

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ УЧРЕЖДЕНИЕ  
ФЕДЕРАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЦЕНТР  
«ИНФОРМАТИКА И УПРАВЛЕНИЕ» РОССИЙСКОЙ АКАДЕМИИ НАУК

На правах рукописи

Чернов Тимофей Сергеевич

**Математические модели и алгоритмы оценки качества  
изображений в системах оптического распознавания**

Специальность 05.13.01 —  
Системный анализ, управление и обработка информации  
(информационно-вычислительное обеспечение)

Диссертация на соискание ученой степени  
кандидата технических наук

Научный руководитель  
доктор технических наук  
Славин Олег Анатольевич

Москва — 2018

## Оглавление

	Стр.
<b>Введение . . . . .</b>	<b>5</b>
 <b>Глава 1. Искажения изображений и их влияние на системы</b>	
<b>распознавания . . . . .</b>	<b>13</b>
1.1 Системы распознавания изображений . . . . .	13
1.1.1 Документооборот и долгосрочное хранение документов . .	13
1.1.2 Контроль доступа и безопасности на предприятии . . . . .	14
1.1.3 Распознавание изображений в технологических процессах .	15
1.1.4 Архитектура систем распознавания . . . . .	16
1.1.5 Системы сжатия изображений документов . . . . .	18
1.1.6 Постановка задач распознавания . . . . .	22
1.1.7 Надежность распознавания . . . . .	24
1.2 Типы искажений входных данных систем распознавания . . . . .	28
1.2.1 Шум и абберации . . . . .	28
1.2.2 Сжатие и интерполяция изображений . . . . .	29
1.2.3 Дефокусировка и размытость . . . . .	29
1.2.4 Тени, блики и голографические элементы . . . . .	31
1.2.5 Механическая деформация и артефакты объекта . . . . .	31
1.2.6 Неустойчивость систем распознавания к искажениям . . . .	32
1.3 Оценка качества изображений . . . . .	33
1.3.1 Виды и приложения методов оценки качества изображений	33
1.3.2 Субъективные методы оценки качества изображений . . . .	35
1.3.3 Объективные методы оценки качества изображений . . . . .	37
1.3.4 Специализированная оценка качества изображений . . . . .	42
1.4 Выводы по главе 1. Основные задачи диссертации . . . . .	47
 <b>Глава 2. Оценка качества изображений в системах распознавания . . .</b>	
2.1 Постановка задачи оценки качества изображений . . . . .	49
2.2 Модель системы распознавания с модулями оценки качества . . . .	51
2.3 Выбор лучших кадров для распознавания в видеопотоке . . . . .	53
2.4 Эксперименты . . . . .	56

	Стр.
2.4.1 Алгоритм оценки степени размытости кадра . . . . .	56
2.4.2 Моделирование распознавания в процессе автоматической фокусировки камеры . . . . .	57
2.4.3 Экспериментальные результаты . . . . .	58
2.5 Выводы по главе 2 . . . . .	61

### **Глава 3. Оценка качества цветовой сегментации изображения**

документа . . . . .	62
3.1 Постановка задачи . . . . .	62
3.2 Общий подход к оценке качества цветовой сегментации . . . . .	64
3.3 Оценка качества текстового слоя с помощью OCR-систем . . . . .	65
3.4 Оценка качества графического слоя . . . . .	66
3.5 Анализ пригодности разработанных алгоритмов оценки цветовой сегментации . . . . .	69
3.6 Автоматическая оптимизация качества цветовой сегментации . . . . .	70
3.7 Выводы по главе 3 . . . . .	72

### **Глава 4. Оценка качества изображений при контроле доступа и безопасности на предприятии . . . . .**

документов с мобильных устройств . . . . .	73
4.1.1 Влияние бликов на системы распознавания документов . . . . .	74
4.1.2 Детектирование областей бликов . . . . .	75
4.1.3 Интерполяция фона на поврежденных бликами областях . . . . .	76
4.1.4 Экспериментальные результаты . . . . .	77
4.2 Анализ устойчивости алгоритма проективного оптического потока к бликам . . . . .	80
4.2.1 Подавление бликов преобразованием цветового пространства . . . . .	81
4.2.2 Алгоритм вычисления оптического потока . . . . .	82
4.2.3 Описание тестового набора данных . . . . .	85
4.2.4 Описание экспериментального исследования . . . . .	86
4.2.5 Экспериментальные результаты . . . . .	88
4.3 Детектирование и локализация периодических фоновых элементов на изображении документа . . . . .	90

	Стр.
4.3.1 Модель сигнала изображения . . . . .	92
4.3.2 Детектирование периодического шаблона . . . . .	97
4.3.3 Локализация периодических шаблонов . . . . .	103
4.3.4 Анализ устойчивости к ошибкам нахождения границ документа . . . . .	112
4.3.5 Экспериментальные результаты . . . . .	114
4.3.6 Общая схема алгоритма . . . . .	120
4.4 Выводы по главе 4 . . . . .	121
 <b>Глава 5. Контроль качества входного сигнала двухканального рентгенографического сепаратора . . . . .</b>	
5.1 Рентгенографическая сепарация . . . . .	122
5.1.1 Линейная модель рентгенографического изображения . . .	123
5.1.2 Устройство рентгенографического сепаратора . . . . .	124
5.2 Контроль качества входного сигнала . . . . .	125
5.3 Динамический контроль чувствительности канала . . . . .	127
5.4 Обнаружение заслонения детектора . . . . .	130
5.5 Обнаружение оптического загрязнения ленты транспортера . . .	132
5.6 Выводы по главе 5 . . . . .	135
 <b>Заключение . . . . .</b>	 136
 <b>Список литературы . . . . .</b>	 138



## Введение

**Актуальность темы.** Развитие науки и технологий послужило причиной тому, что современное общество вступило в новый этап использования систем распознавания образов. Большой вклад в развитие методов распознавания внесли отечественные и зарубежные ученые М.А. Айзерман, В.Л. Арлазаров, Э.М. Браверман, В.Н. Вапник, Ю.И. Журавлев, Н.Г. Загоруйко, А.Г. Ивахненко, В.Д. Мазуров, В.А. Сойфер, А.Я. Червоненкис, Y. Bengio, G. Hinton, M. Jones, Y. LeCun, P. Viola и многие другие.

Как и несколько десятилетий назад, системы распознавания и машинного зрения широко задействуются в анализе документов и текстов, медицине, биометрии, обороне, промышленности и многих других сферах. Тем не менее, процесс их использования и требования к ним существенно изменились.

Точность, устойчивость и производительность методов распознавания образов повышаются, за счет чего расширяется область их применения в автоматизации процессов жизнедеятельности, где системам распознавания доверяются все более важные и критические задачи с высокой ценой ошибки. В существующих профессиональных сферах их применения снижается время и степень участия ответственного работника или оператора. В этих условиях система распознавания должна стабильно выдавать сигнал о недостаточной достоверности результата распознавания, также называемый отказом, при подозрении на некорректность результата или при возникновении нештатной ситуации, что является запросом на передачу контроля дальнейших действий оператору.

Помимо профессиональных сфер, системы распознавания получают свое распространение и среди обычных пользователей, все более плотно проникая в повседневную жизнь людей и становясь неотъемлемой частью быта. Важную роль в этом играют повсеместно используемые мобильные устройства, с каждым годом повышающие свою производительность, доступность и простоту в использовании. Вместе с тем, взаимодействие систем распознавания с непрофессиональной аудиторией открывает ряд новых задач.

Обычный пользователь не является оператором в привычном смысле этого слова: он не знает, как устроен тот или иной алгоритм распознавания и не понимает, как выглядит допустимое множество входных данных, на которых следует ожидать корректный результат. Как следствие, теряется контроль над

изображениями, поступающими в систему распознавания, из-за чего алгоритмам необходимо подстраиваться под всевозможные искажения входных данных. Кроме того, пользователь не привык контролировать выход систем распознавания, что снова перекладывает ответственность за информирование пользователя об ошибке на разработчика системы.

Даже при текущем уровне прогресса построение систем распознавания с абсолютной точностью на практике невозможно. Тем не менее, разные задачи имеют разную цену ошибки или требуемый уровень надежности, который можно обеспечить с помощью правил определения достоверности результатов распознавания. Проблема заключается в том, что при искаженных входных данных поведение систем распознавания не всегда предсказуемо. Следовательно, для построения систем распознавания заданной надежности необходима разработка методов контроля качества входных изображений.

На текущий момент вопросам оценки качества изображений уделяется достаточно большое внимание вследствие их высокой актуальности. Среди ученых, внесших значительный вклад в развитие области оценки качества изображений, можно отметить A. Bovik, Z. Wang, D. Chandler, H. Sheikh, L. Zhang, E. Simoncelli, P. Ye, D. Doermann, J. Kumar и других. Разработано множество объективных (автоматических) методов оценки и коррекции качества изображений. Несмотря на это, практическое применение имеющихся методов в промышленных системах обработки и распознавания изображений имеет существенные ограничения.

Во-первых, не существует универсального метода оценки качества изображений, применимого в любых приложениях, поскольку требования к качеству различаются как среди классов входных изображений, так и среди классов конкретных систем.

Во-вторых, отсутствуют модели, методологии или рекомендации к созданию систем распознавания, учитывающих качество входных изображений с целью повышения точности, надежности, производительности и других характеристик систем.

В-третьих, для многих прикладных систем распознавания изображений и видеопотока актуальна не только традиционная постановка задачи оценки и коррекции качества входных изображений, но и ее вариации, актуальные при выборе лучшего кадра в видеопотоке, отказе в обработке поврежденных областей, интеграции результатов распознавания на разных кадрах и т.д.

Таким образом, дальнейшее исследование вопросов контроля качества изображений в системах распознавания является актуальным.

В диссертации исследуются три важных различных промышленных приложения систем распознавания: в системах документооборота и долгосрочного хранения документов, обеспечения контроля доступа и безопасности, а также распознавания изображений в непосредственном технологическом процессе.

Основные результаты диссертации были получены в процессе выполнения работ по следующим научным грантам РФФИ:

- №17-29-03170 – «Исследование быстродействующих методов и алгоритмов обработки изображений и оптического распознавания для использования в мобильных устройствах с ограниченной вычислительной производительностью»;
- №16-07-01051 – «Создание методологии разработки систем массового ввода и распознавания изображений документов»;
- №16-07-00616 – «Исследование возможности создания новых методов поиска нечетких дубликатов в видеопотоке»;
- №15-07-06520 – «Методы контроля подлинности документов и их фрагментов в гибридных системах обработки, передачи и хранения документов»;
- №14-07-00730 – «Математическое моделирование шумовых помех при распознавании»;
- №13-07-12173 – «Исследование методов сегментации изображений документов на структурные блоки методами цветового и морфологического анализа»;
- №13-07-12172 – «Распознавание документов удостоверяющих личность с помощью веб камер и камер мобильных устройств».

**Объектом исследования** являются системы оптического распознавания.

**Предметом исследования** является оценка качества входных изображений систем оптического распознавания.

**Целью** данной работы является разработка математических моделей и алгоритмов улучшения характеристик систем оптического распознавания путем анализа и учета неравномерности качества входных данных.

Для достижения поставленной цели необходимо было решить следующие **задачи**:

1. Исследовать влияние искажений входных изображений на промышленные системы распознавания и провести обзор современных методов оценки качества изображений.

2. Построить математическую модель системы распознавания изображений и видеопотока с модулями оценки качества и обратной связью на каждом этапе обработки.

3. Разработать алгоритм автоматической оценки качества цветовой сегментации изображений печатных документов.

4. Исследовать влияние детектирования и фильтрации бликов на точность поиска и распознавания документа на изображении, а также влияние бликов на устойчивость алгоритма вычисления взвешенного проективного оптического потока.

5. Разработать алгоритм детектирования и локализации периодических фоновых элементов известной структуры на изображении документа.

6. Разработать алгоритм контроля качества входного сигнала двухканального рентгенографического сепаратора.

7. Реализовать разработанные в диссертации алгоритмы для обеспечения их внедрения в промышленные системы распознавания и провести их экспериментальный анализ.

**Методология и методы исследования** основаны на системном анализе, математическом моделировании, методах цифровой обработки изображений и распознавания образов, Фурье-анализе.

**Основные положения, выносимые на защиту:**

1. Построена математическая модель системы распознавания изображений и видеопотока с модулями оценки качества и обратной связью на каждом этапе обработки.

2. Разработан алгоритм автоматической оценки качества цветовой сегментации изображений печатных документов, учитывающий особенности текстовых и графических типов информационных слоев.

3. Разработан алгоритм детектирования и локализации периодических фоновых элементов известной структуры на изображении документа, основанный на анализе Фурье-спектра, учитывающий модель документа с целью увеличения точности и устойчивости детектирования и локализации.

4. Показано положительное влияние подавления бликов на точность работы детектора документа методом Виолы-Джонса, на точность распознавания документов в видеопотоке, а также на устойчивость алгоритма вычисления взвешенного проективного оптического потока.

5. Разработан алгоритм контроля качества входного сигнала двухканального рентгенографического сепаратора на основе анализа входного рентгенографического изображения, позволяющий производить диагностику нештатных ситуаций и способный работать в реальном времени в процессе эксплуатации сепаратора.

**Научная новизна:**

1. Предложена новая математическая модель системы распознавания изображений и видеопотока с модулями оценки качества и обратной связью на каждом этапе обработки.

2. Разработан новый алгоритм автоматической оценки качества цветовой сегментации изображений печатных документов.

3. Разработан новый алгоритм детектирования и локализации регулярной пространственной помехи на изображении документа.

4. Выполнено оригинальное исследование воздействия подавления бликов на точность поиска и распознавания изображений документов, а также на устойчивость вычисления взвешенного проективного оптического потока.

5. Разработан новый алгоритм контроля качества и детектирования разладки регистрирующей системы оптического распознавания алмазов в рентгеновском диапазоне.

**Практическая значимость.** Реализации разработанных алгоритмов выбора наилучшего кадра для распознавания в видеопотоке, автоматической оценки качества цветовой сегментации изображений печатных документов, детектирования и фильтрации бликов на изображениях документов и детектирования и локализации периодических фоновых элементов известной структуры на изображении документа внедрены в программное обеспечение «Smart PDF/A», «Smart 3D OCR MRZ», «Smart PassportReader» и «Smart IDReader» компании ООО «Смарт Энджинс Сервис». Данные продукты интегрированы в информационные решения крупнейших российских банков, страховых компаний и организаций различных государственных структур, в том числе – международных. Реализация разработанного алгоритма контроля качества входного сигнала двухканального рентгенографического сепаратора внедрена в программный комплекс сепарации алмазов в рамках коммерческого проекта компании ООО «Визиллект Сервис». Предложенная модель системы распознавания изображений и видеопотока с модулями оценки качества и обратной связью использовалась при проектировании данных программных продуктов.

**Достоверность** полученных результатов обеспечивается согласованностью разработанных моделей и алгоритмов с экспериментальными результатами, их успешной апробацией и внедрением в коммерческие программные системы распознавания.

**Апробация работы.** Основные результаты работы докладывались на следующих конференциях и семинарах:

1. Seventh International Conference on Machine Vision (Milan, Italy, 2014),
2. Школе-конференции «Информационные технологии и системы – 2014» (Нижний Новгород, Россия, 2014),
3. 29th European Conference on Modelling and Simulation (Varna, Bulgaria, 2015),
4. Eighth International Conference on Machine Vision (Barcelona, Spain, 2015),
5. Школе-конференции «Информационные технологии и системы – 2015» (Сочи, Россия, 2015),
6. Техническом семинаре лаборатории зрительных систем ИППИ РАН (Москва, Россия, 2015),
7. Международном научно-исследовательском семинаре «Анализ и понимание изображений (Математические, когнитивные и прикладные проблемы анализа изображений и сигналов)» (Москва, Россия, 2017),
8. Школе-конференции «Информационные технологии и системы – 2017» (Уфа, Россия, 2017),
9. 14th IAPR International Conference on Document Analysis and Recognition (Kyoto, Japan, 2017),
10. 7th International Workshop on Camera Based Document Analysis and Recognition (Kyoto, Japan, 2017),
11. Tenth International Conference on Machine Vision (Vienna, Austria, 2017).

**Личный вклад.** Результаты, изложенные в диссертации, принадлежат лично автору. В совместных работах автор принимал непосредственное участие в выборе направлений и задач исследований, разработке и обсуждении результатов.

**Публикации.** Основные результаты по теме диссертации изложены в 24 публикациях, в том числе: 8 изданы в журналах, рекомендованных ВАК, 10 — в сборниках трудов конференций (8 работ в сборниках, входящих в международные базы цитирования Scopus или Web of Science), 1 патент на изобретение, 1 патент на полезную модель и 4 свидетельства о государственной регистрации программы для ЭВМ.



**Объем и структура работы.** Диссертация состоит из введения, пяти глав и заключения. Полный объем диссертации составляет 157 страниц, включая 63 рисунка и 3 таблицы. Список литературы содержит 206 наименований.

**В первой главе** описываются промышленные приложения систем распознавания изображений и анализируется их общее устройство. Рассматриваются типы искажений входных данных, приводящие к снижению точности и надежности систем распознавания. Приводится обзор существующих методов оценки качества изображений. Ставятся основные задачи диссертации.

**Во второй главе** приводится оригинальная модель системы оптического распознавания с встроенными модулями оценки качества входных данных, вводятся необходимые определения, сущности и связи между ними. В рамках модели рассматривается задача выбора наилучших кадров при распознавании в видеопотоке и описывается способ вычисления приоритета кадра с учетом его качества и времени его захвата. Представлены экспериментальные результаты для системы распознавания документов в видеопотоке при искусственно моделируемых условиях автоматической фокусировки камеры, демонстрирующие значительное увеличение точности и производительности системы при выборе кадра в соответствии с введенной функцией приоритета.

**В третьей главе** представлен алгоритм автоматической оценки качества цветовой сегментации изображений печатных документов, актуальный при их сжатии для долговременного хранения. Алгоритм основан на независимой оценке текстового слоя через точность распознавания и графического слоя с помощью вычисления разницы плотности принадлежащих слою пикселей на размеченной карте насыщенности изображения. Экспериментально подтверждается пригодность алгоритма для автоматической оптимизации параметров цветовой сегментации.

**В четвертой главе** рассмотрены приложения оценки качества изображений в системах распознавания при обеспечении контроля доступа и безопасности.

В разделе 4.1 приводится анализ негативного влияния бликов на работу системы поиска и распознавания документов, разрабатываются алгоритмы детектирования бликов и интерполяции поврежденного ими фона, экспериментально показывается увеличение точности поиска и распознавания документов после фильтрации бликов.

Раздел 4.2 посвящен проблеме бликов при вычислении проективного оптического потока, применяющегося для отслеживания движения плоских ригидных

документов в видеопотоке. Экспериментально подтверждается неустойчивость алгоритма к бликам в  $RGB$  пространстве и возможность его заметного улучшения при переходе в  $\alpha\beta$  пространство и исключении зашкаливающих областей.

В разделе 4.3 разработаны оригинальные алгоритмы детектирования и локализации периодических фоновых элементов известной структуры на изображении документа. Алгоритмы основаны на введенной математической модели изображения документа, предварительной обработке изображения и спектра двумерного дискретного преобразования Фурье, анализе спектра и итерационном решении нелинейной системы уравнений для получения итогового результата локализации. Экспериментальные результаты демонстрируют высокую точность алгоритмов детектирования и локализации. Приводятся результаты анализа устойчивости разработанных алгоритмов к ошибкам нахождения границ документа с использованием искусственного моделирования, показавшие устойчивость к небольшим ошибкам.

**В пятой главе** обсуждаются основные проблемы, возникающие при формировании входного сигнала рентгенографического сепаратора. Разработаны алгоритмы детектирования нештатных ситуаций для случаев раскалибровки пары источник/детектор, заслонения детектора и загрязнения конвейерной ленты, экспериментально показавшие высокую точность и позволяющие их применение в промышленных решениях для работы в реальном времени за счет высокой производительности и простоты реализации.



## **Глава 1. Искажения изображений и их влияние на системы распознавания**

### **1.1 Системы распознавания изображений**

Распознаванием изображений называется научная область, направленная на автоматическое определение закономерностей на изображениях с целью их дальнейшего использования в таких задачах, как классификация и идентификация объектов, характеризующихся набором некоторых свойств и признаков [1]. Распознавание изображений часто упоминается в контексте методов машинного зрения и является частным случаем распознавания образов, поскольку входными данными являются изображения или видеопоток [2]. Распознавание изображений находит приложения в различных дисциплинах: анализе документов и текстов, медицине, биометрии, обороне, промышленности, робототехнике и многих других. Таким образом, данная область является актуальной и важной при автоматизации процессов жизнедеятельности.

В диссертации особое внимание уделяется системам анализа документов, в которых распознавание необходимо практически на всех стадиях обработки изображения: детектирования и классификации документа, непосредственного оптического распознавания символов (Optical Character Recognition, OCR). Кроме того, будет рассмотрено приложение распознавания изображений в промышленности на примере рентгенографической сепарации алмазов из руды.

Распознавание изображений широко используется как непосредственно в основных технологических процессах производства, так и в неотъемлемых от нормальной работы предприятия вспомогательных. Рассмотрим некоторые из них.

#### **1.1.1 Документооборот и долгосрочное хранение документов**

Работа с документами является основной частью процессов управления и принятия управленческих решений на предприятиях и имеет место на протяжении всего процесса принятия управленческого решения: получения информации, ее переработки и анализа, подготовки и принятия решения [3].

Прикладные компоненты управления опираются на документооборот предприятия, который является документационным обеспечением управления и по [4] определяется, как процесс прохождения документов внутри системы управления. Система документооборота базируется на программно-технических платформах предприятия и включает средства и правила создания документов, поддержки их движения, электронный архив для долгосрочного их хранения и т.д. [5].

С точки зрения технологии о современном документообороте принято говорить, как об электронном, хотя часть документов и сохраняет бумажную форму на большинстве стадий своего жизненного цикла [6]. Электронные документы превосходят бумажные по многим характеристикам: время транспортировки документа, скорость поиска информации по атрибутам и содержанию документа, удобство восприятия информации, эффективность внесения временных пометок, создания копий, а также сохранность и долговечность хранения документов.

Методы распознавания образов и обработки изображений широко используются при переводе документов из бумажного в электронный вид: автоматическом вводе, архивировании и сжатии для долгосрочного хранения, что будет рассмотрено более подробно в разделе 1.1.5.

### **1.1.2 Контроль доступа и безопасности на предприятии**

Одной из важнейших задач обеспечения безопасности жизнедеятельности человека является контроль и управление перемещением людей или предметов по определенным маршрутам и зонам [7], например, контроль допуска служащих на предприятие, обнаружение выноса неоплаченных товаров или проноса к неразрешенным к провозу предметов и т.д.

Системы контроля управления доступом (СКУД) предназначены для решения ряда задач [8]: противодействие промышленному шпионажу, воровству, саботажу, умышленному повреждению материальных ценностей, учет рабочего времени, контроль своевременности прихода и ухода сотрудников, защита конфиденциальности информации, регулирование потока посетителей, контроль въезда и выезда транспорта.

Удостоверяющие личность документы [9] содержат требуемую информацию для идентификации владельца документа, такую, как его имя, фамилию, фотографию, и являются распространенным идентификатором доступа. За последнее десятилетие технологии распознавания документов развились до такого уровня надежности, что стали стандартными компонентами информационных систем, в том числе – на мобильных устройствах и установках с малоразмерными цифровыми камерами [10].

