025 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 166152

АВТОНОМНОЕ АВТОМАТИЗИРОВАННОЕ РАБОЧЕЕ МЕСТО КОНТРОЛЯ ПАСПОРТНЫХ ДОКУМЕНТОВ

Патентообладатель(ли): ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Автор(ы): ***Арлазаров Владимир Викторович (RU), Гладков Андрей Павлович (RU), Николаев Дмитрий Петрович (RU), Усилин Сергей Александрович (RU)***

Заявка № **2016122432**

Приоритет полезной модели **07 июня 2016 г.**

Зарегистрировано в Государственном реестре полезных моделей Российской Федерации ***26 октября 2016 г.***

Срок действия патента истекает **07 июня 2026 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 163168

ТЕХНОЛОГИЧЕСКАЯ ПЛАТФОРМА ЭЛЕКТРОННОГО ДОКУМЕНТООБОРОТА ОСМОТРА АВТОМОБИЛЯ ДЛЯ ОФОРМЛЕНИЯ СТРАХОВКИ

Патентообладатель(ли): ***Арлазаров Владимир Викторович (RU), Арлазарова Анна Рудольфовна (RU), Славин Олег Анатольевич (RU), Усилин Сергей Александрович (RU)***

Автор(ы): ***см. на обороте***

Заявка № **2015148236**

Приоритет полезной модели **10 ноября 2015 г.**

Зарегистрировано в Государственном реестре полезных моделей Российской Федерации ***17 июня 2016 г.***

Срок действия патента истекает **10 ноября 2025 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

ПАТЕНТ

НА ПОЛЕЗНУЮ МОДЕЛЬ

№ 166038

АВТОМАТИЗИРОВАННОЕ РАБОЧЕЕ МЕСТО КОНТРОЛЯ ПАСПОРТНЫХ ДОКУМЕНТОВ

Патентообладатель(ли): ***Общество с ограниченной ответственностью "СМАРТ ЭНДЖИНС СЕРВИС" (RU)***

Автор(ы): ***Арлазаров Владимир Викторович (RU), Гладков Андрей Павлович (RU), Николаев Дмитрий Петрович (RU), Усилин Сергей Александрович (RU)***

Заявка № **2016106183**

Приоритет полезной модели **25 февраля 2016 г.**

Зарегистрировано в Государственном реестре полезных моделей Российской Федерации ***24 октября 2016 г.***

Срок действия патента истекает **25 февраля 2026 г.**

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2019665480

Программа точной локализации и типизации объекта страницы документа в видеопотоке

Правообладатель: ***Федеральное государственное учреждение «Федеральный исследовательский центр «Информатика и управление» Российской академии наук» (ФИЦ ИУ РАН) (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Яншевский Игорь Михайлович (RU)***

Заявка № **2019664400**

Дата поступления **13 ноября 2019 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***22 ноября 2019 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2018615343

Программа распознавания признаков подлинности "Smart Document Forensics"

Правообладатель: ***Общество с ограниченной ответственностью «Смарт Энджинс Сервис» (RU)***

Авторы: ***Усилин Сергей Александрович (RU), Арлазаров Владимир Викторович (RU), Алиев Михаил Александрович (RU), Маталов Даниил Павлович (RU)***

Заявка № **2018612851**
Дата поступления **23 марта 2018 г.**
Дата государственной регистрации
в Реестре программ для ЭВМ ***07 мая 2018 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2018615952

Программа поиска плоских ригидных объектов "Smart ARTour"

Правообладатель: ***Общество с ограниченной ответственностью «Смарт Энджинс Сервис» (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Булатов Константин Булатович (RU), Николаев Дмитрий Петрович (RU), Скорюкина Наталья Сергеевна (RU)***

Заявка № **2018612805**

Дата поступления **23 марта 2018 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***18 мая 2018 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев

РОССИЙСКАЯ ФЕДЕРАЦИЯ

СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2016616961

Программа для распознавания идентификационных карт личности "Smart IDReader"

Правообладатель: ***Общество с ограниченной ответственностью «Смарт Энджинс Сервис» (RU)***

Авторы: ***Арлазаров Владимир Викторович (RU), Николаев Дмитрий Петрович (RU), Усилин Сергей Александрович (RU), Булатов Константин Булатович (RU), Чернов Тимофей Сергеевич (RU), Слугин Дмитрий Геннадьевич (RU), Ильин Дмитрий Алексеевич (RU), Безматерных Павел Владимирович (RU), Муковозов Арсений Александрович (RU), Лимонова Елена Евгеньевна (RU)***

Заявка № **2016612014**

Дата поступления **10 марта 2016 г.**

Дата государственной регистрации в Реестре программ для ЭВМ ***22 июня 2016 г.***

Руководитель Федеральной службы по интеллектуальной собственности

Г.П. Ивлиев