Вместе с классическими идентификаторами личности в виде материальных носителей (ключей, карт, документов, автомобильных номеров) или знаний (например, паролей) популярность набирают биометрические технологии идентификации личности [11], такие, как отпечатки пальцев, форма кисти, радужная оболочка глаза, форма лица, почерк, походка, и другие [12]. Преимуществами использования биометрической информации в СКУД являются трудности их подделки и невозможность их утери или передачи другому человеку [13]. С другой стороны, особенности биометрической информации приводят к существенно повышенным требованиям к надежности используемых алгоритмов распознавания образов и к их быстродействию.

### **1.1.3 Распознавание изображений в технологических процессах**

Системы машинного зрения, также известного как технологическое зрение, надежно закрепили свои позиции в промышленном секторе, используя видеокамеры как основной источник сигнала и программное обеспечение с алгоритмами распознавания образов [14]. В процессе конвейерной сборки системы машинного зрения сообщают роботу ориентации деталей для их корректного соединения, исследуют оптоволоконно на предмет искажений, осколков и изломов и имеют множество других приложений [15].

Распознавание образов в технологических процессах промышленности позволяет решать множество задач, которые условно можно классифицировать на несколько групп [16].

Во-первых, это автоматическая визуальная инспекция, т.е. подтверждение определенных свойств, измерение различных физических параметров (линейный размер, диаметр, кривизна, площадь, высота и количество и других), контроль

качества, определение дефектов. Примерами являются определения наличия или отсутствия этикетки на бутылке, болтов для проведения операции сборки.

Во-вторых, это контроль и поддержка процесса производства: скоростную сортировку продуктов, контроль пикселей на дисплеях, поворот заготовок для кредитных карт нужной стороной и т.п.

В-третьих, это определение пространственного местоположения или статического положения объекта и передача информации о положении и ориентации объекта в систему управления или контроллер. Примером такого приложения может служить погрузочно-разгрузочный робот, перемещающий объекты различной формы из бункера.

В-четвертых, это идентификация и классификация, т.е. считывание различных кодов или буквенно-цифровых обозначений с целью отбраковки просроченных продуктов, автоматическая сортировка продуктов и автомобильных запчастей, разгрузка деталей из контейнеров и т.д.

#### 1.1.4 Архитектура систем распознавания

Системы распознавания и машинного зрения применяются среди многих приложений и имеют различное внутреннее устройство и архитектуру, но в их структуре можно выделить функциональные блоки, присущие большинству из них. Можно сказать, что общий подход к построению архитектур систем распознавания остается неизменным на протяжении нескольких десятков лет [15; 17]. На рисунке 1.1 показана обобщенная схема работы систем распознавания, состоящая из последовательных этапов. Рассмотрим их подробнее.

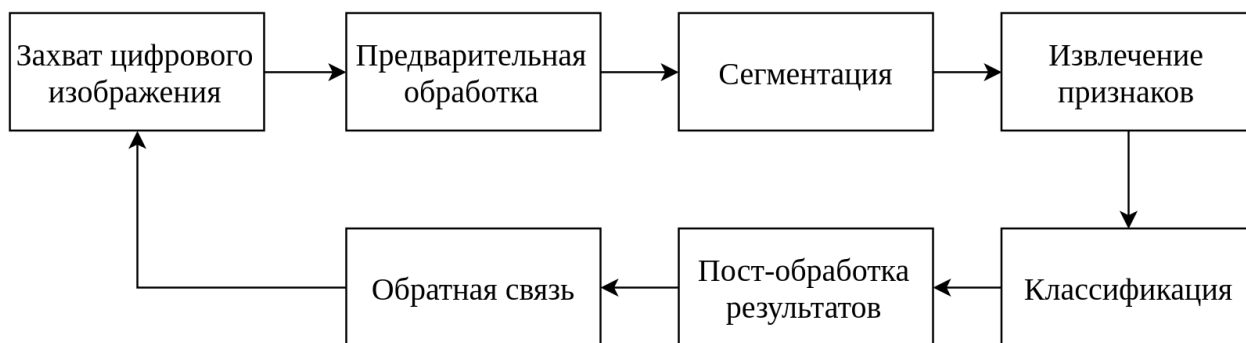


Рисунок 1.1 — Обобщенная схема работы систем распознавания

В любой системе оптического распознавания первым этапом является захват цифрового изображения сцены, освещенной некоторым источником света, предположительно содержащей распознаваемый объект – как со сканера, камеры мобильного устройства или стационарной камеры, так и с других типов сенсоров. Следующим этапом является предварительная обработка изображения, например, приведение его к фиксированному размеру, конвертация цветового пространства, выравнивание баланса белого, яркости, контраста, резкости и другие способы обработки, с целью приведения изображения в более удобный для последующей работы вид [18].

Затем, производится этап сегментации [19], цель которого – выделить регионы изображения, соответствующие объектам на сцене или их участкам. Результатом сегментации в общем случае являются битовые маски принадлежности регионов изображения объектам, но если объекты рассматриваются отдельно друг от друга на дальнейших этапах обработки, то возможно сопоставить каждому объекту отдельное его изображение: выделить зону интереса ограничивающим прямоугольником или выполнить проективную коррекцию для объектов прямоугольной формы. При этом производится смена координатной системы относительно полученных изображений.

Процесс сегментации имеет иерархичную структуру в случае вложенности объектов интереса друг в друга. Например, при построении систем распознавания лиц [20] до непосредственного детектирования лиц происходит сегментация участков кожи и их объединение с помощью морфологических операций. В системе распознавания документов [21] сначала определяется местоположение документа и идентифицируется его тип с помощью модификаций метода Виолы-Джонса, методов особых точек и дескрипторов, классических методов поиска границ и других. Когда документ найден, производится его проективная коррекция и выделяются фиксированные зоны, на которых, с помощью методов математической морфологии, ищутся текстовые поля. Наконец, найденные поля передаются в модуль сегментации полей, возвращающий отдельные изображения символов.

Изображения или участки найденных на этапе сегментации объектов передаются на этапы извлечения признаков и классификации: распознавания или получения дескрипторов объектов для их последующей идентификации [22]. Современные подходы, например, с использованием глубоких сверточных нейронных сетей, позволяют совместить эти два этапа в один путем отказа от

«ручных» способов извлечения признаков на этапе классификации. Однако, использование последних все еще актуально во многих приложениях.

Финальным в рамках распознавания одного кадра этапом является пост-обработка результатов классификации. Примерами такой пост-обработки являются контекстная коррекция распознанного текста [23] или речи [24] в соответствии с моделью языка, проверка контрольных сумм на стандартизованных документах, а также другие обработки, использующие семантические правила предметной области. После этого этапа происходит возвращение текущего результата пользователю вместе с обратной связью системы о возможных улучшениях условий съемки: рефокусировки или изменении угла камеры, контроль освещения и т.п.

При разработке систем распознавания в видеопотоке открывается как ряд возможностей повышения качественных характеристик системы, недоступных при распознавании одного кадра, так и ряд новых задач [21]. Например, становится возможной интеграция результатов на текущем кадре с результатами распознавания предыдущих кадров на этапе пост-обработки, отслеживание положения объектов в пространстве и т.д. С другой стороны, встает вопрос о том, когда останавливать процесс распознавания.

В конце обработки каждого из этапов, ответственная за этап подсистема может выдать как корректный результат обработки вместе с скалярным уровнем своей уверенности в его корректности, так (для некоторых подсистем) и сигнал об отказе в обработке – при низком качестве рассматриваемого изображения или низкой уверенности в результате на текущем шаге. Преимуществ у отказа при заведомо невозможной дальнейшей обработке несколько [25]: экономится время, некорректные результаты не проходят дальше и не снижают качество работы системы, что наиболее актуально при интеграции результатов во время распознавания в видеопотоке.

### **1.1.5 Системы сжатия изображений документов**

Рассмотрим пример строения системы сжатия и распознавания изображений печатных документов, использующейся при автоматическом вводе,



архивировании и долгосрочном хранении документов в электронном виде, актуальность чего уже обсуждалась в разделе 1.1.1.

Несмотря на снижающуюся стоимость хранителей информации (жестких дисков и т.д.) и популярность облачных хранилищ, проблема эффективного сжатия печатных документов остается актуальной. Изображение документа, сжатое стандартными (например, JPEG [26]) алгоритмами, может занимать несколько мегабайт, что непригодно для долговременного хранения и чревато искажением текстовой информации (рисунок 1.2), из-за чего был разработан комбинированный подход.



Рисунок 1.2 — Результаты сжатия текста документа разными алгоритмами и артефакты алгоритма JPEG

### 1.1.5.1 Модель смешанного растрового содержимого

Модель смешанного растрового содержимого (Mixed Raster Content, MRC) – модель, применяющаяся в системах сжатия для эффективного кодирования комбинированных (содержащих текст и графику) растровых изображений. Ее смысл в том, чтобы разбить изображение на непересекающиеся слои, каждый из которых содержит в себе объекты определенного класса: текст, изображения или фон. Такой подход позволяет значительно повысить эффективность сжатия за счет применения оптимальных алгоритмов сжатия для каждого слоя в отдельности и последующего их комбинирования.

Базовая трехслойная модель MRC (рисунок 1.3) представляет цветное растровое изображение как два цветных растровых слоя (изображение переднего плана и фон) и слой битовой маски. Маска требуется, чтобы восстановить

итоговое изображение из этих двух слоев: значение пикселя 1 говорит, что соответствующий пиксель должен принадлежать изображению переднего плана, а 0 – наоборот, фонового [27].

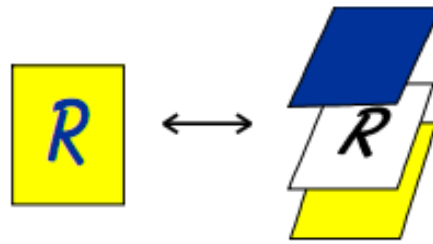


Рисунок 1.3 — Базовая трехслойная модель MRC

Идеология MRC применяется в технологии DJVU, где возможно сжатие с 25 мегабайтного изображения в 300 dpi до 30-80 килобайт [27; 28], а также в стандарте PDF/A (англ. Portable Document Format / A), предназначенном для долгосрочного хранения электронных документов.

**Технология сжатия PDF/A** Опишем одну из технологий сжатия изображений цветных печатных документов в формат PDF/A, предложенную в [29] (рисунок 1.4). Первым этапом обработки является расслоение исходного изображения, в результате которого изображение разбивается на два непересекающихся «слоя» или изображения с текстом и графикой, кодирующимися единичными и нулевыми значениями битовой маски принадлежности слою. Затем, производится распознавание текстового слоя с помощью OCR-систем, после чего происходит сжатие каждого слоя соответствующим образом (алгоритмами JPEG и JBIG2) и их упаковка в PDF/A файл.

**Цветовая модель HSI** Для описания процесса разбиения изображения на слои требуется ввести цветовую модель HSI, в которой цветовой тон (Hue) описывает базовый цвет, насыщенность (Saturation) – мера степени разбавления белым базового цвета, а световая интенсивность (Intensity) – субъективная характеристика, которая практически не поддается измерению, но соответствует полутонной яркости в ахроматическом случае [30]. Преобразование из RGB компонент производится по следующим формулам [31]:

$$\theta = \arccos \frac{(R - G) + (R - B)}{2 \cdot ((R - G)^2 + (R - B)(G - B))^{1/2}}, \quad (1.1)$$



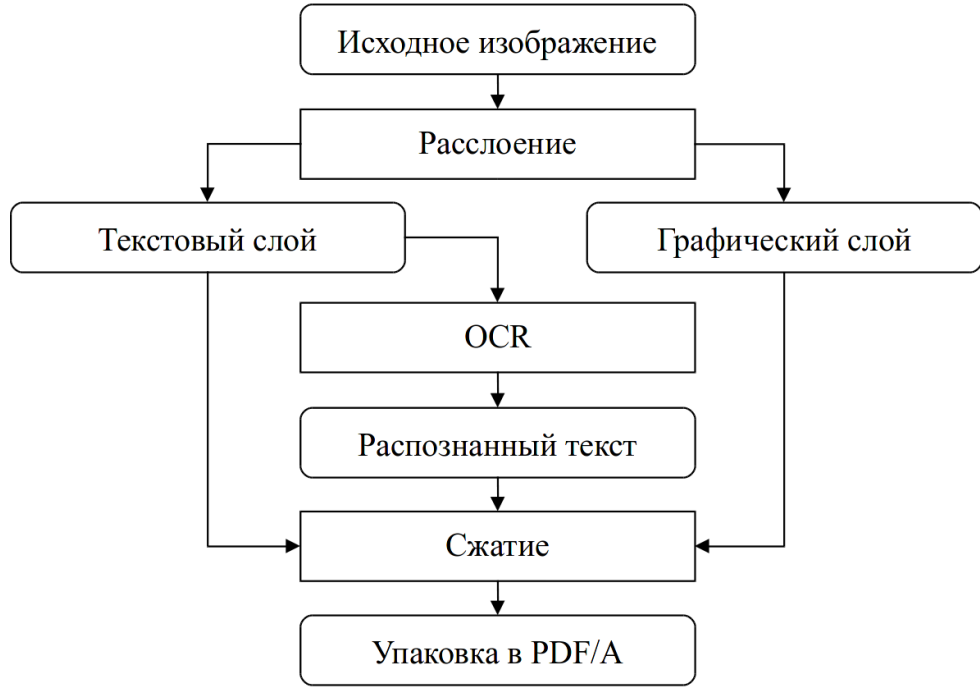


Рисунок 1.4 — Процесс обработки и сжатия изображения в формат PDF/A

$$S = 1 - \frac{3 \cdot \min(R, G, B)}{R + G + B}, \quad I = \frac{1}{3}(R + G + B), \quad H = \begin{cases} \theta, & B \leq G, \\ 360 - \theta, & B > G, \end{cases} \quad (1.2)$$

где  $H$ ,  $S$ ,  $I$  – значения цветового тона, насыщенности и интенсивности, а  $R$ ,  $G$ ,  $B$  – значения красной, зеленой и синей компоненты соответственно.

**Схема расслоения изображения цветного печатного документа** В цветовой системе HSI черные области, соответствующие черному тексту, будут иметь близкий к нулю показатель насыщенности, тогда как графическая составляющая документа, наоборот – высокую насыщенность. Поэтому, построим логарифмическую гистограмма цветовой насыщенности изображения, т.е. зависимость  $h(x) = \log N_x$ , где  $N_x$  – количество пикселей изображения с показателем насыщенности, равным  $x$ . Найдем порог разделения  $t^*$  двух выделяющихся на гистограмме классов насыщенности классическим методом Оцу [32]:

$$t^* = \arg \max \sigma_b^2(t), \quad \sigma_b^2(t) = \omega_0(t)\omega_1(t)(\mu_0(t) - \mu_1(t))^2 \quad (1.3)$$

$$\omega_0(t) = \sum_{i=0}^t \frac{N_i}{N}, \quad \omega_1(t) = \sum_{i=t+1}^L \frac{N_i}{N}, \quad \mu_0(t) = \sum_{i=0}^t \frac{iN_i}{N\omega_0}, \quad \mu_1(t) = \sum_{i=t+1}^L \frac{iN_i}{N\omega_1}, \quad (1.4)$$

где  $t^*$  – оптимальный порог разделения,  $\sigma_b^2(t)$  – межклассовая дисперсия,  $\omega_0$  и  $\omega_1$  – относительные частоты классов,  $\mu_0$  и  $\mu_1$  – средние уровни классов,  $N_i$  – количество пикселей с насыщенностью  $i$ ,  $L$  – максимальный порог насыщенности.

Зададим минимальное значение порога  $t_{\min}$ , и, если  $t^*$  оказался меньше  $t_{\min}$ , то будем считать, что изображение содержит только черный текст и не содержит графических областей. Иначе, сгенерируем битовую маску принадлежности соответствующего пикселя текстовому или графическую слою, сравнив значение его насыщенности с  $t^*$ .

Правильное разбиение на слои является важнейшим этапом работы системы и качество результирующего упакованного изображения документа зависит, в основном, именно от него.

### 1.1.6 Постановка задач распознавания

Основными задачами распознавания изображений, возникающими в практических приложениях, являются задачи детектирования, локализации и классификации объектов на изображении. При анализе видеопотока к предыдущим задачам добавляется возможность постановки специфических задач, например отслеживания объекта [33].

Пусть имеется образ (изображение)  $I$  и конечный набор классов  $A = \{C_1, \dots, C_M\}$  размера  $M$ , также называемый алфавитом распознавания. Классификацией называется сопоставление образу  $I$  вектора альтернатив  $\vec{a}$  [34]:

$$C(I) = \vec{a} = (a_1, \dots, a_M), \quad a_k = (i_k, g_k), \quad (1.5)$$

$$i_k \in \{1, \dots, M\}, \quad g_i \in [0, 1], \quad (1.6)$$

где  $(i_k, g_k)$  – оценка («уверенность» классификатора)  $g_k$  принадлежности образа к  $C_{i_k}$ -му классу из  $A$ . Для удобства, вектор  $\vec{a}$  можно отсортировать по убыванию  $g_k$ , чтобы считать первую альтернативу  $i_1$  результатом распознавания (классификации).

При решении задачи детектирования требуется определить, находится ли заданный объект на изображении без необходимости определения его непосредственного местонахождения:

$$D(I) = g, \quad g \in [0, 1], \quad (1.7)$$

где  $g$  – оценка присутствия объекта на изображении. Задача детектирования является частным случаем классификации на два класса: положительный и отрицательный.

Если при детектировании был получен положительный результат или заранее известно, что объект заданного класса на изображении присутствует, то возможно поставить задачу локализации, т.е. определения непосредственного местонахождения объекта или объектов:

$$L(I) = \vec{x} = (x_1 \dots x_X), \quad x_k = (l_k, g_k), \quad g \in [0, 1], \quad (1.8)$$

где  $l_k$  – координаты  $k$ -го объекта на образе  $I$ , а  $g_k$  – как и ранее, «уверенность» метода в данной альтернативе вектора альтернатив  $\vec{x}$ .

#### 1.1.6.1 Распознавание в видеопотоке

Видеопотоком  $V$  размера  $N$  называется последовательность изображений (кадров), упорядоченных во времени:

$$V = (I_1, \dots, I_t, \dots, I_N). \quad (1.9)$$

В качестве дополнительного ограничения будем считать, что кадры видеопотока были получены с одного и того же устройства съемки.

Задачи распознавания в видеопотоке аналогичны приведенным ранее задачам распознавания на отдельных изображениях, но при распознавании  $I_t$ -го кадра можно учитывать все предыдущие. Например, задача классификации 1.5 может быть поставлена так:

$$C(I_1, \dots, I_t) = \vec{a} = (a_1, \dots, a_M). \quad (1.10)$$

Важными преимуществами распознавания объектов в видеопотоке по сравнению с их распознаванием на единичных изображениях являются возможность выбрать в каком-то смысле наилучшие кадры, построить единую систему координат, интегрировать результаты на различных кадрах и другие [35].

### 1.1.7 Надежность распознавания

Задачи распознавания могут быть поставлены и без необходимости возвращать оценки  $g_k$ . Однако, оценки уверенности распознавателя позволяют проводить дополнительный анализ результата, используя оценки как распределение вероятностей соответствующих результирующих классов – в частности, при анализе достоверности результатов распознавания, что будет рассмотрено далее. К тому же, данные оценки возникают естественным образом при использовании методов машинного обучения, к примеру, нейронных сетей, являющихся одним из наиболее распространенных классификаторов в различных задачах распознавания [36].

#### 1.1.7.1 Постановка задачи минимизации ожидания ошибки

Пусть имеется пространство признаков (входных изображений)  $\mathcal{X} = \mathbb{R}^d$ , пространство ответов (результатов распознавания)  $\mathcal{Y} = \mathbb{R}^q$  и пространство  $\mathcal{F}$  распознавателей  $f : \mathcal{X} \rightarrow \mathcal{Y}$ . Также имеется обучающий набор  $T = ((x_1, y_1), \dots, (x_N, y_N))$ , полученный из некоторого неизвестного распределения  $\pi$  и состоящий из пар векторов  $(x_i, y_i) \in \mathcal{X} \times \mathcal{Y}$  входных признаков и соответствующих им результатов. Наконец, имеется функция потерь или штрафа  $E : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$ , как правило, неотрицательная и равная 0 при совпадении первого параметра (прогнозируемого результата) и второго (истинного ответа). Примером функции потерь может быть квадратичный штраф:  $E(a, b) = \|a - b\|^2$ . Тогда, задача построения распознавателя  $f \in \mathcal{F}$  ставится следующим образом [37]:

$$E_\pi(f) = \int_{(x,y) \in \mathcal{X} \times \mathcal{Y}} E(f(x), y) d\pi(x, y) \rightarrow \min_{f \in \mathcal{F}}, \quad (1.11)$$

тем самым минимизируя математическое ожидание потерь по распределению  $\pi$ . Так как распределение  $\pi$  неизвестно, на практике ожидание штрафа 1.11 приближают суммой ошибки по обучающему набору  $T$ :

$$E(f, T) = \frac{1}{N} \sum_{i=1}^N E(f(x_i), y_i) \rightarrow \min_{f \in \mathcal{F}}. \quad (1.12)$$

При классификации, векторы  $y_i$  обучающего набора  $T$  составляются так, что содержат нули на всех позициях, кроме позиций, соответствующих верным классам для  $x_i$  – там содержится единица. В процессе обучения, т.е. минимизации функционала 1.12, проводится выбор или настройка параметров распознавателя  $f$ . Понятно, что для минимизации ошибки распознаватель будет стараться выдавать ответы со значениями, близкими к нулю в позициях неверных классов и близкими к единице в позициях верных. Отметим, что нормировка на  $[0, 1]$  часто достигается в выходных векторах моделей за счет использования softmax-преобразования [1]:

$$\sigma_j(y_i) = \frac{e^{y_{i,j}}}{\sum_{k=1}^q e^{y_{i,k}}}, \quad (1.13)$$

что позволяет считать значения  $y'_{i,j} = \sigma_j(y_i)$  вероятностными оценками [38] принадлежности  $i$ -го образа к  $j$ -му классу. Ограничимся рассмотрением векторов альтернатив, содержащими оценки вероятностной природы.

#### 1.1.7.2 Определение достоверности результатов распознавания

Приведенная постановка задачи обучения модели  $f$  не учитывает многих деталей, необходимых для построения эффективных методов распознавания (например, борьбы с переобучением на обучающей выборке), но позволяет заметить особенность, присущую подобным моделям. А именно, поскольку построить систему распознавания, имеющую идеальную точность работы для любых входных данных практически невозможно, требуется способ определения достоверности или надежности распознавания, что по умолчанию не предоставляется системой [34].

Задача определения надежности распознавания исследуется на протяжении многих лет и актуальна до сих пор [34; 39–41]. Под задачей определения надежности [41] результатов распознавания понимается построение правил отбраковки ошибочных результатов, т.е. двухклассовой классификации результатов на надежные и ненадежные, используя оценки (в частности, вероятностные) принадлежностей к соответствующим классам вектора альтернатив.

Ошибки самих правил отбраковки делятся на два типа: ошибки первого рода, когда результат признан недостоверным при верном распознавании объекта и, наоборот, ошибки второго рода – результат признан достоверным при неверном распознавании объекта. Пусть  $P_1$  и  $P_2$  – апостериорные вероятности ошибок первого и второго рода, а  $W_1$  и  $W_2$  – стоимости ошибок первого и второго рода, соответственно. Стоимости ошибок зависят от конкретных систем распознавания. Тогда, задача построения правил отбраковки может быть поставлена следующим образом [34]:

$$P_1W_1 + P_2W_2 \rightarrow \min. \quad (1.14)$$

### 1.1.7.3 Проблема оценок уверенности распознавания

Так как правила отбраковки руководствуются оценками выходных векторов альтернатив, то критическим требованием для их эффективной работы является адекватность вероятностных оценок апостериорному распределению результатов при этих оценках [42; 43]. Распределение выходных оценок зависит как от устройства модели распознавания, так и от ее параметров, подбираемых в процессе обучения – минимизации функции потерь, например, методом стохастического градиентного спуска.

Распространенной проблемой является чрезмерная «уверенность» распознавателя в выдаваемых им оценках [44], что ведет, помимо прочего, к невозможности применения методов отбраковки, тем самым понижая надежность распознавания. Подобное поведение происходит вследствие того, что результирующие оценки не имеют по своей природе вероятностного характера, что проявляется не только при назначении слишком высоких оценок неверным классам, но и наоборот – при назначении слишком низких оценок верным классам. Повышение способности нейронных сетей производить более гладкое распределение оценок широко исследуется как независимый от отбраковки метод регуляризации [45] для борьбы с переобучением и предотвращения попадания в локальные минимумы при обучении, подтверждая актуальность проблемы.

#### 1.1.7.4 Повышение надежности систем распознавания

В работе [44] была предложена модель релевантности распознавателя, позволяющая контролировать количество получаемой от него информации. Контроль информации производится за счет регулировки значений оценок вектора альтернатив с помощью преобразования, аналогичного 1.13:

$$\sigma_j(y_i, g, \tau) = \frac{e^{y_{i,j}g_j\tau}}{\sum_{k=1}^q e^{y_{i,k}g_k\tau}}, \quad (1.15)$$

где  $g_j$  – поправки на известные распределения классов [38], а  $\tau \in [0, 1]$  – параметр релевантности. Преобразование 1.15 можно рассматривать, как «гамма-коррекцию» вектора псевдо-вероятностных оценок.

Важной особенностью предложенного в работе [44] метода вычисления параметра  $\tau$  для конкретного применения в задаче распознавания символов является использование оценок неравномерности качества входных изображений, таких, как дефокусировка, смаз и контрастность. Результаты работы [44] демонстрируют повышение точности распознавания за счет увеличения влияния работы остальных подсистем (например, контекстной пост-обработки) для символов с низкой оценкой качества входных изображений. Заметим, что использование оценок качества изображений в качестве предиктора достоверности результата также широко исследуется на протяжении долгого времени [46].

Контроль качества входных данных интересен в более общем контексте повышения надежности систем распознавания, поскольку общей проблемой снижения надежности является потенциально низкое качество входных данных, приводящее к заведомо непредсказуемым результатам на выходе [47].

Таким образом, для разработки точных и надежных систем распознавания необходимо проводить оценку неравномерности качества входных данных. В том числе, это позволяет контролировать количество информации, получаемой от методов (подсистем) распознавания. Обзор основных причин снижения качества входных данных приведен в следующем разделе.



## 1.2 Типы искажений входных данных систем распознавания

Входные изображения систем оптического распознавания могут подвергаться большому количеству различных искажений, особенно в неконтролируемых или естественных условиях съемки [48]: дефокусировка, смаз, цифровой шум, разладка регистратора, артефакты сжатия, блики, проективные искажения, повреждения объекта обработки и т.д. Рассмотрим подробнее типы искажений и их влияние на системы распознавания.

### 1.2.1 Шум и аберрации

Шум неизбежно появляется при регистрации изображений устройствами с фотосенсорами и заметен в виде наложенной маски пикселей случайного цвета и яркости [10]. Шум появляется в процессе оцифровки аналогового сигнала в результате теплового шума, ошибок квантования сигнала и переноса заряда на матрице с последующим усилением сигнала, а также из-за загрязнений матрицы, дефектных элементов сенсора и дальнейшего применения алгоритмов сжатия или обработки изображений. Наиболее сильное проявление шума наблюдается на однотонных участках изображения, в особенности – на темных.

В результате дефектов видеокамеры, при съемке могут проявляться различные аберрации оптических систем, например, радиальная или тангенциальная дисторсия [49]. Радиальная дисторсия возникает из-за неидеальной формы линзы и приводит к тому, что прямые линии на изображении по мере удаления от центра становятся более выпуклыми, что также является источником эффектов «бочка» или «рыбий глаз». При радиальной дисторсии искажения отсутствуют в оптическом центре сенсора и увеличиваются по мере приближения к периферии. Производственные дефекты, вызывающие тангенциальную дисторсию, приводят к неперпендикулярности расположения экрана к оптической оси объектива.



### 1.2.2 Сжатие и интерполяция изображений

Методы сжатия изображений применяются во множествах приложений при хранении и передаче изображений. В целях сокращения передаваемой информации наиболее популярными являются методы сжатия с потерями, использующие визуальную избыточность изображений, т.е. информацию, которую можно изъять из изображения, не сильно нарушая визуально воспринимаемое качество изображений [50]. Тем не менее, при сильном сжатии возможна чрезмерная потеря качества результирующего изображения.

Наиболее распространенными при сжатии изображений с потерями являются методы семейства JPEG [26], получившие популярность из-за высокой степени сжатия и невысокой вычислительной сложности, хотя с момента появления оригинального стандарта алгоритма разработано большое количество других методов [51]. Данный класс алгоритмов основан на дискретном косинусном преобразовании и оперирует квадратными блоками изображения размером 8 пикселей, что при высокой степени сжатия приводит к видимым артефактам специфичного вида. Непригодность JPEG-алгоритмов при сжатии изображений текстов более подробно рассмотрена в разделе 1.1.5.

Схожей проблемой, возникающей при изменении размера или разрешения изображения, являются артефакты интерполяции, приводящие к искажению или размытию границ объектов на изображении [52].

### 1.2.3 Дефокусировка и размытость

В зависимости от характеристик объектива и взаимного расположения камеры и объекта, изображение может быть частично или полностью размыто в результате дефокусировки. Также, движение объектов или камеры относительно друг друга чревато появлением размытости из-за смещения объекта во время экспозиции [10].

При моделировании степени дефокусировки большинство работ используют модель толстой линзы [53], устройство которой показано на рисунке 1.5.

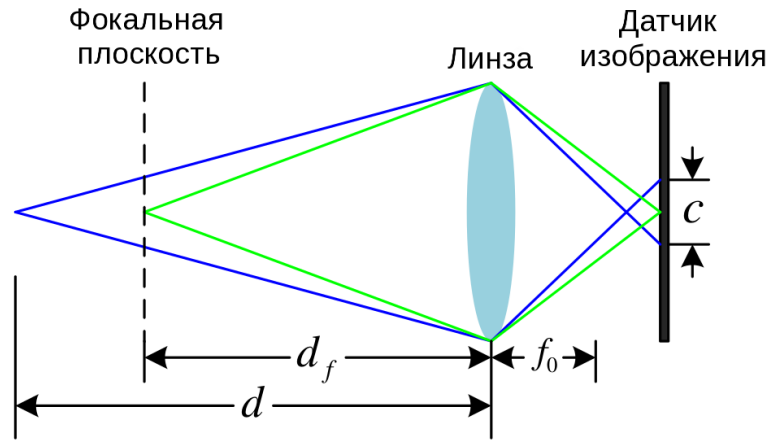


Рисунок 1.5 — Модель толстой линзы [54]

Когда объект находится на расстоянии  $d_f$  до главной фокальной плоскости, все лучи из точки объекта сходятся в одной точке датчика, благодаря чему изображение выглядит четким. Лучи от объекта, находящегося на другом расстоянии  $d$ , попадают в множество точек датчика, что приводит к размытости изображения. Форма областей размытости зависит от формы апертуры и называется кружком (пятном) рассеивания [53]. Диаметр кружка рассеивания характеризует степень дефокусировки и может быть вычислен, как:

$$c = \frac{|d - d_f|}{d} \frac{f_0^2}{N(d_f - f_0)}, \quad (1.16)$$

где  $f_0$  – фокусное расстояние камеры, а  $N$  – ее относительное отверстие. Таким образом, диаметр кружка рассеивания монотонно и нелинейно увеличивается с увеличением расстояния  $d$  до объекта.

Эффект размытия границы  $i(x)$  в результате дефокусировки может быть смоделирован [54] как свертка идеально сфокусированного изображения  $f(x)$  с функцией точечного распределения, представляющей пятно рассеивания и обычно приближающейся Гауссовым распределением  $g(x, \sigma)$ :

$$i(x) = f(x) * g(x, \sigma). \quad (1.17)$$

Среднеквадратичное отклонение  $\sigma = kc$  регулирует степень дефокусировки и пропорционально диаметру  $c$  пятна рассеивания.

### 1.2.4 Тени, блики и голографические элементы

При съемке камерой в естественных сценах на изображениях часто возникают перепады яркости, вызванные тенями и бликами – особенно, на объектах с глянцевой поверхностью. Данный тип искажений приводит к потере существующих или появлению фальшивых границ объектов, затрудняющих обработку изображения и распознавание [10]. Кроме того, из-за ограничений динамического диапазона камер, область блика может быть полностью зашкалена, делая невозможным восстановление изображения в этой области и означая необратимые потери информации. Смежным типом помех являются голографические элементы, использующиеся, например, в качестве защиты документов от подделок [55]. Будучи нанесенными поверх текстового заполнения, они изменяют характеристики распознаваемых областей изображений и заслоняют объекты интереса.

Если распознавание происходит в видеопотоке, то перепады яркости, вызванные описанными помехами, могут проявляться только на некоторой части кадров. Однако, помимо возможных необратимых потерь информации, их «движение» по объекту нарушает общий вид объекта между соседними кадрами, что усложняет отслеживание объекта и построение единой системы координат.

### 1.2.5 Механическая деформация и артефакты объекта

Наконец, помимо искажений, вызываемыми «внешними» причинами по отношению к объекту распознавания, сам объект может содержать дефекты, негативно влияющие на процесс распознавания. Физические объекты подвергаются механическим деформациям [10], изменяющим их структуру и значительно усложняющим разработку модели их поверхности, поскольку подобные деформации нелинейны и сложно описываются математической моделью объекта. Например, в результате механического изгиба плоского документа, его ранее параллельные строки текста начинают изгибаться [48].

Смежным классом искажений являются артефакты на поверхности объекта в результате его износа, низкого качества производства: плохо пропечатанные

на бумаге или вытесненные на пластике или металле символы текста, опознавательные знаки, мятая страница книги и т.п. Кроме того, на объекте могут присутствовать непредвиденные начертания или гравировка, нанесенные как случайно, так и намеренно – для мошеннических целей и обмана систем распознавания.

### **1.2.6 Неустойчивость систем распознавания к искажениям**

На основании обзора основных искажений входных изображений систем распознавания можно сказать, что они являются актуальной проблемой и могут приводить к полной потере информации. Рассмотрим подробнее проблему неустойчивости работы систем распознавания в присутствии искажений.

Особое внимание проблеме низкого качества входных изображений уделено в области биометрии. В работах [56] и [57] приводится анализ неустойчивости систем верификации и распознавания лиц к перепадам яркости. В работе [58] исследуется неустойчивость систем распознавания лиц к изменениям позы человека с целью обеспечения сигнала отказа в распознавании.

Исследования показывают, что современные глубокие нейронные сети оказываются неустойчивыми к размытию [59] и шуму [60]. В исследовании [61] рассмотрена проблема неустойчивости нейронных сетей при распознавании небольших вырезанных регионов изображений в сравнении с возможностями человека на той же задаче.

Также в последние годы активно исследуются возможности «обмана» глубоких нейронных сетей искусственно модифицированными или созданными целиком изображениями. Например, в работе [62] показывается, что нейронная сеть может выдавать на 99.99% уверенные ошибочные результаты распознавания на изображениях со случайным шумом или на модифицированных примерах, что подтверждает проблему чрезмерной уверенности в неверном результате, описанную ранее. Подобное ошибочное поведение наблюдается даже если модифицированное изображение было напечатано на принтере, а затем повторно снято камерой [63]. Более того, бывает достаточно изменить всего лишь один или несколько пикселей изображения для достижения такого эффекта [64]. Вместе с этим развиваются и способы повышения устойчивости к подобным методам [65].

Таким образом можно сделать вывод, что системы распознавания, в том числе использующие нейронные сети, могут быть неустойчивы к непредвиденным пертурбациям или искажениям, присутствующим на входных изображениях.

### **1.3 Оценка качества изображений**

Как было показано в разделе 1.2, изображения подвергаются различным типам искажений, ухудшающих их качество. Качеством изображения называется характеристика, измеряющая его наблюдаемую и воспринимаемую деградацию [66]. Для того, чтобы поддерживать, контролировать и улучшать качество изображений, системам их передачи, управления, получения и обработки важно оценивать качество работы на каждой из стадий обработки. По этой причине, оценка качества изображений играет важную роль в большинстве приложений обработки изображений, а соответствующая область активно развивается в последние десятилетия, развившись до отдельной отрасли науки обработки изображений.

Большая сложность в оценке качества изображений обусловлена практической невозможностью описания типа и количества ухудшения в восстановленном изображении [67]. Именно по этой причине, несмотря на значительное количество исследований по теме оценки качества изображений, до сих пор не разработано универсального способа оценки качества изображений [68].

#### **1.3.1 Виды и приложения методов оценки качества изображений**

Методы оценки качества изображения могут быть разделены на две категории: субъективные и объективные. Субъективные методы оценки качества изображений основываются на экспертном мнении людей, тогда как цель исследования объективных методов – разработка математических моделей, способных автоматически и с высокой точностью предсказать качество изображения.

Субъективные методы оценки качества обеспечивают высокую точность, но требуют много времени, тяжелы и дороги в проведении, что делает их невозможными или трудными в использовании на практике [69]. Объективные методы оценки качества являются автоматическими, что переводит их на следующий уровень применимости по сравнению с субъективными. Идеальный объективный метод оценки должен быть способен имитировать мнения людей-экспертов [70]. В настоящее время не известно объективной метрики, предсказывающей качество изображения с практически абсолютной точностью [71]. Однако, результаты и общий прогресс в разработке объективных критериев качества за последние годы являются впечатляющими и многообещающими [72].

Методы оценки качества могут быть использованы для мониторинга [73], а также для сравнения результатов обработки изображений и выбора оптимальных их параметров [68]. Методы оценки качества изображений показали свою пригодность в таких приложениях, как захват изображений [74], сегментации [75], в системах печати и представления изображений [76], синтеза [77], биометрии и многих других [70].

#### **1.3.1.1 Предсказание точности распознавания**

Центральным в рамках данной работы применением методов оценки качества изображений является предсказание точности распознавания образов на данном изображении в зависимости от его качества [72], что следует из проблемы неустойчивости алгоритмов распознавания к различным помехам, рассмотренной в разделе 1.2.

Приложения оценки качества в контексте распознавания включают в себя оценку в целях избежания передачи изображений низкого качества на удаленный сервер для распознавания, контроль действий и обратная связь с пользователем (низкая освещенность, объект слишком далеко и т.п.) [78], отказ в распознавании изображений заведомо низкого качества [79]. Предварительная оценка качества используется для выбора из нескольких алгоритмов распознавания отпечатков пальцев [80] и радужной оболочки [81], по-разному устойчивых к деградации качества. Подобные приложения реального времени предъявляют к методам оценки качества особенно строгие требования к производительности. Более



подробный обзор методов оценки качества изображений и их применений в контексте систем распознавания будет приведен в последующих разделах работы.

### 1.3.2 Субъективные методы оценки качества изображений

В средах обработки изображений, где человек является основным приемником, наилучшим способом оценить качество изображения является анализ человеческого оценочного мнения [68; 70; 71]. Несмотря на их сравнительно малую область применения в конечных приложениях, субъективные методы оценки качества играют важную роль в разработке объективных методов [82]. Субъективное тестирование предоставляет возможность анализа человеческого восприятия качества изображений, что обеспечивает полезной информацией разработчиков моделей восприятия и объективных метрик, на них основанных [83]. Результаты субъективного тестирования используются для оценки качества объективных методов, оценивая, насколько объективные методы приближаются к точному воспроизведению человеческого восприятия качества изображений и видео [84].

**Методологии субъективного тестирования** Для проведения процедур субъективного тестирования разработано множество методологий, в том числе стандартизованных [70; 82], наиболее простыми из которых методы одностимульной (средней экспертной оценки) и двустимульной категориальной оценки [85], получающие дальнейшее развитие в методах, использующих пост-обработку результатов по причине неравнозначности оценочной экспертной шкалы [86]. Методы попарного сравнения требуют  $C_n^2 = 0.5 \cdot n \cdot (n + 1)$  сравнений, что может быть неприемлемо при большом объеме тестовой выборки. В целях снижения количества требуемых сравнений используются методы, основанные, например, на  $O(n \log n)$  алгоритмах сортировки [87], в качестве компаратора запрашивающие экспертный выбор из двух изображений, а также методы кластеризации на основании предварительного эксперимента [88] и активного сэмпирования [89].

**Анализ и обработка субъективных оценок** По завершении процедуры субъективного тестирования требуется преобразование всего множества полученных оценок в интегральную оценочную характеристику для каждого изображения.

Пусть  $s_{i,j}$  – категориальная оценка на дискретной или непрерывной шкале, выставленная  $i$ -м экспертом  $j$ -му тестовому изображению, которое всего оценивало  $N_j$  экспертов. В качестве начального способа преобразования можно взять усреднение оценок в соответствии с методом средней экспертной оценки:  $\text{MOS}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} s_{i,j}$ .

Исследования [82; 90] показали, что непосредственное использование оценок качества изображений ненадежно: эксперты склоняются к назначению различных шкал оценок как среди отдельно взятых изображений, так и для различных типов искажений в рамках одного изображения. Первым шагом к корректированию такого поведения экспертов является анализ разницы оценок между исходным и искаженным изображением. Пусть  $s_{i,j}^*$  – оценка, аналогично выставленная  $i$ -м экспертом соответствующему  $j$ -му оригинальному (референтному) изображению. Тогда, показатель дифференциальной экспертной оценки для  $j$ -го изображения вычисляется, как  $d_{i,j} = s_{i,j}^* - s_{i,j}$ , а затем применяется  $z$ -нормализация:

$$z_{i,j} = \frac{d_{i,j} - \mu_i}{\sigma_i}, \quad \mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} d_{i,j}, \quad \sigma_i = \sqrt{\frac{1}{N_i - 1} \sum_{j=1}^{N_i} (d_{i,j} - \mu_i)^2}, \quad (1.18)$$

где  $N_i$  – количество изображений, оценивавшееся  $i$ -м экспертом,  $\mu_i$  – среднее значение его оценок и  $\sigma_i$  – их стандартное отклонение. Итоговый субъективный показатель качества изображения вычисляется, как среднее среди полученных  $z$ -показателей:  $\text{DMOS}_j = \frac{1}{N_j} \sum_{i=1}^{N_j} z'_{i,j}$ .

В [82; 84] приводятся более подробные методы проведения и анализа эксперимента, учитывающие различные варианты его условий, повторений и т.п., а также способы обработки результатов методов попарного сравнения.

**Меры соответствия объективных оценок субъективным** Степень соответствия разработанных объективных оценок субъективным на некоторой тестовой выборке в [91] рекомендуется вычислять по трем критериям: точность, монотонность, и согласованность [92]. Точность предсказания может быть вычислена двумя способами: как средняя ошибка (RMSE) между непосредственными предсказанными значениями  $x$  и субъективными MOS/DMOS  $y$ , или же измерением корреляции (линейной корреляции Пирсона [93]) между ними. Так как субъективные и объективные оценки качества изображений имеют заведомо разные шкалы, перед вычислением точности предсказания каким-либо из предложенных



способов следует выполнить монотонное нелинейное преобразование  $f(x)$  над предсказанными значениями объективного алгоритма, например, логистическую функцию [70] с четырьмя или пятью параметрами  $\bar{\beta} = \beta_1 \dots \beta_5$ , оптимизируемыми методом наименьших квадратов:

$$f(x) = \frac{\beta_1 - \beta_2}{1 + e^{\frac{x - \beta_3}{\beta_4}}} + \beta_5, \quad \bar{\beta} = \arg \min_{\bar{\beta}} \sum_{i=1}^N (f(x_i, \beta) - y_i)^2 \quad (1.19)$$

Монотонность предсказания определяет, насколько успешно объективный алгоритм сохраняет относительный порядок субъективных оценок. Примерами способов измерения монотонности, коэффициент ранговой корреляции Спирмена, является частным случаем коэффициента линейной корреляции Пирсона  $\rho$ , но для порядковых переменных, а также коэффициент ранговой корреляции Кендалла  $\tau$ . Согласованность предсказания показывает, насколько предсказания качества изображений объективным алгоритмом консистентны на данной тестовой выборке. Наиболее часто используемым показателем согласованности является процент выбросов, вычисляющийся, как:  $R_{\text{out}} = \frac{N_{\text{out}}}{N}$ , где  $N_{\text{out}}$  – количество объективных предсказаний, отличающихся от соответствующих субъективных оценок более, чем на два стандартных отклонения  $2\sigma_{y_i}$ . Диапазон в  $2\sigma_{y_i}$  был выбран потому, что он покрывает 95% субъективных оценок для данного изображения (при предположении нормального распределения) [92].

### 1.3.3 Объективные методы оценки качества изображений

Объективные методы оценки качества изображений можно классифицировать по доступности оригинального (без искажений) изображения, с которым сравнивается искаженное (например, сжатое) изображение. Большинство существующих подходов известны как full-reference, или эталонные, т.е. с полной информацией об исходном изображении. Однако, во многих практических приложениях, когда невозможно получить исходное изображение, приходится использовать no-reference («слепые», blind) подходы по оценке качества изображений. Кроме того, существуют комбинированные методы, т.н. методы с частичной информацией об исходном изображении, в которых известны некоторые свойства оригинального изображения [94].

В работе 1940 года [95] были описаны факторы, влияющие на качество изображения: разрешение, контраст, оттенок, яркость, мерцание, геометрическое искажение, размер, цвет и шум. Хотя в той работе и не приводилась явная формула оценки качества, многие сегодняшние алгоритмы оценки качества изображений в той или иной степени задействуют один или несколько из этих факторов.

### 1.3.3.1 Методы оценки качества с полной информацией

Покажем базовые статистические метрики оценки качества изображений с полной информацией об исходном изображении, используемые в [96]. Ими являются ( $x$ ,  $y$  – последовательности пикселей исходного и восстановленного изображения соответственно,  $N$  – количество пикселей) среднеквадратичное отклонение (СКО, Mean Squared Error, MSE), пиковое отношение сигнала к шуму (Peak Signal-to-Noise Ratio, PSNR):

$$MSE = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2, \quad PSNR = 20 \log_{10} \left( \frac{MAX_I}{\sqrt{MSE}} \right) \quad (1.20)$$

и другие, подробный обзор которых приведен в [70]:

$$NK = \frac{\sum_{i=1}^N x_i y_i}{\sum_{i=1}^N x_i^2}, \quad AD = \frac{1}{N} \sum_{i=1}^N (x_i - y_i), \quad SC = \frac{\sum_{i=1}^N x_i^2}{\sum_{i=1}^N y_i^2}, \quad MD = \max_{i=1 \dots N} (|x_i - y_i|). \quad (1.21)$$

Рассмотрим первые две метрики подробнее.

**Среднеквадратичное отклонение** Среднеквадратичное отклонение в теории вероятностей и статистике – наиболее распространенный показатель рассеивания значений случайной величины относительно ее математического ожидания. Среднеквадратичное отклонение является одной из самых популярных метрик в области обработки сигналов: до сих пор оно занимает основную позицию в оценке их качества и точности. СКО используют по многим причинам: метод очень прост в реализации, не требователен к памяти и имеет высокую производительность. СКО получило обширное применение в оптимизации, где часто

является объектом (функцией) минимизации, а также и в статистике, прогнозировании. Большое значение MSE подразумевает сильное различие в изображениях, и, наоборот, маленькое – схожесть. При равенстве изображений MSE равно нулю.

Несмотря на все преимущества СКО в обработке сигналов и статистике, в случае попытки предсказать ошибку или искажение изображения с позиции человеческого зрения СКО не так эффективно. На рисунке 1.6 изображено сравнение метрики СКО (MSE) с SSIM и CW-SSIM метриками, которые будут рассмотрены далее. Для сравнения с оригинальным изображением (а) были использованы изображения с различными искажениями: изменение контраста (b), яркости (c), гауссовый шум (d), импульсивный шум (e), компрессия JPEG (f), размытие (g), пространственное изменение масштаба (уменьшение) (h), пространственный сдвиг вправо (i), влево (j), поворот против часовой стрелки (k), по часовой стрелке (l). Можно увидеть, что значение СКО одинаково на изображениях как достойного качества, так и полностью испорченных, например, JPEG-компрессией с сильным коэффициентом сжатия (f). С другой стороны, MSE принимает большое значение там, где картинка визуальна почти не изменилась, а, например, сдвинулась на несколько пикселей. Для человеческого зрения подобные искажения практически незаметны, чего нельзя сказать про изменение в значении СКО [97].

**Пиковое отношение сигнала к шуму** Пиковое отношение сигнала к шуму является инженерным термином, означающим соотношение между максимумом возможного значения сигнала и мощностью шума, искажающего значения сигнала [98; 99] и по сути является корректировкой MSE для различной битности изображений. Поскольку многие сигналы имеют широкий динамический диапазон, PSNR обычно измеряется в логарифмической шкале в децибелах. Обычно, значения PSNR после сжатия изображений или видео с потерями находятся в пределах 30-50 дБ (больше – лучше) [100]. Для идентичных изображений MSE в знаменателе равно нулю, поэтому PSNR не определено.

MSE и PSNR метрики подвергаются критике за то, что они не всегда коррелируют с субъективными оценками людей, поэтому с трудом применимы в области оценке качества изображений, плохо применимы для сравнения изображений, сжатых различными кодеками и т.п.

**Индекс структурного сходства** Сигналы изображений имеют выраженную структуру: их пиксели находятся в сильной зависимости друг от друга, особенно

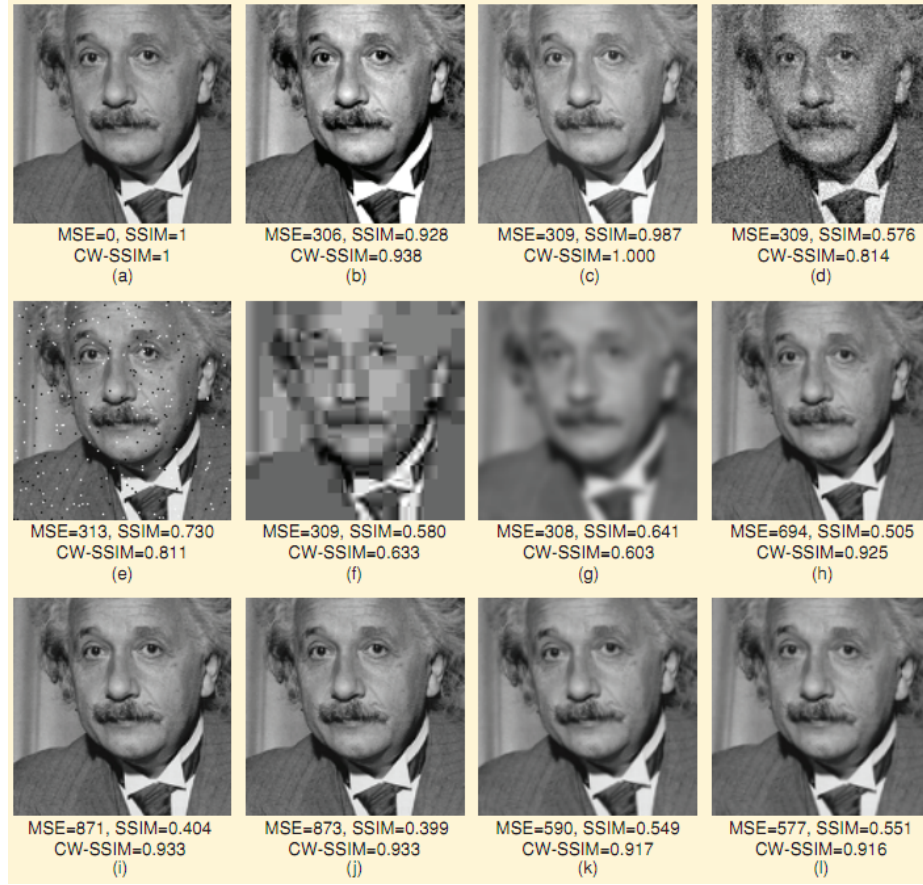


Рисунок 1.6 — Сравнение среднеквадратичного отклонения на изображениях с различными типами искажений

при их пространственной близости. Такие зависимости несут важную информацию о структуре объектов и о сцене в целом. Идея метрики SSIM, в отличие от MSE и PSNR, в том, что метод учитывает «восприятие ошибки», благодаря учету структурного изменения информации.

Индекс структурного сходства (Structural Similarity, SSIM) для двух изображений  $x$  и  $y$  вычисляется, как комбинация функций яркостного ( $l$ ), контрастного ( $c$ ) и структурного ( $s$ ) сравнения [94]:

$$SSIM(x, y) = l(x, y)^\alpha \cdot c(x, y)^\beta \cdot s(x, y)^\gamma, \quad (1.22)$$

где  $\alpha > 0$ ,  $\beta > 0$ ,  $\gamma > 0$  – параметры, требуемые для изменения относительной важности трех компонент, в свою очередь вычисляемых следующим образом:

$$l(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad c(x, y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad s(x, y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (1.23)$$

где  $\mu_x$  – среднее значение сигнала,  $\sigma_x$  – его дисперсия, а  $\sigma_{xy}$  – ковариация двух сигналов. Константы  $C_1$ ,  $C_2$  и  $C_3$  используются для избегания неустойчивости дробей при близких к нулю значениях знаменателей. Если положить  $C_3 = C_2/2$ ,

то получим итоговую формулу структурного сходства:

$$SSIM = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy})}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}. \quad (1.24)$$

На практике часто используют усредненную-SSIM метрику, усредняя полученные значения SSIM в окнах, размера, например, 8x8.

### 1.3.3.2 Методы оценки качества без информации об исходном изображении

Методы с полной информацией об исходном изображении, описанные ранее, являются простыми в реализации, дают достаточно хорошую универсальную оценку степени деградации качества изображения после его обработки и могут быть использованы в случаях, когда исходное изображение доступно. Однако, в системах распознавания актуальны методы, оценивающие степень присутствия помех на изображении без какой-либо дополнительной информации, поскольку исходное изображение во многих случаях недоступно или не существует.

Согласно обзору в работе [101], данный класс методов оценки качества можно разделить на два типа. Методы первого типа являются универсальными и не анализируют конкретные типы искажений, а второго типа, наоборот, – подразумевают наличие специфичной задаче модели искажений: размытие, блики и т.д. Рассмотрим подходы к универсальной оценке качества изображений.

Большинство универсальных методов оценки аналогичны методам машинного обучения в том смысле, что обычно имеют две стадии: извлечение признаков и предсказание качества. Их цель – научиться предсказывать человеческие субъективные оценки на имеющихся наборах данных [102].

Один из популярных методов такого типа – BRISQUE [103], использует в вычислениях нормализованные яркостные значения в окне:

$$\hat{I}(i, j) = \frac{I(i, j) - \mu(i, j)}{\sigma(i, j) + C}, \quad (1.25)$$

$$\mu(i, j) = \sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} I_{k,l}(i, j), \quad (1.26)$$

$$\sigma(i, j) = \sqrt{\sum_{k=-K}^K \sum_{l=-L}^L w_{k,l} (I_{k,l}(i, j) - \mu(i, j))^2}, \quad (1.27)$$



где  $w$  представляет собой двумерную функцию весов в соответствии с Гауссовым распределением. Идея метода заключается в том, что на основании данной статистики можно построить модель (а именно – обобщенное Гауссово распределение), сохраняющие свои свойства для естественных изображений и значительно их меняющие при различных искажениях. Авторы метода показывают, что анализ распределения позволяет в некоторой степени определить тип встречающейся помехи.

В методе GMLOGQA [104] используется подход с вычислением карт модулей градиентов и Лапласиана от Гауссиана и их последующей адаптивной нормализацией. В работе [105] предложен метод ILNIQE, моделирующий модули градиентов через распределение Вейбулла.

После извлечения признаков каким-либо способом [101], необходимо произвести подбор параметров модели для финального предсказания качества. Для этого могут использоваться стандартные подходы машинного обучения: регрессия, метод опорных векторов, и прочие. В последние годы популярность набирают методы, основанные на глубоких нейронных сетях, что позволяет использовать предварительно обученные модели и обучать только последний слой сети на тренировочных наборах данных с субъективными оценками качества [106].

#### **1.3.4 Специализированная оценка качества изображений**

Введенные ранее методы оценки качества изображений благодаря своей универсальности могут использоваться во многих сферах обработки изображений. Однако их универсальность приводит к тому, что диагностика помех становится невозможной, из-за чего пропадает возможность информативной обратной связи с пользователем, применения методов корректировки помех и т.д. По этой причине для внедрения методов оценки качества изображений в системы распознавания наиболее интересны специальные методы поиска искажений заданного вида.

В разделе 1.2 были описаны основные типы искажений, возникающие при распознавании в неконтролируемых условиях съемки. Рассмотрим менее универсальные объективные методы оценки качества, разработанные специально для поиска или оценки степени проявления данных типов искажений, а также приложения этих методов.

### 1.3.4.1 Методы и приложения детектирования и фильтрации бликов

Задачи детектирования и фильтрации бликов широко распространены при разработке различных систем захвата, обработки и анализа изображений и видео. Существует множество методов удаления с изображения зеркальных отражений источника света без привязки к конкретной предметной области [107–110], подробный обзор которых приведен в [111].

Детекция бликов и других шумовых областей на изображениях актуальна при реконструкции архивов фото и видеоданных [111; 112]. В [113] для этого используется специальная модификация физического устройства камеры. Системы внешнего ночного наблюдения [114] и автоматического детектирования дорожных происшествий [115] задействуют детекцию и удаление бликов по причине того, что блики ведут за собой большое количество отказов и ложных срабатываний в этих системах.

Детектирование и закрашивание световых отражений широко применяются в медицинских и биометрических сферах. Блики являются серьезной проблемой при сегментировании радужной оболочки глаза, что требует их детекции и закрашивания перед работой остальных подсистем – в частности, при использовании AdaBoost-каскада Виолы-Джонса [116]. В дерматологии детекция бликов задействуется при анализе изображений кожи пациентов на предмет наличия меланомы [117], в том числе, при съемке с мобильных устройств [118].

Множество работ опубликовано о детекции и удалении бликов на изображениях, полученных в процессе различных видов эндоскопии. Основными направлениями детекции и удаления бликов в видеопотоке и на отдельных изображениях являются коррекция видео на мониторе в реальном времени [119–124], увеличение точности восстановления структуры из видеопотока [119], коррекция кадров для последующего автоматического или ручного анализа на предмет рака и других заболеваний [123; 125; 126]. В [127] детектирование бликов является важной частью подсистемы выделения информативных кадров среди результатов эндоскопии для оптимизации работы как системы автоматического анализа этих кадров, так и ручной проверки врачом.

Влияние бликов при сканировании документов также исследуется в контексте систем общего контроля качества изображений [128], модификации процесса сканирования [129] и непосредственного детектирования бликов [130].



### 1.3.4.2 Периодические помехи

Рассмотренные в разделе 1.2 голографические и прочие фоновые элементы во многих случаях имеют периодическую структуру, что позволяет производить их поиск с высокой точностью по сравнению с одиночными фоновыми элементами.

Большое число работ посвящено проблематике детектирования и удаления шумов [131], но большинство из них имеют дело с неструктурированным шумом [132] или помехами на документах с простым фоном [133]. Для достаточно малых (по сравнению с размером текста) периодических шумовых шаблонов возможно применение морфологических операций [134] для их удаления. Однако, некоторые документы имеют достаточно сложный фоновый рисунок с большими и высококонтрастными шумовыми шаблонами. Также, периодический шум может быть детектирован с помощью инженерного подхода, основанном на сглаживании сигнала и накоплении ошибки [135], или же непосредственно анализируя заранее известное изображение периодической структуры методами динамического программирования [136]. Для периодических шумов достаточно маленького размера (по сравнению с размером текста) возможно применить морфологические операции для их удаления [134].

При взаимодействии с различными периодическими структурами широко распространены подходы с использованием дискретного преобразования Фурье (ДПФ, [137]), где для детектирования и локализации происходит анализ пиковых структур на ДПФ спектре. Данный класс подходов исследуется уже продолжительное время [138].

В [139] приводится алгоритм обнаружения квазипериодических текстур с целью дальнейшего сжатия или уточнения оценки движения при обработке видеопоследовательностей, а в [140] – его продолжение с оценкой координат узлов решетки квазипериодических структур, позволяющее восстанавливать форму объектов по текстуре и автоматически устранять решетчатую преграду на фотографиях. Поиск повторяющихся текстур с использованием ДПФ для увеличения устойчивости стабилизации видео также исследуется в [141], а в [142] предлагается использование ДПФ для получения изображения отпечатка подошвы обуви в крайне зашумленных условиях на местах преступлений. В [143] анализ

пиковой структуры ДПФ проводится с целью контроля качества, заключающегося в детектировании, наоборот, непериодических структур. Анализ ДПФ может применяться при классификации образов, пример чего описан в работе [144], где в качестве вектора признаков используются амплитудные коэффициенты спектра. Наконец, в [145] исследуется влияние различных шумов, возникающих при съемке периодических текстур с камер мобильных устройств, на пиковую структуру ДПФ, что затрагивает широкий класс описанных выше алгоритмов.

#### **1.3.4.3 Дефокусировка и размытость**

Основными подходами при оценке степени размытия изображения являются анализ градиента изображения [146; 147], а также преобразований Фурье [148], Радона [149] и вейвлет-преобразования [150], более устойчивых к шумам на изображении. Простые, но менее точные подходы также актуальны в приложениях с критичными запросами к производительности [44].

С целью восстановления изображения после дефокусировки и размытия в результате движения общим методом является аппроксимация ядра свертки, с которым, согласно модели, было свернуто исходное (резкое) изображение, для последующей обратной фильтрации. В случае дефокусировки ядро свертки представляет функцию точечного распределения из формулы 1.17.

Скалярная оценка степени размытия является важной оценкой качества изображений, для которой существуют различные тестовые выборки, предназначенные для анализа работоспособности алгоритмов стандартными процедурами [151]. Большое внимание уделяется предсказанию точности распознавания на основании степени размытия текста [78].

Методы оценки степени размытия также используются при автофокусировке камер, построении карт глубины изображения [54] и предсказании движения на изображении аналогично оптическому потоку [152].

#### 1.3.4.4 Оценка качества изображений печатных документов

В разделе 1.1.1 были рассмотрены изображения печатных документов и их области применения. Изображения данного класса имеют особенности их использования, создания, возможных искажений, оцифровки и обработки, позволяющие производить более специфичную и точную оценку качества.

В работах [153] и [154] приведены обзоры основных методов оценки качества изображений документов, их коррекции, а также моделирования различных типов искажений. Можно выделить два общих подхода к определению качества изображения документа: точность распознавания присутствующего на нем текста с помощью OCR-систем и качество в смысле визуального человеческого восприятия.

Помимо ранее рассмотренных в данной работе искажений (размытие, неравномерность освещения, шум и др.), большинство методов оценки качества изображений документов рассматривают такие признаки, как контрастность, параметры шрифта, штрихов, касание, пересечение и повреждение символов, выраженность границ, равномерность свойств компонент связности и прочие. Минусом данных методов является необходимость в ручной разработке метрик оценки для каждого из типов искажений. По этой причине наблюдается рост популярности методов оценки качества, основанных на машинном обучении: как для комбинирования созданных вручную признаков изображения, так и для составления словаря признаков [155]. Обзор данного класса методов представлен в [156].

## 1.4 Выводы по главе 1. Основные задачи диссертации

В главе 1 были рассмотрены основные типы искажений, которым подвергаются входные данные систем обработки и распознавания изображений, а также соответствующие методы оценки качества изображений. Было показано, что этим вопросам уделяется достаточно большое внимание вследствие их высокой актуальности. На текущий момент разработано множество объективных (автоматических) методов оценки и коррекции качества изображений.

Несмотря на это, практическое применение имеющихся методов в промышленных системах обработки и распознавания изображений имеет существенные ограничения.

Во-первых, не существует универсального метода оценки качества изображений, применимого в любых приложениях, поскольку требования к качеству различаются как среди классов входных изображений, так и среди классов конкретных систем.

Во-вторых, отсутствуют модели, методологии или рекомендации к созданию систем распознавания, учитывающих качество входных изображений с целью повышения точности, надежности, производительности и других характеристик систем.

В-третьих, для многих прикладных систем распознавания изображений и видеопотока актуальна не только традиционная постановка задачи оценки и коррекции качества входных изображений, но и ее вариации, актуальные при выборе лучшего кадра в видеопотоке, отказе в обработки поврежденных областей, интеграции результатов распознавания на разных кадрах и т.д.

В диссертации исследуются три важных промышленных приложения систем распознавания: в системах документооборота и долгосрочного хранения документов, обеспечения контроля доступа и безопасности, а также распознавания изображений в непосредственном технологическом процессе. Для этих приложений рассматриваются следующие задачи:

1. Исследовать влияние искажений входных изображений на промышленные системы распознавания и провести обзор современных методов оценки качества изображений.

2. Построить математическую модель системы распознавания изображений и видеопотока с модулями оценки качества и обратной связью на каждом этапе обработки.

3. Разработать алгоритм автоматической оценки качества цветовой сегментации изображений печатных документов.

4. Исследовать влияние детектирования и фильтрации бликов на точность поиска и распознавания документа на изображении, а также влияние бликов на устойчивость алгоритма вычисления взвешенного проективного оптического потока.

5. Разработать алгоритм детектирования и локализации периодических фоновых элементов известной структуры на изображении документа.

6. Разработать алгоритм контроля качества входного сигнала двухканального рентгенографического сепаратора.

7. Реализовать разработанные в диссертации алгоритмы для обеспечения их внедрения в промышленные системы распознавания и провести их экспериментальный анализ.

Экспериментальный анализ систем распознавания документов в главах 2 и 4 будет вестись на примере системы, описанной в [21], во-первых, реально применяющейся на большом количестве предприятий, а во-вторых – достаточно разнообразной, чтобы отразить основные особенности систем распознавания и проблемы контроля качества входных изображений. В главе 3 в экспериментах будет рассмотрена реализация аналогичной [29] системы сжатия документов для их долговременного хранения. В качестве системы распознавания, применяющейся в промышленном технологическом процессе, в главе 5 рассматривается система рентгенографической сепарации алмазов.

## Глава 2. Оценка качества изображений в системах распознавания

В данном разделе исследуется оригинальный подход к построению систем распознавания образов с встроенными модулями оценки качества входных изображений и обратной связью на каждом этапе обработки. Данный подход обеспечивает возможность контроля уверенности результата и отказа в дальнейшей обработке в случае низкого качества входного изображения на рассматриваемом этапе, что позволяет увеличить точность, устойчивость и скорость работы систем распознавания в неконтролируемых условиях съемки. Введены необходимые определения и построена модель архитектуры систем распознавания, применимые для программной реализации подхода в проектируемых и существующих системах.

Значимость подхода иллюстрируется на примере решения задачи выбора наилучших кадров для распознавания в видеопотоке с помощью функции приоритета кадра, зависящей от его качества и времени съемки. Приводятся экспериментальные результаты решения данной задачи для реализации системы распознавания удостоверяющих личность документов в условиях автоматической фокусировки камеры, приводящих к размытию кадра. Результаты демонстрируют значительное увеличение точности распознавания с различными ограничениями на продолжительность видеопотока при выборе кадров для распознавания в соответствии с их приоритетом.

### 2.1 Постановка задачи оценки качества изображений

Введем определения оценки качества изображений, необходимые для построения модели их внедрения в системы распознавания. Как упоминалось ранее в разделе 1.3.3, 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  $\mathcal{Q}$ , а за  $\mathcal{D}$  – множество выходных оценок качества. Пусть имеется

изображение  $I \in \mathcal{I}$ . Объективной функцией оценки качества в случае отсутствия информации об исходном изображении назовем функцию  $Q_{NR}$ , принимающую одиночное входное изображение:

$$Q_{NR} : \mathcal{I} \rightarrow \mathcal{D} \quad (2.1)$$

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

$$Q_{FR} : \mathcal{I} \times \mathcal{I} \rightarrow \mathcal{D} \quad (2.2)$$

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

$$Q_{PR} : \mathcal{I} \times \mathcal{P} \rightarrow \mathcal{D} \quad (2.3)$$

Рассмотрим подробнее множество  $\mathcal{D}$  выходов методов. В самом распространенном для общей оценки качества случае  $\mathcal{D} = \mathcal{D}_R$  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:

$$Q(I) = D \in \mathcal{D}_R \quad (2.4)$$

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

$$Q(I) = D \in \mathcal{D}_B, D = \{d_{x,y} \in \{0,1\} \mid x, y : I_{x,y} \in I\}, \quad (2.5)$$

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

$$Q(I) = D \in \mathcal{D}_I, D = \{d_{x,y} \in \mathbb{R}_{[0,1]} \mid x, y : I_{x,y} \in I\} \quad (2.6)$$

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:

$$Q(I) = D \in \mathcal{D}_C, D = \{(q_i, C_i)\}, C_i = \{(x, y) \mid x, y : I_{x,y} \in I\}, C_i \cap C_j = \emptyset, \quad (2.7)$$

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

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

В дальнейшем под множеством  $\mathcal{D}$  будет подразумеваться одно из вышеперечисленных множеств:

$$\mathcal{D} \in \{\mathcal{D}_R, \mathcal{D}_B, \mathcal{D}_I, \mathcal{D}_C, \mathcal{D}_O\}, \quad (2.9)$$

в зависимости от конкретного приложения задающих информацию об искажениях.

## 2.2 Модель системы распознавания с модулями оценки качества

На основании обзора в разделе 1.1.4 построим общую модель системы распознавания, содержащей подсистему оценки качества, и опишем их взаимодействие между собой.

Рассмотрим систему  $G = \langle S, E \rangle$ , представленную в виде ориентированного графа подсистем обработки  $S = \{S_0, \dots, S_N\}$  и связей между ними  $E = \{(S_i, S_j) \mid S_i, S_j \in S\}$ . Каждая подсистема (этап) обработки  $S_i$  может быть произвольной процедурой обработки, анализа (в том числе – распознавания), синтеза, компрессии, декомпрессии изображений и иметь своими входами и выходами как изображения, так и другие типы данных: вектора признаков, строки результатов распознавания, найденные контуры объектов и другие. Пусть  $S_0$  – особый этап, имеющий пустой вход и заключающийся в регистрации изображения сцены, которое и возвращается на выходе этого этапа.

Выходом этапа  $S_i$  является упорядоченный набор объектов  $r_i = (r_{i,0}, \dots, r_{i,N_i})$ . Каждый отдельный объект  $r_{i,j}$  содержит непосредственный набор выходных данных  $v_{i,j}$ , тип которых зависит от устройства  $S_i$  и может быть,

например, изображением или вектором альтернатив распознавания. Кроме того, он содержит скалярную оценку уверенности подсистемы в корректности своего результата  $c_{i,j} \in [0, 1]$  и опциональную информацию  $l_{i,j}$  о местонахождении результата в глобальной системе координат – например, аффинное или проективное преобразование для получения позиции символа после модуля сегментации строки в координатах текущего кадра. Итого,  $r_{i,j} = \langle v_{i,j}, c_{i,j}, l_{i,j} \rangle$ . Данные выходы поступают на вход соединенным дугами из  $E$  подсистемам обработки, которые, в том числе, принимают решение о целесообразности их обработки с учетом оценок уверенности  $c_{i,j}$  и учитывают их в своих последующих оценках.

Опишем процесс встраивания модулей оценки качества изображений в систему. Ограничимся модулями исходной системы  $S_i$ , и теми их 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  $j$ -го выхода подсистемы  $S_i$ , принимающий на вход  $r_{i,j}$ . Данный модуль реализует одну из описанных в предыдущем подразделе функций  $Q(I)$  и возвращает оценки качества  $q_{i,j} \in \mathcal{D}_{i,j}$  ( $\mathcal{D}_{i,j} \in \mathcal{D}$  по формуле 2.9), которые, как было обозначено ранее, являются пространственным распределением оценок качества для поступившего изображения  $v_{i,j}$  в системе координат, заданной  $l_{i,j}$ .

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  $r_{i,j}^*$ , тип которых совпадает с типом  $r_{i,j}$ , но компонента уверенности в корректности результата  $c_{i,j}^*$  может отличаться в соответствии с логикой учета пространственного распределения оценок качества  $q_{i,j}$ . Важной особенностью  $\Psi_{i,j}$  является возможность выдачи отказа в дальнейшей обработке за счет присвоения  $c_{i,j} = 0$ , когда повреждение покрывает большую часть изображения  $v_{i,j}$ , и передачи сообщения об этом родительской подсистеме  $S_i$  с помощью обратной связи. Другим способом модификации уверенности при повреждениях может быть снижение значения уверенности  $c_{i,j}$  путем умножения на некоторое скалярное значение  $\theta(q_{i,j})$ , вычисляемое как средний уровень деградации качества изображения от эталонного. Кроме того, модуль возвращает восстановленное от дефектов изображение  $v_{i,j}^*$  при таковой возможности.

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, что будет описано подробнее в следующем подразделе. На вход данному модулю подаются оценки качества  $q_{i,j}$  и модифицированные результаты  $r_{i,j}^*$  работы подсистемы, причем в модуле хранится история о результатах на предыдущих кадрах, т.е. выходная карта является интегральной по видеопотоку на текущий момент. Обновление карты интереса может происходить путем усреднения обратных показателей качества  $(1 - q_{i,j})$  или уверенности в корректности работы  $(1 - c_{i,j}^*)$  в областях, пересчитанных обратно в координатах исходного объекта с помощью  $l_{i,j}$ . Если  $\Psi_{i,j}$  принимает решение об отказе дальнейшей обработки, модуль  $A_i$  все равно обновляет свою карту интереса в сторону его увеличения в поврежденных областях. Применение карты интереса особенно актуально при распознавании объектов с фиксированными формами, например, документов – низкий интерес в регионе означает, что система уже имеет высокий уровень уверенности в результатах распознавания в этом регионе, или же что в нем отсутствуют какие-либо информативные участки.

Наконец, по завершении процесса оценки качества и коррекции, модифицированные результаты  $r_{i,j}^*$  передаются на следующий этап обработки в соответствии с существующими связями между подсистемами.

На рисунке 2.1 проиллюстрирована модель графа подсистем и их связей после добавления модулей оценки качества изображений в процесс обработки.

При реализации данной модели в программных системах добавляемые модули оценки качества и принятия решений могут быть многократно использованы при однотипных выходах соответствующих подсистем.

### 2.3 Выбор лучших кадров для распознавания в видеопотоке

При распознавании в видеопотоке становится актуальной задача выбора наилучших по какому-либо критерию качества кадров для распознавания из целого потока кадров. Это обусловлено несколькими причинами.

Во-первых, большинство современных камер (как мобильных устройств, так и других) имеет возможность захватывать кадры с частотой 30-60 кадров в

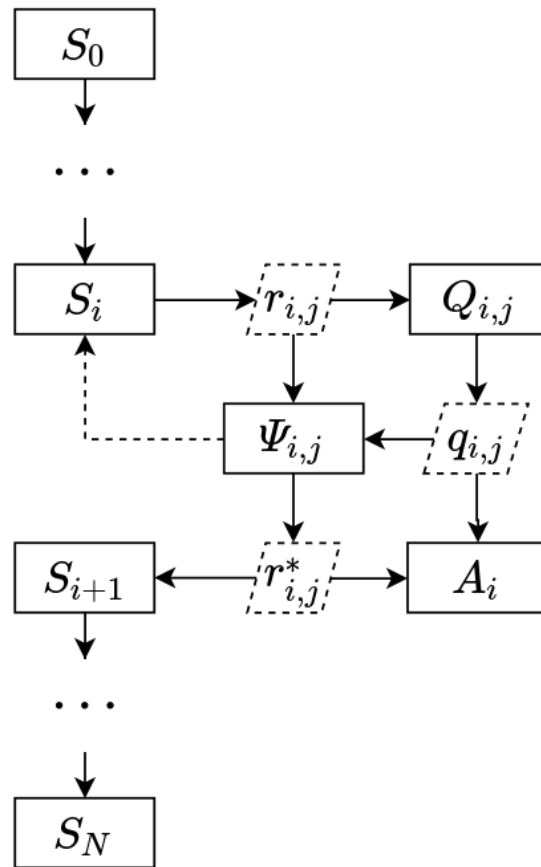


Рисунок 2.1 — Граф системы обработки с модулями оценки качества изображений

секунду и выше, что может в разы превышать производительность обработки одного кадра системой распознавания [21].

Во-вторых, помимо распознавания на устройстве, также популярным является сценарий только съемки фотографии или видеопотока с интересующим объектом на устройстве, а затем посылка этих данных на удаленный сервер для последующего распознавания [157]. Отсылать весь захваченный видеопоток целиком неблагоприятно как при слабом канале связи, так и при необходимости обеспечить постоянную обратную связь с пользователем о процессе распознавания.

В-третьих, остаются верными предыдущие аргументы о целесообразности оценки качества в системах распознавания: передавая кадр с заведомо низким качеством на распознавание существует риск снижения общей точности распознавания.

Модуль выбора кадров из избыточного видеопотока может быть встроен в соответствии с ранее приведенной схемой на рисунке 2.1 – на основании результатов подсистемы захвата кадров  $S_0$  и модуля предварительной оценки качества  $Q_0$ .

Пусть  $\Psi_0$  – модуль выбора кадров, имеющий устройство очереди с приоритетом, накапливающий входные кадры и при запросе выдающий кадр с наибольшим приоритетом. Если после добавления очередного кадра размер очереди превышает  $L$ , из нее удаляется кадр с наименьшим приоритетом. Введем дополнительный модуль оценки приоритета кадра  $P$ , зависящий от результатов оценки качества  $q(x, y)$ , времени  $t$  нахождения кадра в очереди, а также от карты интереса  $a(x, y)$  системы распознавания в соответствующих регионах входного изображения ( $0 \leq q(x, y), a(x, y) \leq 1$ ). На основании пространственных распределений  $q(x, y)$  и  $a(x, y)$  вычислим коэффициент  $w$ , показывающий пригодность данного кадра с точки зрения качества интересующих систему регионов:

$$w = \sum_{x,y} q(x, y)a(x, y). \quad (2.10)$$

В качестве способа вычисления итогового приоритета кадра  $p$  с учетом времени нахождения в очереди  $t$  и пригодности кадра  $w$  возьмем способ, использующий гауссово ядро [158]:

$$p(w, t) = w^\alpha e^{\frac{-\beta t^2}{\sigma^2}}, \quad (2.11)$$

где коэффициент  $\alpha$  регулирует важность оценки качества, а  $\beta$  и  $\sigma$  контролирует период релевантности кадра относительно времени захвата. На рисунке 2.2 показаны графики функции  $p(w, t)$  для различных значений  $w$  и  $t$ .

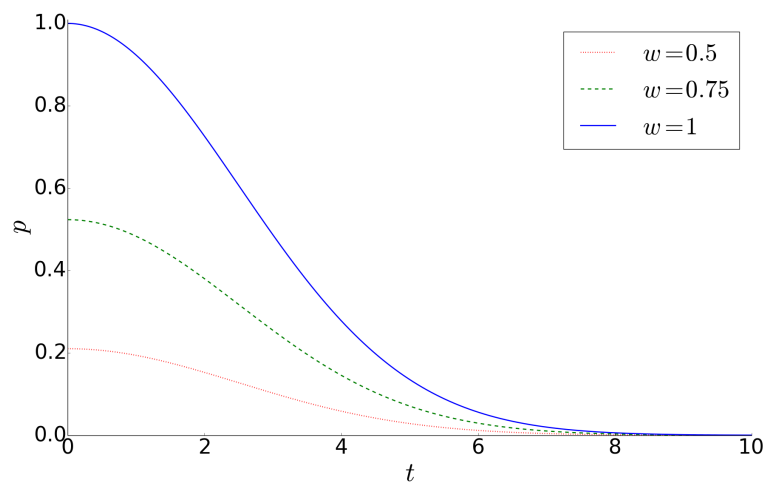


Рисунок 2.2 — Графики функции  $p(w, t)$  при  $\alpha = 2.25$ ,  $\beta = 0.5$ ,  $\sigma = 2.5$

Таким образом, с помощью вычисления приоритета  $p(w, t)$  становится возможным выбор кадров видеопотока наилучшего качества для распознавания.

## 2.4 Эксперименты

Цель экспериментов – показать, что с помощью выбора кадров видеопотока в соответствии с введенной функцией приоритета можно добиться большей точности и скорости распознавания, чем при произвольном выборе кадров. В экспериментах использовалась реализация системы распознавания документов, удостоверяющих личность, в видеопотоке, описанной в [21]. Используемый в экспериментах модуль оценки качества оценивает степень размытия (при движении) и расфокусировки кадра. При этом, также проводится моделирование размытия во время процесса фокусировки камеры.

### 2.4.1 Алгоритм оценки степени размытости кадра

Для оценки размытости и расфокусировки входного кадра  $I$  в экспериментах использовался алгоритм, предложенный в работе [44].

Построим карты  $G_1, \dots, G_4$  модулей производных изображения  $I$  в вертикальном, горизонтальном и двух диагональных направлениях, приведя изображение  $I$  к полутоновому со значениями пикселей от 0 до 1. Для текущего изображения модуля градиента  $G_k$  вычислим его гистограмму  $H_k$  и возьмем верхний  $g$ -й квантиль  $r_k$ . Финальной оценкой  $w$  сфокусированности кадра будет квантиль для самого выраженного смаза, т.е.  $\min(r_1, \dots, r_4)$ , экспоненциально нормированная на  $(0, 1]$ .

Для проверки соответствия вычисленных алгоритмом значений настоящему размытию был проведен эксперимент: часть выборки (100 изображений) была подвергнута Гауссовому размытию с радиусом  $\sigma$ , изменяющимся с шагом в заданном интервале, в гистограмме выбирался  $g = 0.05$ -й квантиль. На рисунке 2.3 показана зависимость усредненных значений  $w$  по выборке для фиксированных  $\sigma$ .

Начиная с  $\sigma_0$  размытие становится заметным, а с  $\sigma_1$  – текст на документе становится нечитаемым. Видно, что значения вычисленной оценки монотонно убывают при возрастании степени размытия, что доказывает пригодность использования данного алгоритма.

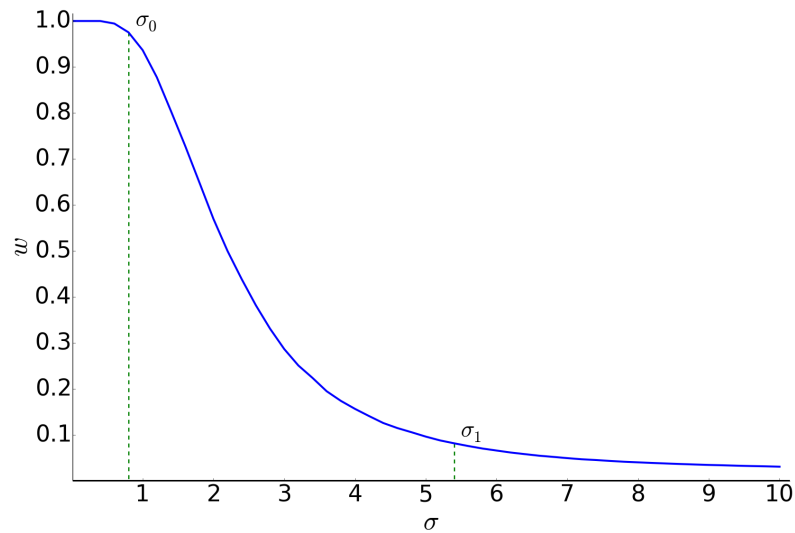


Рисунок 2.3 — График зависимости вычисленной оценки фокусировки  $w$  от радиуса  $\sigma$  размытия фильтром Гаусса

#### 2.4.2 Моделирование распознавания в процессе автоматической фокусировки камеры

Рассмотрим приложение систем распознавания, при котором расстояние до объекта не известно заранее, что актуально при распознавании с мобильных устройств, веб-камер и т.д. В таких случаях используются алгоритмы автоматического фокусирования камеры на объекте, сканирующие диапазон фокусного расстояния и измеряя некоторый критерий четкости кадра, в конце возвращаясь к его оптимуму [159]. Это осложняет процесс распознавания и вместе с тем является идеальным примером применения выбора наилучших кадров для распознавания, поскольку при автоматической фокусировке с камеры поступают как сильно, так и слабо размытые кадры и важно распознавать именно последние.

Для моделирования размытия в процессе автоматического фокусирования камеры возьмем функцию

$$\sigma(t) = a|\text{sinc}(b(t - t_0))| = \frac{a|\sin(b(t - t_0))|}{b|t - t_0|} \quad (2.12)$$

зависимости радиуса фильтра Гаусса от времени захвата кадра  $t > 0$ , график которой для различных коэффициентов  $a$ ,  $b$  и  $t_0$  показан на рисунке 2.4.

На рисунке 2.5 показан пример части видеопотока с искусственно подготовленным шаблоном ID-карты Германии, подвергаемый размытию в соответствии



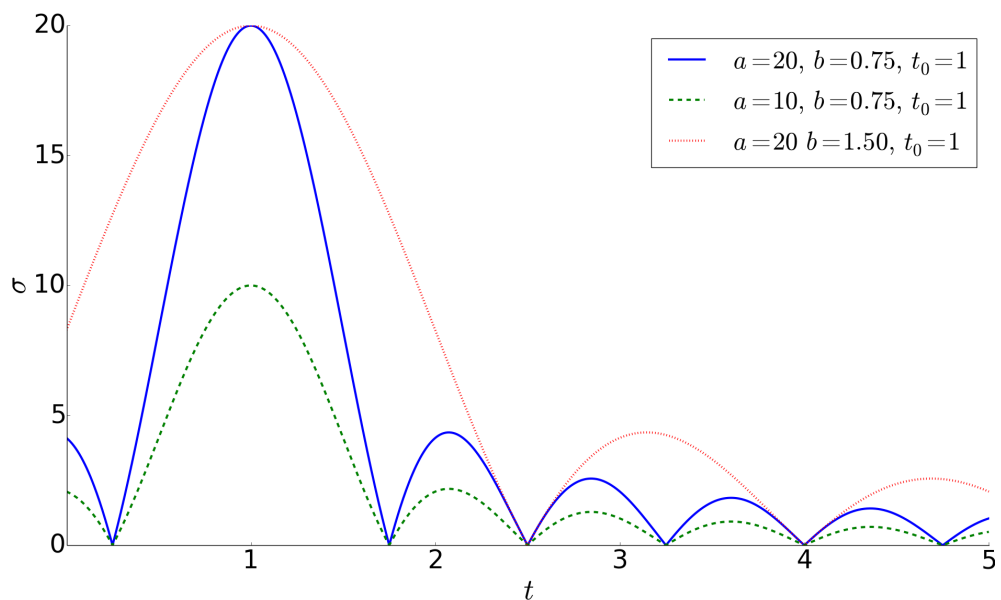


Рисунок 2.4 — График функции  $\sigma(t)$  моделирования радиуса размытия при автоматической фокусировке

с формулой 2.12. Искаженные таким образом видеоролики подаются в систему распознавания.

### 2.4.3 Экспериментальные результаты

Для экспериментов с введенной ранее системой [21] использовались тестовые выборки из 217 и 39 видеороликов удостоверяющих личность документов (ID-карты Германии, паспорт гражданина РФ) длительностью 5 секунд каждый, снятых с использованием различных мобильных устройств. Для каждого видеоролика была подготовлена идеальная разметка полей документа: номера, имени, фамилии, национальности, даты и места рождения, даты окончания действия документа, а также полной машиночитаемой зоны для ID-карт Германии и серии, номера и кода подразделения для паспортов РФ, поскольку они присутствуют на обеих страницах разворота. Поле считалось правильно распознанным в случае полного совпадения возвращенного системой значения с идеальным значением на момент обработки всего видеоролика с документом. Итоговой точностью работы системы на выборке положим среднюю точность распознавания вышеперечисленных полей.

а)  $\sigma = 0$ б)  $\sigma = 2$ в)  $\sigma = 4$ г)  $\sigma = 8$ 

Рисунок 2.5 — Пример размытия при моделировании процесса автоматической фокусировки для видеопотока

Каждый видеоролик разбивался на кадры с фиксированной частотой  $F = 5$  кадров в секунду. Кадры передавались на распознавание с периодом  $T = (1, \dots, 8)$  кадров (т.е. с итоговой частотой  $\frac{F}{T}$  кадров в секунду), тем самым имитируя продолжительность распознавания одного кадра. Помимо этого, были введены искусственные ограничения на максимальную продолжительность видео  $M = (5, 3, 1)$  секунд для проверки потенциала скорости распознавания, т.е. точности при фиксированном максимальном времени обработки видеопотока. Эксперимент проводился в двух режимах: с простым выбором каждого  $T$ -го кадра («Fixed») и выбором наилучшего кадра в каждый  $T$ -й момент времени в соответствии с предложенным способом вычисления приоритета («Priority»). При моделировании автоматической фокусировки по формуле 2.12 использовались значения  $a = 20$ ,  $b = 0.75$ , а время  $t_0$  пика функции размытия в секундах был равномерно распределено на  $[0, 3]$ .

На рисунках 2.6 (ID-карты Германии) и 2.7 (паспорта гражданина РФ) показаны полученные результаты точности распознавания видеопотока для исходной (вверху) и размытой (снизу) выборки изображений с различными ограничениями  $M$  на максимальную продолжительность видео. По горизонтальной оси отложен

период между кадрами  $T$ , а по вертикальной – средняя точность  $ACC$  распознавания всех полей. Каждый график содержит сравнение методов с фиксированным и приоритетным выбором кадра.

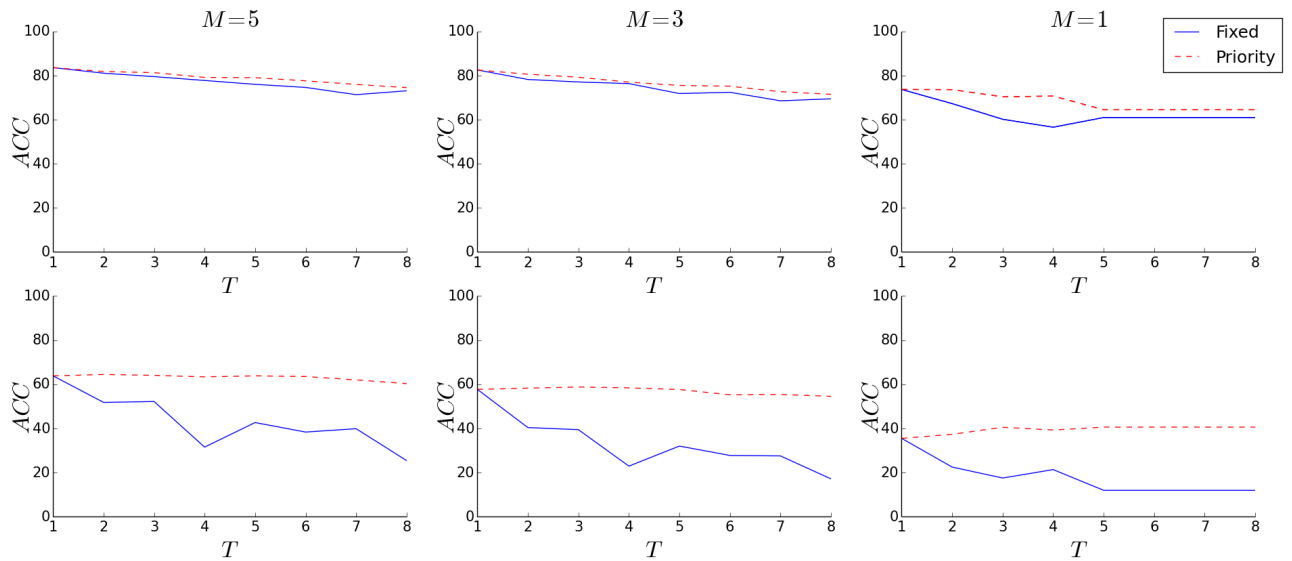


Рисунок 2.6 — Графики точности распознавания видеопотока для исходной (сверху) и размытой (снизу) выборки изображений

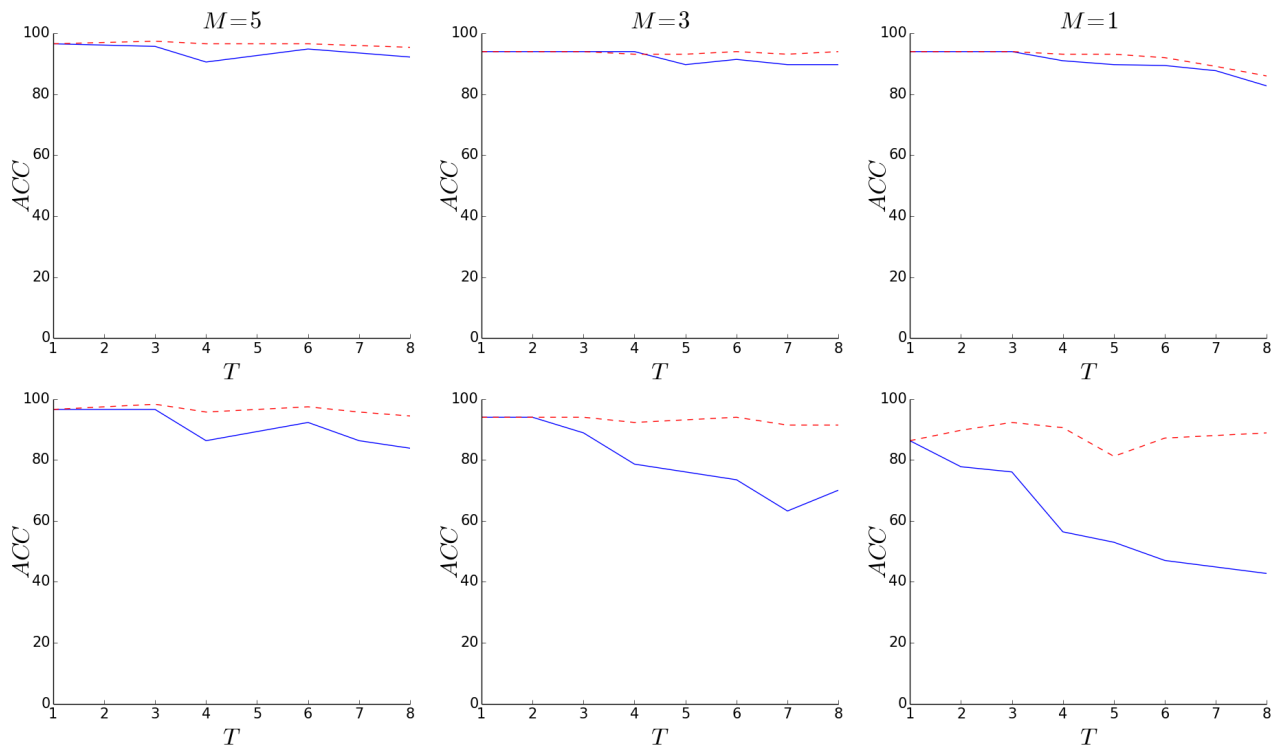


Рисунок 2.7 — Графики точности распознавания видеопотока для исходной (сверху) и размытой (снизу) выборки изображений паспортов гражданина РФ

Графики демонстрируют, что точность распознавания видеопотока при использовании предложенного алгоритма с вычислением приоритета значительно выше при моделировании процесса автоматической фокусировки, чем в случае выбора фиксированных кадров. При этом, оно выше и на исходной выборке изображений, не подвергнутой искусственному размытию. Данный эффект наблюдается при всех выбранных ограничениях на максимальную длину видеопотока.

Таким образом, можно сделать вывод о целесообразности применения предложенного алгоритма выбора наилучших кадров для распознавания в видеопотоке: как в процессе автоматической фокусировки камеры (или в присутствии других искажений), где он дает наибольший прирост точности и скорости работы, так и в обычных условиях.

## **2.5 Выводы по главе 2**

В данной главе представлена оригинальная модель архитектуры систем оптического распознавания с встроенными модулями оценки качества входных данных, введены необходимые определения, сущности и связи между ними. В рамках данной модели рассматривается задача выбора наилучших кадров при распознавании в видеопотоке и предлагается способ вычисления приоритета кадра с учетом его качества и времени его захвата. Экспериментально показано значительное увеличение точности и скорости работы реализации системы распознавания документов, удостоверяющих личность, в видеопотоке, при моделируемых условиях автоматической фокусировки камеры, приводящих к размытию кадров.

### Глава 3. Оценка качества цветовой сегментации изображения документа

Автоматическая цветовая сегментация (расслоение) изображения на информационные слои является важнейшим этапом работы основанных на модели MRC систем сжатия, что было детально рассмотрено в разделе 1.1.5. Ошибки на этапе расслоения могут приводить к нежелательному для долгосрочного хранения увеличению размера выходных файлов, появлению различных визуальных артефактов и, как следствие, к ошибкам распознавания текстовых данных документа, делающим невозможным эффективное занесение документа в базу данных и эффективный по ней поиск в дальнейшем. Поэтому, системы расслоения необходимо оптимизировать по точности и устойчивости работы.

#### 3.1 Постановка задачи

Рассмотрим упомянутую ранее технологию сжатия изображений цветных печатных документов в формат PDF/A [29]. При внесении очередного изменения в систему расслоения, основанную на данной технологии, разработчики вынуждены собственноручно (возможно, с помощью группы экспертов) проверять качество получившихся в результате работы системы сжатых изображений документов и убеждаться в их приемлемости. Этот процесс занимает много времени, потраченного на рутинную работу, результаты которой впоследствии являются совершенно бесполезными. Такая же ситуация происходит при внесении в логику работы системы различных параметров, пороговых значений и других: требуется ручная настройка для достижения максимального качества работы системы на выборке тестовых изображений. Кроме того, подобная настройка параметров не является точной: при достижении приемлемого результата часто опускаются попытки достижения оптимального результата по причине трудоемкости работы.

На рисунке 3.1 проиллюстрирована текущая схема оценки и оптимизации качества при разработке системы.

Для увеличения эффективности и снижения стоимости контроля качества и настройки параметров системы сжатия изображений документов необходимо использование соответствующих автоматических модулей. На рисунке 3.2

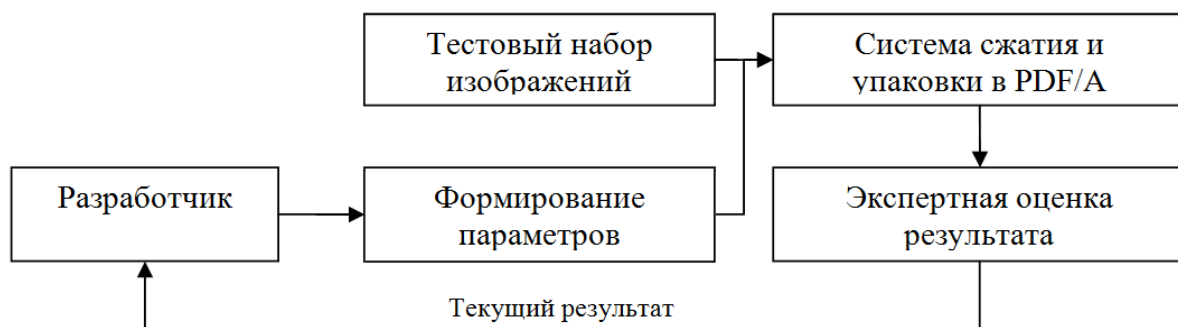


Рисунок 3.1 — Схема ручного процесса оценки качества и настройки параметров системы

представлена желаемая схема с автоматическим процессом оценки качества и настройки параметров.

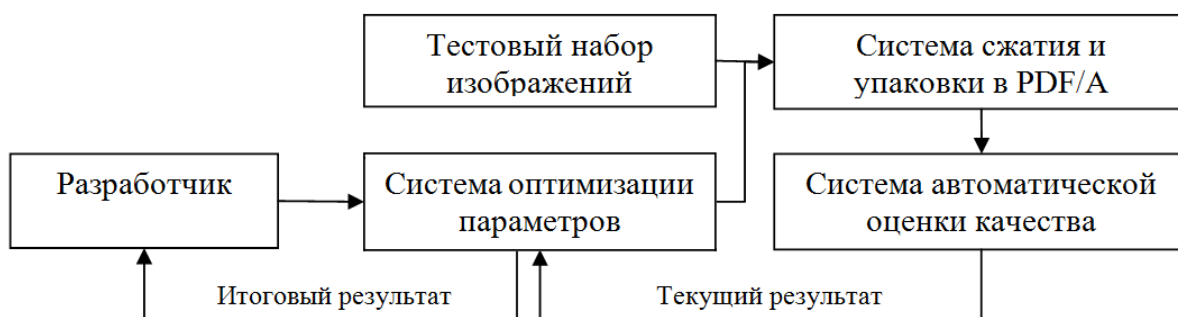


Рисунок 3.2 — Схема автоматического процесса оценки качества и настройки параметров системы

Выделим целевой класс цветных изображений документов, с одной стороны, благоприятных для сжатия основанными на модели смешанного растрового содержимого методами, а с другой стороны – чувствительных к ошибкам цветовой сегментации. Ограничимся изображениями документов, полученных в результате сканирования; у документов однородный белый фон, черный печатный текст (возможно наличие таблиц, форм ввода и т.п.), возможны цветные печати и подписи (рисунок 3.3).

Таким образом, задача заключается в построении автоматических алгоритмов оценки качества цветовой сегментации изображений документов, принадлежащих к целевому классу, которые бы наиболее точно соответствовали экспертным оценкам или другим критериям качества. Критерии качества могут варьироваться в зависимости от того, как результирующие изображения будут использоваться в дальнейшем.



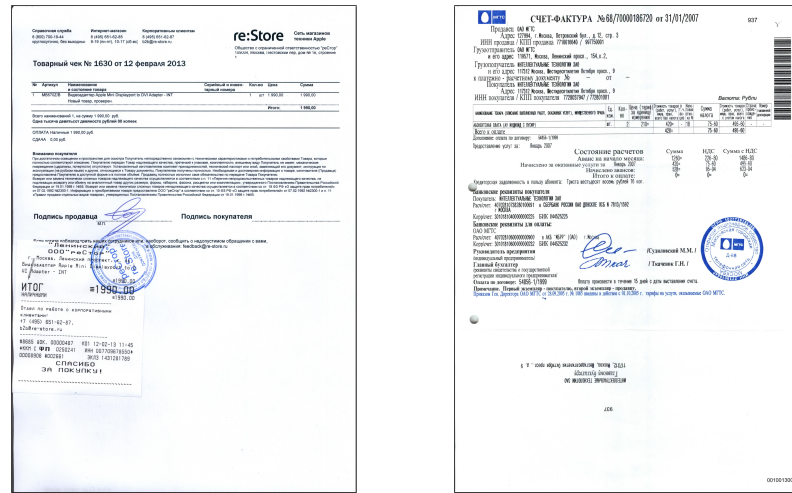


Рисунок 3.3 — Примеры изображений документов рассматриваемого класса

### 3.2 Общий подход к оценке качества цветовой сегментации

Информационные слои изображения печатного документа содержат качественно различную информацию, поэтому проведем их независимую оценку с последующим комбинированием результатов. Обобщенная функция оценки  $Q(I)$  качества цветовой сегментации изображения документа  $I$  представлена в формуле 3.1 и является скалярным произведением вектора весовых коэффициентов слоев на вектор результатов оценки соответствующих слоев:

$$Q(I) = \sum_i w_i Q_i(I), \quad (3.1)$$

где  $w_i$  — задаваемый экспертом весовой коэффициент значимости  $i$  информационного слоя, а  $Q_i$  — функция оценки качества  $i$  типа слоя.

Основной идеей разрабатываемых в данном разделе функций оценок информационных слоев является использование подготовленной вручную разметки, задающей модели качества информационных слоев с учетом специфики принадлежащих к ним объектов. В рамках работы рассматривается оценка качества текстового и графического типа информационного слоя.



### 3.3 Оценка качества текстового слоя с помощью OCR-систем

Текстовый слой имеет важные отличительные особенности: он должен содержать только текстовую информацию, а его изображение удобно представлять монохромным (бинарным, однобитовым). Эти особенности позволяют проверять качество выделения и сжатия изображения в текстовом слое путем попытки его распознавания автоматической системой оптического распознавания символов (OCR). Для оценки качества бинаризации подобный подход описан в работе [160].

Оценка с помощью OCR-систем интуитивно ясна: чем больше точность распознавания (отношение количества верно распознанных символов к общему количеству символов в разметке документа), тем более высокую оценку качества получит изображение. Проблема такого подхода в том, что необходимы большие трудозатраты на предварительную посимвольную разметку текста на изображениях из тестовой выборки.

Модифицируем подход: будем проверять не посимвольное качество распознавания, а точность распознавания слов в некоторых областях изображений (рисунок 3.4). Для каждого изображения из тестовой выборки размечается набор прямоугольных областей, для которых указан набор значимых для распознавания OCR-системой слов.

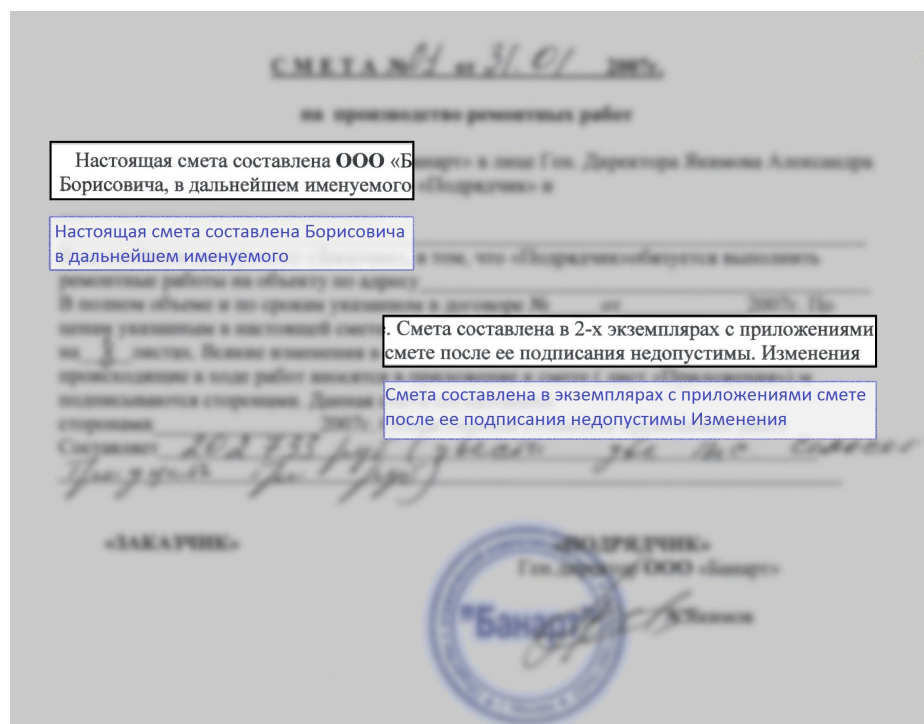


Рисунок 3.4 — Пример разметки областей для последующего распознавания

Для подсчета оценки используется следующая формула:

$$Q_{OCR} = \frac{\sum_{w_i} |w_i| \text{recognized}(w_i)}{\sum_{w_i} |w_i|}, \quad (3.2)$$

$$\text{recognized}(w_i) = \begin{cases} 1, & w_i \text{ распознано полностью,} \\ 0, & \text{иначе,} \end{cases} \quad (3.3)$$

где  $w_i$  – слово из разметки,  $|w_i|$  – его длина (количество символов),  $\text{recognized}(w_i)$  – показатель полностью успешного распознавания слова  $w_i$ .

Такой подход нормализует результаты по отношению к длинам слов, поскольку полностью верно распознанное длинное слово должно давать больший вклад, чем короткое. С другой стороны, неизбежны ошибки в одном символе в длинных словах, которые огрубляют оценку, но подобные ошибки выровняются среди областей всех изображений тестовой выборки. Важно, что данная статистика считается глобально по всем областям разметки изображения.

В качестве внешней OCR-системы для проверки алгоритма была выбрана популярная кросс-платформенная система Tesseract OCR [161].

### 3.4 Оценка качества графического слоя

Графический слой изображения печатного документа должен содержать цветные информативные участки изображения, такие как печати, штампы, подписи и т.п. Наличие в маске слоя не принадлежащих ему по смыслу пикселей ухудшает качество работы алгоритма: размер выходного файла увеличивается в силу применения менее эффективных алгоритмов сжатия; объекты, подлежащие сжатию без потерь, получают различные артефакты. Попадание информативных пикселей графического слоя в текстовый слой означает их последующую бинаризацию, что сильно скажется на визуальном качестве выходного документа.

Требуется разработать алгоритм, оценивающий качество построенной маски слоя с точки зрения отклонения ее от некоторой размеченной вручную маски. Классической функцией оценки такого отклонения является среднеквадратичное отклонение (MSE) для случая двух монохромных изображений  $x$  (построенной

маски) и  $y$  (размеченной маски) размера  $N$ :

$$MSE(x, y) = \frac{1}{N} \sum_{i=1}^N b(x_i, y_i), \quad (3.4)$$

$$b(x_i, y_i) = \begin{cases} 1, & x_i \neq y_i, \\ 0, & x_i = y_i. \end{cases} \quad (3.5)$$

Проблема такого подхода в высокой трудоемкости попиксельной разметки маски графического слоя для тестовой выборки изображений за разумное время.

В разделе 1.1.5 была описана схема расслоения изображения печатного документа технологии PDF/A [29]. Она основывается на построении гистограммы насыщенности изображения и последующей бинаризации карты насыщенности. Картой насыщенности (рисунок 3.5) называется полутоновое изображение, в котором каждому пикселю  $(R, G, B)$  исходного изображения соответствует значение его насыщенности  $S$ , при работе с 8-битными изображениями вычисляемое по формуле:

$$S = 255 - \frac{3 \cdot \min(R, G, B)}{R + G + B}. \quad (3.6)$$

Для упрощения вычислений можно использовать формулу:

$$S = \max(R, G, B) - \min(R, G, B). \quad (3.7)$$

Группа инвалидности первая (указывается прописью)  
 Причина инвалидности общее заболевание (указывается прописью)  
 Степень ограничения способности к трудовой деятельности третьей (указывается прописью)  
 Инвалидность установлена на срок до 01.01.2007 (указывается прописью)  
 Дата очередного освидетельствования 29.12.2006  
 Причина пропуска срока переосвидетельствования за прошлое время с \_\_\_\_\_ по \_\_\_\_\_ признана уважительной (неуважительной) (нужное подчеркивается)  
 Инвалидность с ограничением способности к трудовой деятельности (категория «ребенок-инвалид») за прошлое время с \_\_\_\_\_ по \_\_\_\_\_ установлена (не установлена) (нужное подчеркивается)  
 Основание: акт освидетельствования в учреждении Государственной службы медико-социальной экспертизы № 43 от 29.12.2006 2004 г.  
 Дата выдачи выписки 14.01.2007 2005 г.  
 Руководитель учреждения Государственной службы медико-социальной экспертизы (подпись) \_\_\_\_\_ (расшифровка подписи) \_\_\_\_\_  
 М.П. \_\_\_\_\_



Рисунок 3.5 — Исходное изображение и его карта насыщенности

С целью повышения удобства работы для пользователя, полученная карта насыщенности инвертируется:  $\bar{S}_i = 255 - S_i$  (рисунок 3.6a).

Следующим шагом схемы расслоения является глобальная пороговая бинаризация (метод Оцу [32]), т.е. вычисляется некоторое единственное пороговое значение, в соответствии с которым пиксели после вычисления насыщенности попадают в текстовый или графический слой. Идею пороговой бинаризации удобно использовать для эффективной разметки маски: при ручном изменении порога насыщенности происходит монотонное добавление (удаление) в разметку графического слоя всех удовлетворяющих (не удовлетворяющих) этому порогу пикселей.

Для разметки инвертированной карты насыщенности экспертом выделяется набор непересекающихся прямоугольных областей, и для области устанавливается оптимальный с точки зрения визуального восприятия порог бинаризации пикселей, принадлежащих этой области (рисунок 3.6б).

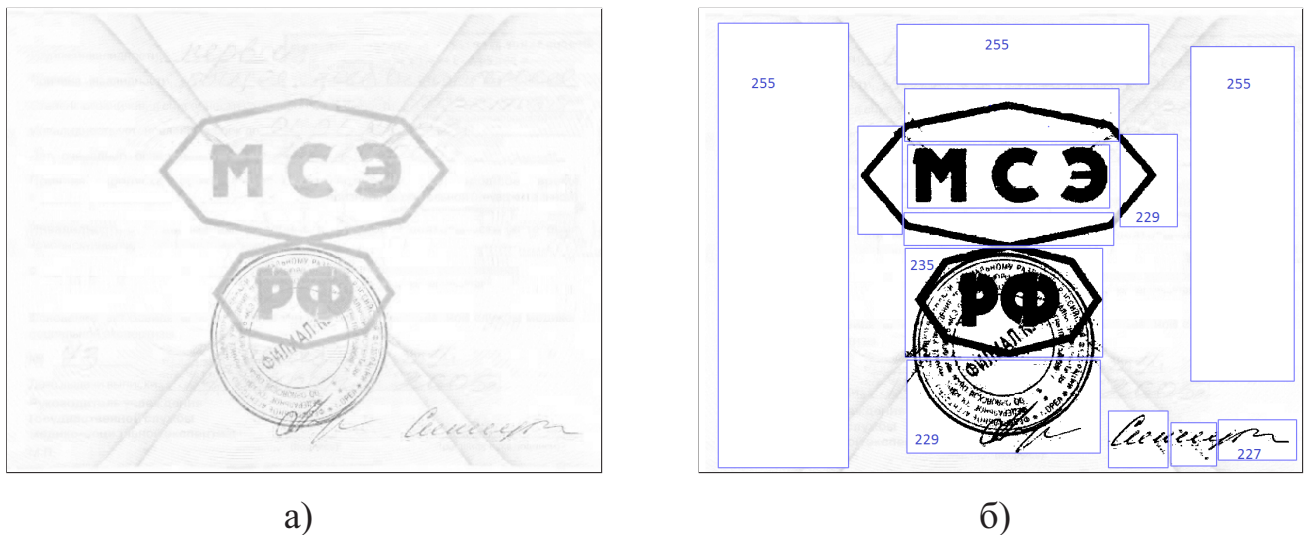


Рисунок 3.6 — Инвертированная карта насыщенности и разметка порогов бинаризации для ее областей

Для непосредственной оценки качества будем использовать две метрики, основанные на описанном подходе. Первой из них является вычисление суммарного MSE для размеченных областей:

$$Q_{BINMSE}(X, Y) = \sum_{i=1}^{N_A} MSE(X_i, Y_i), \quad (3.8)$$

где  $X$  – исходное изображение (инвертированная карта насыщенности),  $Y$  – бинаризованное изображение (полученная маска графического слоя),  $X_i$  – соответствующая  $i$  области часть разметки карты насыщенности,  $Y_i$  – соответствующая  $i$  области часть графического слоя,  $N_A$  – количество областей разметки.

У данной метрики есть недостаток – вообще говоря, идеальную разметку расслоения составить невозможно: от присутствия или отсутствия в маске слоя некоторых пикселей (чаще всего, на границе объектов) визуальное восприятие качества может остаться прежним. Используем следующий способ оценки: вместо расчета попиксельного отклонения проводить расчет отклонений плотностей пикселей для целых областей.

Введем величину  $\rho(A)$  плотности пикселей для области разметки  $A$  размерами  $W_A \times H_A$ :

$$\rho(A) = \frac{N_1(A)}{N_0(A) + N_1(A)} = \frac{N_1(A)}{W_A \cdot H_A}, \quad (3.9)$$

где  $N_k(A)$  – количество пикселей в области разметки  $A$  со значением  $k \in \{0, 1\}$ .

Тогда, второй функцией оценки будет:

$$Q_{BINCNT}(X, Y) = \frac{\sum_{i=1}^{N_A} (\rho(X_i) - \rho(Y_i))^2 \cdot (W_{A_i} \cdot H_{A_i})}{\sum_{i=1}^{N_A} (W_{A_i} \cdot H_{A_i})}. \quad (3.10)$$

Плотность пикселей – относительная величина, поэтому отклонения плотностей взвешиваются по площадям областей.

### 3.5 Анализ пригодности разработанных алгоритмов оценки цветовой сегментации

Для измерения пригодности алгоритма оценки слоя используем коэффициент линейной корреляции Пирсона между оценками алгоритма и средней экспертной оценкой (MOS) на сегментированной фиксированным алгоритмом тестовой выборке изображений. В качестве экспериментальной системы расслоения возьмем реализация технологии PDF/A из [29].

Экспертное оценивание результатов цветовой сегментации собранной и размеченной тестовой выборки из 15 изображений документов проводилось среди 5 экспертов и показало высокую их согласованность. В опросе использовалась 20-балльная шкала. На рисунке 3.7 представлены вычисленные значения корреляции, отмечающие, что для оценки качества графического слоя в самом деле выгоднее использовать алгоритм оценки отклонений плотностей пикселей.



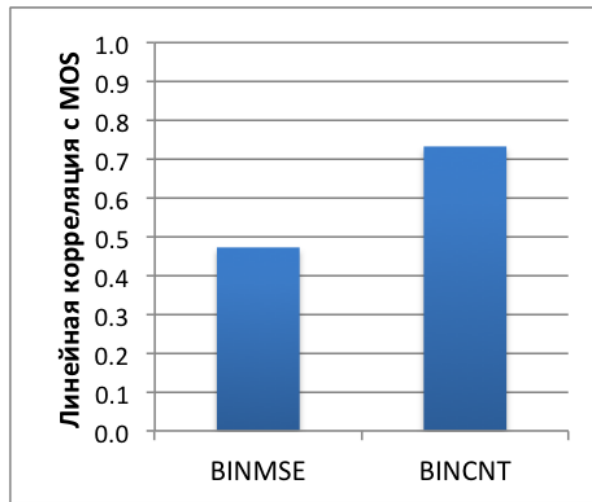


Рисунок 3.7 — Вычисленные значения линейной корреляции между MOS и оценками разработанных алгоритмов на тестовой выборке изображений

Таким образом, для оценки графического слоя выберем функцию  $Q_{BINCNT}$ , от чего итоговая функция оценки качества определяется, как следующая линейная комбинация независимых оценок слоев:

$$Q(I) = w_1 \cdot Q_{BINCNT}(I) + w_2 \cdot Q_{OCR}(I), \quad (3.11)$$

где веса  $w_1$  и  $w_2$  выставляются экспертом для регулировки значимости графического и текстового слоя. В экспериментах  $w_1$  и  $w_2$  принимались равными 0.5.

### 3.6 Автоматическая оптимизация качества цветовой сегментации

При наличии автоматического алгоритма оценки качества цветовой сегментации становится возможной автоматическая оптимизация качества. По причине сложной структуры как системы цветовой сегментации, так и системы оценки качества, строится система типа «черный ящик». Входами являются некоторые внутренние параметры, влияющие на результат цветовой сегментации и размеченная тестовая выборка изображений документов, а выходом – значение оценки качества для результата расслоения с данными параметрами.

Был проведен эксперимент с 4 параметрами (целыми, с диапазоном от 0 до 255) цветовой сегментации. Первые два параметра представляют собой опорные точки кусочно-линейной функции порога насыщенности на классификацию пикселя к текстовому заполнению в зависимости от значения  $\min(R, G, B)$ , Другие



два параметра аналогичны первым, но определяют принадлежность пикселя к графическим областям.

Для автоматической оптимизации использовалось программное обеспечение оптимизации систем типа «черный ящик» NOMAD, реализующее алгоритм MADS – Mesh Adaptive Direct Search [162]. На рисунке 3.8 приведены примеры фрагментов слоев проблемного изображения тестовой выборки, полученные в результате работы системы при исходных параметрах (начальном приближении), заданными разработчиком.

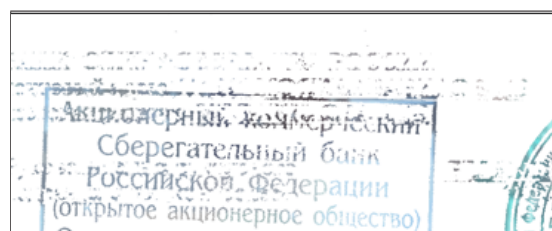
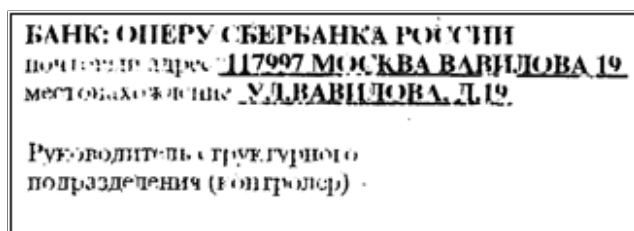


Рисунок 3.8 — Пример проблемных фрагментов текстового и графического слоя до проведения оптимизации на тестовой выборке изображений

На рисунке 3.9 изображены фрагменты того же изображения, полученные при автоматически подобранных оптимальных параметрах.

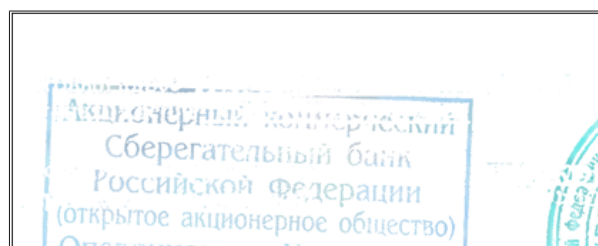
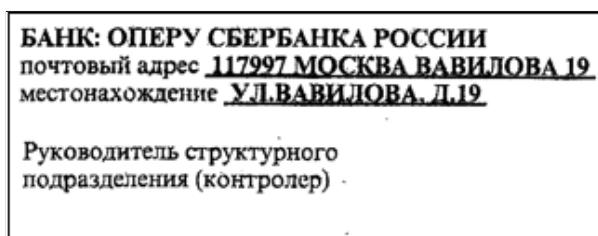


Рисунок 3.9 — Пример оптимизации качества проблемных фрагментов текстового и графического слоя

До оптимизации часть маски текстового слоя попадала в графический, тем самым делая нечитаемым текст, а также создавая сильный шум в области печати на графическом слое. Разработанные в разделе алгоритмы оценки качества учитывают оба этих фактора, поэтому оптимизация прошла успешно с точки зрения визуального восприятия человека. Для всех изображений выборки качество проблемных (при различных вариантах исходных параметров алгоритма цветовой сегментации) изображений также визуально улучшилось, а качество изначально удачных – осталось прежним.

### 3.7 Выводы по главе 3

В данной главе был разработан алгоритм автоматической оценки качества цветовой сегментации изображений печатных документов, учитывающий особенности текстовых и графических типов информационных слоев и проводящий их независимую оценку с целью уточнения соответствия вычисленных оценок критериям качества, следующих из сфер последующего использования данных изображений документов. Дополнительной особенностью алгоритма является ясность и простота трактовки используемых критериев качества, что важно при разработке алгоритмов цветовой сегментации.

## Глава 4. Оценка качества изображений при контроле доступа и безопасности на предприятии

### 4.1 Детектирование и фильтрация бликов в задачах распознавания документов с мобильных устройств

Одним из типов искажений, возникающих при съемке в естественной среде, являются световые отражения и блики [10], рассмотренные в разделе 1.2 (рисунок 4.1).

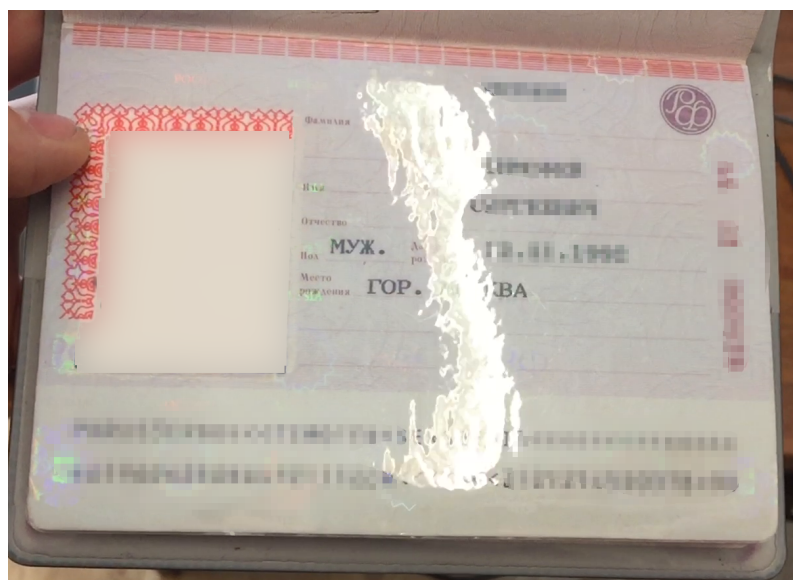


Рисунок 4.1 — Пример выраженного блика на изображении паспорта РФ

Исследовательские системы распознавания и компьютерного зрения обычно разрабатываются с тем допущением, что на поверхностях рассматриваемых объектов отсутствуют блики. Это допущение упрощает их построение ценой снижения устойчивости их работы в условиях наличия бликов, которые воспринимаются как аномальные явления или выбросы [111].

Для повышения устойчивости работы таких систем в условиях наличия бликов требуется детектировать области с бликами на входных изображениях. Теоретически, если в области блика сигнал яркости не вышел за пределы динамического диапазона, то возможно восстановление изображения в области блика, как если бы блик отсутствовал [112]. К сожалению, динамический диапазон камер мобильных устройств, как правило, крайне ограничен. В таком случае область

блика оказывается полностью зашкалена и восстановление такого изображения возможна только в смысле корректной интерполяции фона.

Важным качеством систем распознавания является способность игнорировать соответствующие области изображений, используя информацию о местонахождении бликов, чтобы не возвращать заведомо ошибочный результат, не тратить лишнее время на обработку, а также снизить показатель уверенности в результате или выдать отказ, что было подробно рассмотрено в главе 2.

Системы распознавания документов в видеопотоке [10; 55], например, описанная в работе [21], характерны тем, что даже при частичных повреждениях областей документа на кадрах видеопотока в результате бликов они все равно потенциально способны распознать документ. Это возможно тогда, когда на протяжении видеопотока каждая информативная составляющая документа отчетливо представлена хотя бы на одном или нескольких кадрах, и достигается за счет комбинирования («интеграции») результатов распознавания между различными кадрами видеопотока.

Таким образом, задача заключается в увеличении точности и устойчивости методов распознавания документов с помощью детектирования поврежденных бликами областей изображения документа и интерполяции поврежденного ими фона. При этом необходимо обеспечивать быстроедействие, достаточное для работы с видеопотоком в реальном времени на мобильном устройстве.

#### **4.1.1 Влияние бликов на системы распознавания документов**

Опишем влияние бликов и возможные способы взаимодействия с ними на примере системы распознавания документов [21], близкой по своему устройству к общей схеме, рассмотренной в разделе 1.1.4. Блики могут служить серьезной помехой практически на всех шагах работы системы.

При наличии бликов относительно большой площади, детектор документа (например, алгоритмом Виолы-Джонса [163]) способен отрицательно классифицировать область документа, в случае чего система выдает отказ и дальнейшее распознавание не производится. Такое поведение системы оправдано только тогда, когда в данном видеопотоке присутствуют кадры без засвеченных областей, что в естественных условиях возможно не всегда. В противном случае, даже

если каждый кадр в отдельности и невозможно распознать целиком по причине засвета текстовой информации, система может аккумулировать результаты распознавания видимых частей кадров видеопотока. Для повышения устойчивости детектирования документа и уменьшения количества отказов подсистемы на данном этапе предлагается фильтрация бликов в поврежденных областях путем интерполяции фона.

Методы выделения строк полей зачастую используют алгоритмы математической морфологии, а также проекции на вертикальные оси для выделения областей текста. При замене темных текстовых областей с видимыми границами базовых линий светлыми бликовыми областями, вышеупомянутые алгоритмы ведут себя непредсказуемым образом и выдают либо отказ, либо неточный результат.

Аналогичная ситуация наблюдается с алгоритмами сегментации выделенных строк на символы и последующим их распознаванием: корректное применение этих алгоритмов в областях бликов не определено и система должна дать отказ вместо того, чтобы тратить время на заведомо безуспешную операцию.

Таким образом, общая методология взаимодействия с бликами на всех уровнях распознавания документа такова: если возможно – избежать отказа метода на текущем этапе, сохранив полезную информацию для выделения на следующих этапах. Иначе – выдать отказ и избежать заведомо бесполезной траты вычислительных ресурсов.

#### 4.1.2 Детектирование областей бликов

Проведем грань между произвольными отражениями света, изменяющими наблюдаемые цветовые характеристики документа, и бликами, приводящими к зашкаливающим значениям на полученных с камеры изображениях. Упрощение постановки задачи путем рассмотрения только бликов позволяет разработать простой и быстрый алгоритм детектирования бликовых областей. Результатом алгоритма является бинарное изображение  $G$  – маска принадлежности пикселей исходного изображения бликовым областям.

Принадлежащие блику пиксели изображения могут быть охарактеризованы двумя признаками: высокая яркость и низкая насыщенность [126] в результате

потери цветовой информации о документе. Тогда можно считать, что пиксель  $(x, y)$  изображения точно принадлежит блику, если одновременно выполняется  $I(x, y) > T_I$  и  $S(x, y) < T_S$ , где  $T_I$  и  $T_S$  – пороги на яркость и насыщенность. Помимо непосредственных зашкаливающих областей требуется также включить в маску менее засвеченные области, граничащие с уже найденными, что дополнительно обуславливается неравномерностью поверхности документа – например, из-за защитной пленки. Для расширения границ найденной маски можно применить операции математической морфологии. Первым шагом производится морфологическое закрытие битовой маски с целью закрашивания небольших бассейнов с отрицательными значениями маски, окруженных ранее найденными положительными. Затем, положительные границы маски расширяются с помощью морфологической операции дилатации. На рисунке 4.2 показан результат работы алгоритма детектирования.

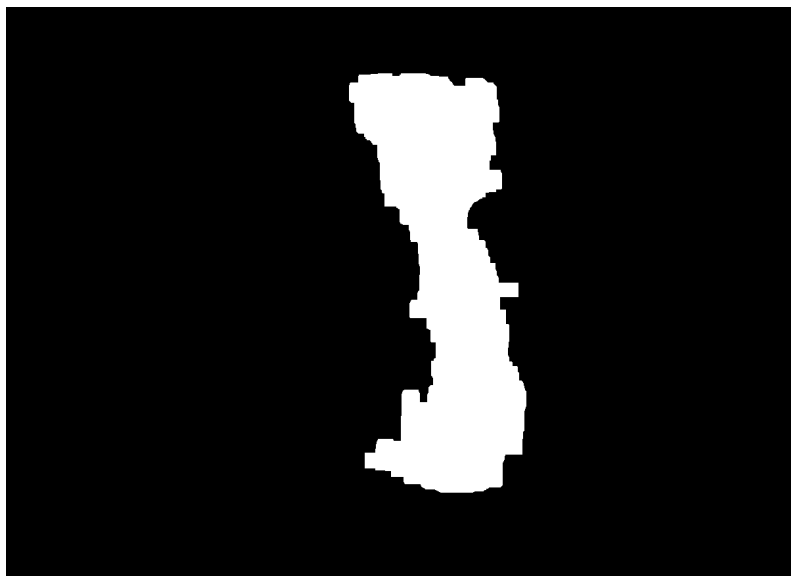


Рисунок 4.2 — Результат детектирования поврежденной бликом области

#### 4.1.3 Интерполяция фона на поврежденных бликами областях

Для восстановления поврежденных областей будем использовать вычислительно простой алгоритм, основанный на обходе в ширину. Компоненты связности бинарной маски бликов обрабатываются независимо друг от друга. Для заданной компоненты связности поддерживается очередь  $Q$  порядка обхода внутренних



пикселей поврежденной области, сначала в нее заносятся все ее внутренние граничные пиксели. На каждой итерации алгоритма текущий незакрашенный поврежденный бликом пиксель  $(x, y)$  достаётся из очереди и его значение для цветовых компонент  $c$  пересчитывается, как:

$$I'(x, y, c) = \frac{1}{|N(x, y, W)|} \sum_{x', y' \in N(x, y, W)} I(x', y', c), \quad (4.1)$$

где  $N(x, y, W)$  – множество не принадлежащих блику, или же уже закрасенных бликовых пикселей в окне с центром  $(x, y)$  и крылом  $W$ . После этого, еще не закрасенные бликовые пиксели, являющиеся непосредственными соседями  $(x, y)$ , добавляются в очередь  $Q$ . На рисунке 4.3 показан пример работы алгоритма.

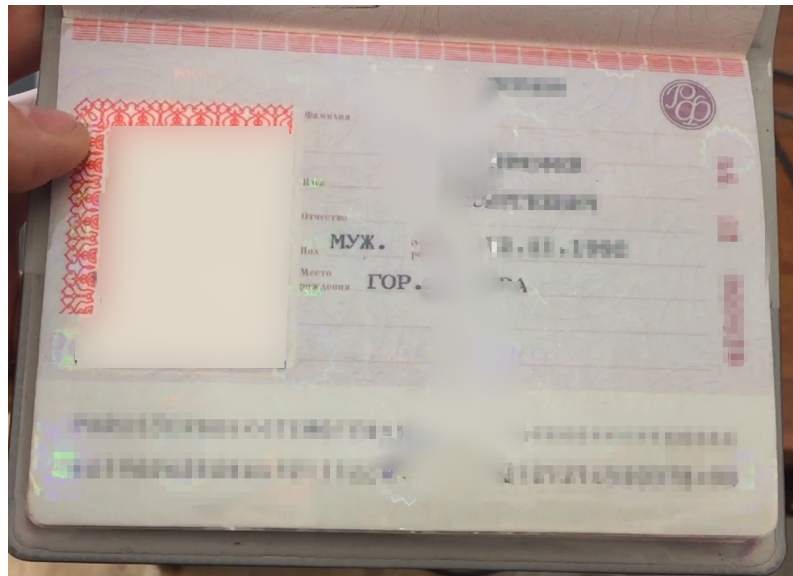


Рисунок 4.3 — Результат работы алгоритма фильтрации областей бликов

#### 4.1.4 Экспериментальные результаты

В экспериментах использовался набор из 579 изображений 3-й страницы (покрытой пленкой) паспортов РФ, полученных с камер мобильных устройств. Съемка проводилась в таких условиях, чтобы на документах проявлялись блики, возникающие от направленных источников света. Блики на изображениях выборки располагаются в различных частях документа, тем самым повреждая его различные информационные области, а также могут иметь достаточно большую площадь для негативного влияния на детектор документа.

Первый этап первого эксперимента заключался в тестировании детектора 3-й страницы паспорта РФ на всех изображениях выборки и отделении негативных изображений, на которых детектор выдал отказ (детектор настроен на практическое отсутствие ложноположительных срабатываний). Затем, вторым этапом, на выделенных изображениях производилась детектирование поврежденных бликами областей и их восстановление. Третий этап заключался в повторном запуске системы детектирования документа, но уже на изображениях с интерполированными областями бликов.

На рисунке 4.4 показаны пропорции количества изображений с найденными документами на первом этапе и не требующих закрашивания бликов (75.3%), изображений с отрицательными результатами детектирования, ставшими положительными после закрашивания бликов (19.4%), а также изображений, на которых даже после закрашивания не удалось получить успешное срабатывание детектора (5.3%).

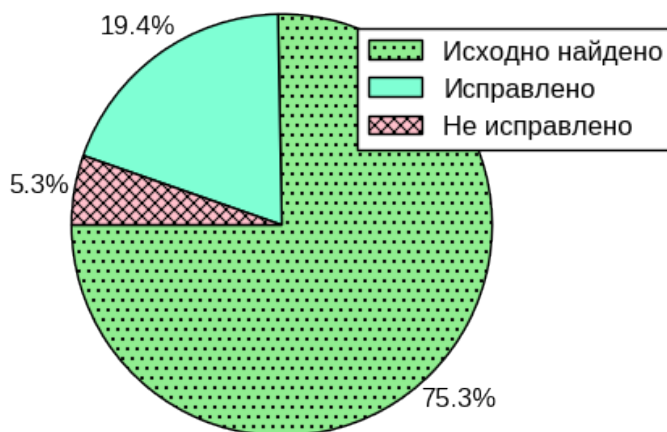


Рисунок 4.4 — Статистика работы детектора документа до и после восстановления поврежденного бликами фона

Поскольку отрицательное срабатывание детектора приводит к невозможности дальнейшего распознавания, исправление детектирования одной пятой части выборки документов, что также является уменьшением числа ошибок в 4.6 раз, можно считать существенным.

Для проверки непосредственного влияния исправления изображений для детектора на общую точность распознавания был проведен второй эксперимент, в котором подсчитывалось среднее количество частично и полностью распознанных полей 3-й страницы паспорта на исходных и восстановленных изображениях. К полям относились: серия, номер, имя, фамилия, пол, дата и место рождения. Результаты эксперимента показаны на рисунке 4.5.

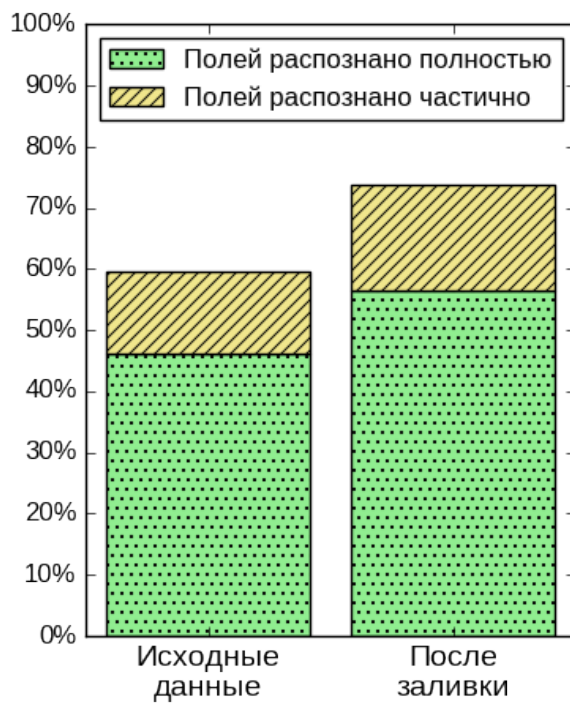


Рисунок 4.5 — Точность распознавания документа до и после восстановления поврежденного бликами фона

Из иллюстрации видно, что общая точность распознавания также увеличилась при рассмотрении восстановленных изображений, ранее не распознаваемых по причине отказа детектора документа.

## **4.2 Анализ устойчивости алгоритма проективного оптического потока к бликам**

Задача нахождения проективного оптического потока возникает при отслеживании движения плоских ригидных документов в видеопотоке, в частности при сканировании документов и распознавании их текстового наполнения.

Вычисление оптического потока позволяет построить единую координатную систему между кадрами, что открывает широкий спектр ранее недоступных подходов к анализу документов, учитывающих не только покадровые результаты распознавания, но и другую накопленную информацию, которую возможно экстраполировать на последующие кадры, например, информацию о положении и форме присутствующих на документе объектов.

Оптический поток может быть использован в случаях, когда время его вычисления меньше, чем время работы алгоритмов поиска документа. В таких случаях, с целью повышения быстродействия, оправданно прореживание запусков дорогостоящих алгоритмов поиска границ пересчетом ранее найденных границ с помощью межкадрового проективного преобразования.

Наконец, оптический поток позволяет решать задачу составления одиночного изображения документа, более высокого качества и разрешения, чем его представление на каждом отдельно взятом кадре видеопотока – задачу сверхразрешения, актуальную даже при обычной съемке документов без их распознавания [164].

Помимо отслеживания документов, задача поиска проективного преобразования между изображениями плоской сцены возникает при съемке земной поверхности с большой высоты с помощью БПЛА для решения задачи визуальной навигации и картирования местности [165].

Одной из существенных проблем при поиске проективного оптического потока является наличие бликов на изображении, в том числе при сканировании документов в естественной среде – например, при съемке с мобильных устройств [166]. Наличие блика на изображениях может приводить к значительным ошибкам при вычислении оптического потока, так как его движение рассогласовано с движением сцены относительно наблюдателя (видеокамеры). Таким образом, возникает задача разработки алгоритмов вычисления проективного оптического потока, устойчивых к бликам.

В данном разделе рассматривается модификация классического алгоритма оптического потока [167] и проводится анализ устойчивости алгоритма к бликам в сравнении с исходным алгоритмом.

#### 4.2.1 Подавление бликов преобразованием цветового пространства

В работе [168] используется линейная модель формирования цветных изображений [169]. Согласно данной модели, вектор-стимул (далее цвет)  $\bar{c}(x, y)$  в пикселе изображения с координатами  $(x, y)$  может быть представлен как сумма двух компонент:

$$\bar{c}(x, y) = g_b(x, y) \cdot \bar{c}_b + g_s(x, y) \cdot \bar{c}_s, \quad (4.2)$$

где  $g_b(x, y), g_s(x, y)$  – геометрические коэффициенты диффузной и зеркальной компонент, содержащие в себе информацию о геометрии сцены;  $\bar{c}_b$  – вектор, зависящий от спектрального состава диффузной компоненты окраски объекта и спектрального состава излучения источника света, а  $\bar{c}_s$  – вектор, аналогичным образом зависящий от зеркальной компоненты и источника [170]. В работе предполагается, что для всех объектов сцены вектора зеркальных компонент  $\bar{c}_s$  коллинеарны.

Важно заметить, что различие значений соответствующих пикселей затрудняет решение задачи вычисления проективного оптического потока, поскольку очевидный признак соответствия – совпадение наблюдаемого цвета – оказывается неверным. На практике считается, что величина  $g_b(x, y)$  слабо зависит от точки наблюдения сцены, в отличие от значения  $g_s(x, y)$ . Поэтому, согласно (4.2), с изменением ракурса съемки меняется и  $\bar{c}(x, y)$ , то есть цвет не инвариантен к смене ракурса и не может быть корректно использован алгоритмами оптического потока или стереосопоставления. В линейной модели любое изображение может быть представлено, как сумма двух: диффузного  $g_b(x, y) \cdot \bar{c}_b$  и зеркального  $g_s(x, y) \cdot \bar{c}_s$ . И лишь диффузное изображение может быть использовано в качестве инварианта для сопоставления, так как  $g_b(x, y)$  слабо зависит от точки наблюдения сцены, и, таким образом, возникает задача его восстановления.

Легко показать, что задача вычисления  $g_b(x, y) \cdot \bar{c}_b$  без дополнительных предположений не имеет единственного решения: считая, что зеркальная компонента  $\bar{c}_s$  известна, система уравнений (4.2) относительно  $g_b(x, y) \cdot \bar{c}_b$  и  $g_s(x, y)$  недоопределена. Для получения единственного решения должны быть использованы предположения о текстурных особенностях поверхности [171], или знания о том, как пиксели на нескольких изображениях соответствуют друг другу. В качестве примера можно привести алгоритмы одновременного вычисления карты диспаратности и диффузного изображения [172].

Предлагается использовать другой цветовой инвариант, являющийся параллельной проекцией цветов из  $RGB$  пространства изображения на плоскость  $\alpha\beta$ , нормаль которой совпадает с направлением зеркальной компоненты [170]. Векторы, полученные в результате описанного преобразования, содержат информацию только о форме и цветности объекта, следовательно, их можно использовать для задач сопоставления пикселей. Для того, чтобы построить такое инвариантное подпространство, должно быть известно направление  $\bar{c}_s$ , либо должен существовать способ его определения (например, в работе [170] описан алгоритм поиска цветности источника для случая «простых» сцен). Кроме того, угол между зеркальной и диффузной компонентами должен быть достаточно большим, чтобы отношение сигнала к шуму не стало критическим для алгоритма сопоставления [170]. Таким образом, если известен вектор зеркальной компоненты и выполнены все вышеописанные условия, изображения, полученные в результате параллельной проекции цветов их пикселей из пространства  $RGB$  на плоскость  $\alpha\beta$ , инвариантны к смене ракурса, и следовательно могут быть использованы при вычислении проективного оптического потока.

#### 4.2.2 Алгоритм вычисления оптического потока

Рассмотрим обобщение классического алгоритма оптического потока Лукаса-Канаде [167]. Введем требуемые обозначения. Пусть  $I_\tau(x)$ ,  $I_\varphi(x)$  – исходные изображения;  $W(x, p)$  – некоторое параметрическое преобразование, где  $p$  – вектор параметров.  $W(x, p)$  задает проективное преобразование из СК изображения  $I_\varphi(x)$  в СК изображения  $I_\tau(x)$ .



С учетом введенных обозначений, задача поиска проективного оптического потока формулируется следующим образом – найти такое «малое приращение»  $\Delta p$  при котором достигается минимум функционала, если известно начальное приближение  $p$ :

$$\sum_x \left[ I_\tau(x) - I_\phi(W(x, p + \Delta p)) \right]^2 \rightarrow \min_{\Delta p}, \quad (4.3)$$

где значения  $I_\phi(W(x, p + \Delta p))$  вычисляются при помощи билинейной интерполяции.

В задаче оптического потока 4.3 все пиксели учитываются в оптимизируемом функционале с одинаковым весом. Однако, оценка достоверности значения в разных пикселях могут различаться. К примеру относительное влияние темнового тока матрицы фотоаппарата различно для пикселей с разными значениями, что может быть учтено в функционале с помощью соответствующей маски весов.

Другим примером является рассмотренный в предыдущем разделе эффект зашкаливания в результате бликов, когда значение яркости по одному из каналов достигает максимально возможного значения. Модификация алгоритма в целях борьбы с бликами вычисляет оптический поток в  $\alpha\beta$  пространстве. В результате эффекта зашкаливания происходит частичная потеря информации о цвете такого пикселя, поэтому переход в подпространство  $\alpha\beta$  некорректен. В таком случае целесообразно использовать функционал, позволяющий нивелировать вклад таких пикселей после детектирования маски бликов.

Таким образом, добавление мультипликативных весов в функционал 4.3 для учета оценки достоверности значений пикселей, теоретически может увеличить точность вычисления преобразования  $W(x, p)$ . Поэтому, в модификации алгоритма учитывается информация о том, для каких пикселей значение яркости по одному из каналов превосходит максимальное допустимое значение. Такие пиксели учитываются при вычислении проективного преобразования с весом 0, остальные – с весом 1. Опишем алгоритм [168] подробнее.

Через  $\omega(x, W(x, p)) = \omega_\tau(x) \cdot \omega_\phi(W(x, p))$  обозначим функцию весов, где  $\omega_\tau(x)$  и  $\omega_\phi(x)$  функции весов изображений  $I_\phi(x)$  и  $I_\tau(x)$ . Тогда, задача 4.3 может быть записана следующим образом:

$$\sum_x \left[ I_\tau(x) - I_\phi(W(x, p + \Delta p)) \right]^2 \cdot \omega(x, p + \Delta p) \rightarrow \min_{\Delta p}. \quad (4.4)$$

Обозначим  $E(x, p) = I_\tau(x) - I_\varphi(W(x, p))$ . Для краткости дальнейшего описания все не зависящие от  $\Delta p$  функции будем приводить без указания их переменных и параметров.

Как и в [167], считая  $\Delta p$  малой величиной, аппроксимируем  $I_\varphi(W(x, p + \Delta p))$  в ряд Тейлора первого порядка по  $\Delta p$ , при этом  $\omega(x, p + \Delta p)$  аппроксимируем через  $\omega(x, p)$ :

$$\sum \left( E - \nabla I_\varphi \frac{\partial W}{\partial p} \Delta p \right)^2 \cdot \omega \rightarrow \min_{\Delta p}. \quad (4.5)$$

Такая аппроксимация  $\omega(x, p + \Delta p)$  разумна, поскольку для малых  $\Delta p$  карта весов будет смещена незначительно, внося малый вклад в изменение функционала 4.5. Для поиска минимума продифференцируем выражение 4.5 и приравняем результат к нулю:

$$\sum \left( E - \nabla I_\varphi \frac{\partial W}{\partial p} \Delta p \right) \omega \left( \nabla I_\varphi \frac{\partial W}{\partial p} \right)^T = 0, \quad (4.6)$$

из чего получается решение оптимизационной задачи 4.5:

$$\Delta p = H^{-1} \sum_x \left( I_\tau(x) - I_\varphi(W(x, p)) \right) \omega(x, p) \left( \nabla I_\varphi \frac{\partial W}{\partial p} \right)^T \quad (4.7)$$

где

$$H = \sum_x \omega(x, p) \left( \nabla I_\varphi \frac{\partial W}{\partial p} \right)^T \nabla I_\varphi \frac{\partial W}{\partial p}. \quad (4.8)$$

Вычисленное по формулам 4.7 и 4.8 приращение  $\Delta p$  затем прибавляется к начальному приближению  $p$ :

$$p := p + \Delta p. \quad (4.9)$$

Полученный вектор параметров затем используется в качестве нового начального приближения и шаги 4.7, 4.8 и 4.9 снова выполняются. Данный процесс повторяется  $n$  раз и предполагается, что такого числа повторений достаточно для его сходимости.

Кроме того, для уменьшения влияния выбора начального приближения на результат вычислений используется следующий прием. Предварительно создается последовательность пар сжатых изображений  $I_\tau(x)$  и  $I_\varphi(x)$ , отсортированных в порядке убывания степени сжатия. Затем последовательно, начиная с самых сжатых пар, происходит вычисление вектора параметров  $p$  по указанной ранее

процедуре. Далее вычисленное  $p$  используется в качестве начального приближения для вычисления параметров проективного преобразования на следующей в списке паре сжатых изображений.

Описанный алгоритм назовем иерархической итеративной процедурой вычисления взвешенного оптического потока. Проведем анализ устойчивости данного алгоритма к бликам.

### 4.2.3 Описание тестового набора данных

Для проведения численных экспериментов были получены наборы синтетических и реальных данных. Все данные были размечены следующим образом: на каждом изображении в рамках одной последовательности кадров отмечались четыре расположенные на максимальном отдалении друг от друга точки объекта (для уменьшения влияния погрешности разметчика), никакие три из которых не лежат на одной прямой. Полученная разметка использовалась для вычисления проективного преобразования между двумя любыми кадрами в рамках одного набора изображений.

**Синтетические данные** Синтетические данные (SD) были получены при помощи программы рендеринга V-Ray (<https://www.chaosgroup.com/vray/maya>). Создана последовательность из двухсот трехканальных изображений (с разрешением 1920x1440) плоского текстурированного прямоугольника (шахматная доска, рисунок 4.6).

Сцена была освещена одним источником света известной цветности. Мощность источника и параметры поверхностей были подобраны таким образом, чтобы на ней присутствовал ярко выраженный блик. Камера двигалась по дуге окружности, так, что оптическая ось камеры в процессе движения проходила через центр наблюдаемого прямоугольника. Угол между плоскостью траектории и плоскостью сцены выбран равным тридцати градусам. Подробное описание конфигурации сцены и параметров движения камеры приведено в [168].

**Реальные данные** Реальные данные представляют собой последовательности изображений сцены с плоским текстурированным объектом на темном фоне,

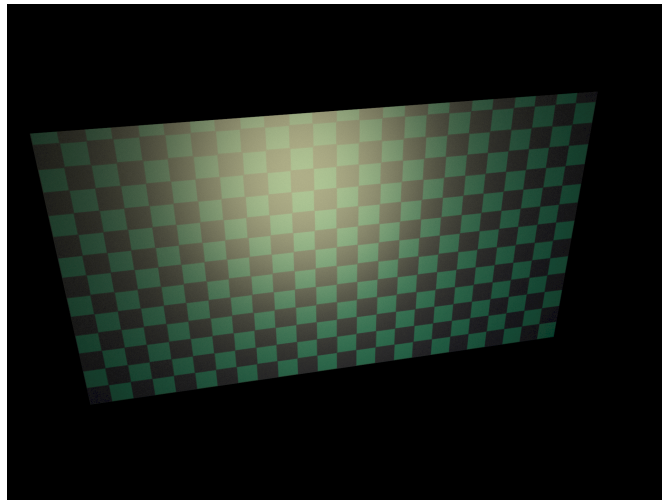


Рисунок 4.6 — Пример изображения из синтетического набора данных

Таблица 1 — Описание характеристик наборов данных

Данные	SD	TD	BD
Размер изображений	$1920 \times 1440$	$1728 \times 1151$	$1728 \times 1151$
Количество изображений	200	56	50
Шаг $s$	25	5	5
Количество пар $O_s$	1592	660	490

освещенной ахроматическим источником света. В качестве объектов использовались карта «Тройка» (TD) и банковская карта (BD). Конфигурации каждой сцены во время съемки оставались неизменными. Все изображения были сняты на фотоаппарат Canon EOS550D. Предварительно была произведена гамма-коррекция изображений для линейаризации цветового пространства. Съемка плоского объекта на фоне темной подложки производилась с различных ракурсов. В таблице 1 приведены основные характеристики данных.

#### 4.2.4 Описание экспериментального исследования

Для удобства описания проведенных экспериментов введем следующие обозначения:

–  $V = \{1, 2, \dots, n\}$  – упорядоченное множество индексов последовательности изображений;

- $O = \{(i, j) : i, j \in V, i \neq j\}$  – множество пар индексов  $(i, j)$ , принадлежащих  $V$ , изображений  $I_i$  и  $I_j$ ;
- $I_i^c$  – изображение  $I_i$ , цвета пикселей которого принадлежат цветовому пространству  $c \in \{RGB, \alpha\beta\}$ ;
- $P_{markup}(I_i, I_j)$  – истинное проективное преобразование между изображениями  $I_i$  и  $I_j$ ;
- $P_{est}(I_i^c, I_j^c)$  – вычисленное проективное преобразование между изображениями  $I_i^c$  и  $I_j^c$ ;
- $P^c = \{P_{est}(I_i^c, I_j^c) : (i, j) \in O\}$  – множество проективных преобразований, вычисленных между изображениями, индексы которых принадлежат  $O$ , и цвета пикселей принадлежат  $c$ ;
- $Q^c = \{Q_{ij} \equiv Q(P_{est}(I_i^c, I_j^c), P_{markup}(I_i, I_j)) : (i, j) \in O\}$  – множество оценок качества вычисленных проективных преобразований.

Использовалось следующее множество пар индексов

$$O_s = \{O : i = \{1, 1 + s, 1 + 2 \cdot s, \dots\}, i < n, j \in \{1, \dots, n\} \setminus i\},$$

где  $s$  – параметр шага. Для каждого такого множества пар  $O_s$  вычислялись  $P^c$  и  $Q^c$  для  $c = \{RGB, \alpha\beta\}$ .

В работе использовалась следующая функция оценки результата работы алгоритма  $Q_{ij}$ :

$$Q(P_{est}, P_{markup}) = \frac{1}{4} \sum_q |q - P_{markup}^{-1} \cdot P_{est} \cdot q|, \quad (4.10)$$

где  $q \in \{(1, 1), (1, -1), (-1, 1), (-1, -1)\}$ .

На основе полученных результатов затем строилась гистограмма распределения ошибки вычисления проективного преобразования на множестве  $O_s$  в интервалах  $[a_i, a_{i+1})$ , где  $a = \{0, 1, \dots, t, \infty\}$ ,  $t$  – пороговое значение. Далее, пары изображений  $(i, j)$  из  $O_s$  считались успешно сопоставленными, если вычисленная ошибка сопоставления  $Q_{ij}$  была меньше порогового значения  $t$ . Пороговое значение  $t$  может задаваться в зависимости от приложения, например, в случае сопоставления символьной информации, в качестве  $t$  выбирается доля характерного размера символа. Пусть  $O_s^c(t)$  множество таких пар индексов.

В качестве основной оценки результата работы алгоритма на данном множестве пар изображений  $O_s$  было выбрано отношение:

$$\psi^c(O_s, t) = \frac{|O_s^c(t)|}{|O_s|}. \quad (4.11)$$

Использовались следующие параметры алгоритма взвешенного поиска проективного оптического потока:

- Порог максимально допустимого значения яркости пикселя для всех каналов – 240;
- Параметры ядра смаза  $\sigma = 1$ , размер окна – 7;
- Отсортированный по убыванию список коэффициентов уменьшения масштаба исходных изображений  $S = [16.0, 8.0, 4.0]$ ;
- Количества итераций, которое должно быть выполнено на каждом уровне пирамиды – 50.

#### 4.2.5 Экспериментальные результаты

Результаты экспериментов на синтетическом и реальном наборах данных представлены на рисунках 4.7, 4.8 и 4.9, которые содержат гистограммы ошибок сопоставления, посчитанных по формуле 4.10. Слева находятся гистограммы для цветового пространства  $RGB$ , справа –  $\alpha\beta$ . По оси  $x$  отложена величина ошибки, по  $y$  – доля от всех элементов множества  $O_s$  с соответствующей величиной ошибки.

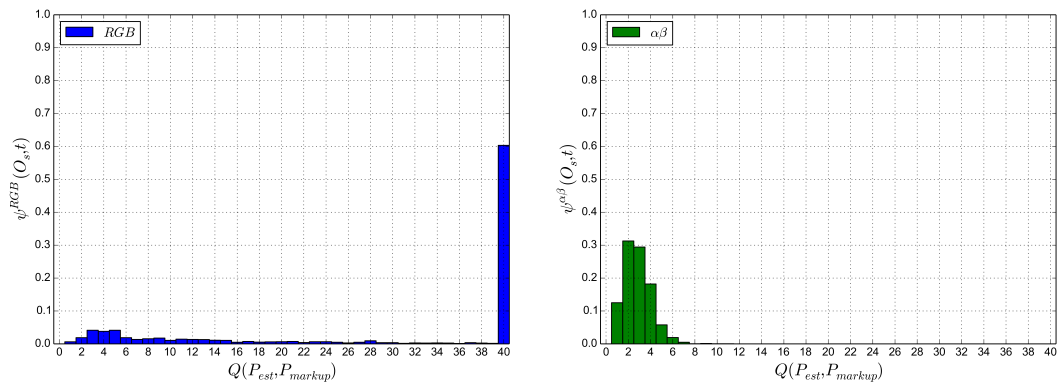


Рисунок 4.7 — Гистограммы ошибок сопоставления на синтетическом наборе данных (SD)

В таблице 2 приведены значения показателей качества  $\psi^{RGB}(O_s, t)$ ,  $\psi^{\alpha\beta}(O_s, t)$  для каждого набора данных.

Как видно из графиков 4.7, 4.8, 4.9, статистические характеристики показателя 4.11 алгоритма вычисления проективного оптического потока для данных



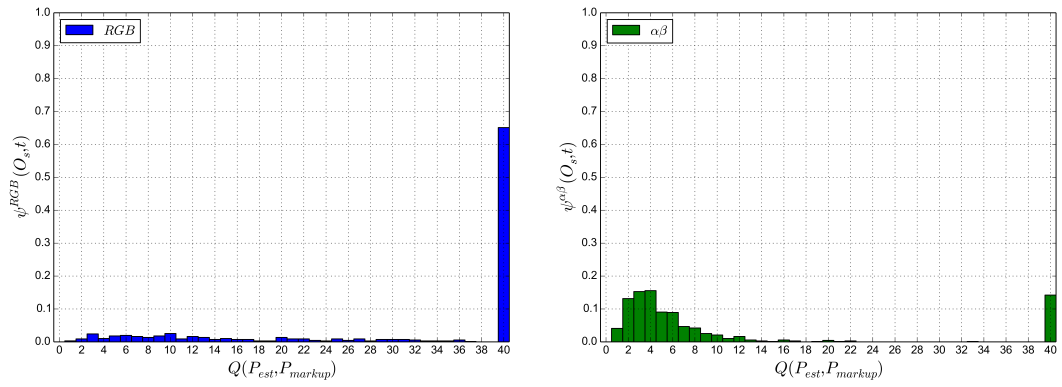


Рисунок 4.8 — Гистограммы ошибок сопоставления на наборе данных «Тройка» (TD)

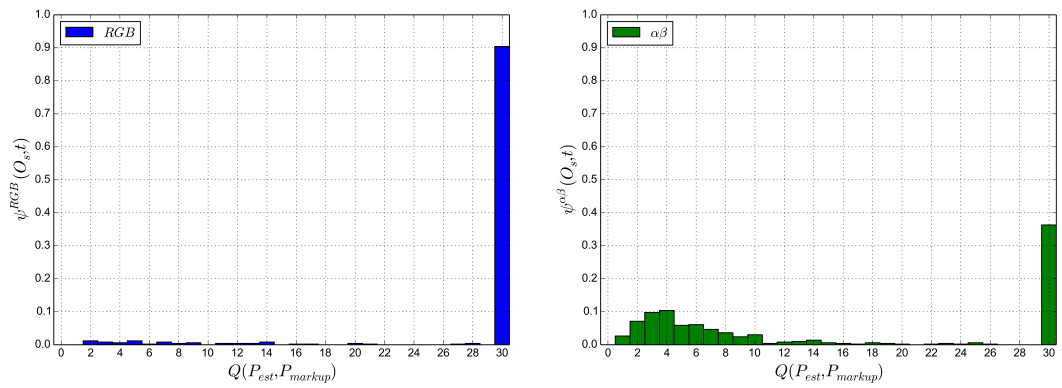


Рисунок 4.9 — Гистограммы ошибок сопоставления на наборе изображений с банковской картой (BD)

Таблица 2 — Показатели качества вычисления проективного оптического потока

Данные	SD	TD	BD
Порог допустимой ошибки $t$	40	40	30
Показатель качества $\psi^{RGB}(O_s, t)$	0.396	0.348	0.096
Показатель качества $\psi^{\alpha\beta}(O_s, t)$	1.000	0.858	0.637

SD, TD и BD выше в цветовом подпространстве  $\alpha\beta$  чем в  $RGB$  (в случае синтетический данных SD в  $\alpha\beta$  все пары изображений были успешно сопоставлены).

### 4.3 Детектирование и локализация периодических фоновых элементов на изображении документа

Бланки официальных личных документов часто содержат особые визуальные элементы, такие, как водяные знаки и гильёши, а поверх документа иногда наносятся голограммы и изображения на ретрорефлективном покрытии (рисунок 4.10). Эти элементы, как правило, уменьшают читаемость реквизитов документа, но обеспечивают дополнительный уровень защиты от подделок.



а) Водяные  
знаки на  
паспорте РФ



б) Голограммы на водительском  
удостоверении

Рисунок 4.10 — Примеры периодических защитных элементов

Детектирование, т.е. определение факта наличия таких элементов на изображении документа, делает возможным автоматическую проверку его подлинности и классификацию его типа [173], особенно актуальную в системах автоматического ввода документов.

В процессе извлечения текстовой информации из документов с помощью систем оптического распознавания символов (OCR), защитные элементы изменяют характеристики распознаваемых областей изображений, тем самым являясь шумом для OCR-систем. Как следствие, наличие фоновых элементов на текстовом заполнении потенциально снижает точность распознавания. Для увеличения точности распознавания возможно регулировать параметры OCR-систем в областях нанесения защитных элементов, а также задействовать методы подавления этих элементов с максимально возможным сохранением текстовой информации. Таким образом, для эффективной адаптации OCR-системам необходимо предоставить точное местонахождение защитных элементов с помощью алгоритмов их локализации.

Сузим класс рассматриваемых защитных элементов до периодических фоновых элементов, структура которых заранее известна. Например, при создании защитного слоя зачастую используется пленка с периодическими визуальными элементами фиксированной структуры, при этом разрез пленки проходит в произвольном месте, и смещение узора относительно бланка документа оказывается различным у разных экземпляров. Так, на паспорте гражданина РФ (рисунок 4.11) присутствует явный периодический голографический шаблон, загораживающий часть персональных данных, подлежащих вводу.



Рисунок 4.11 — Пример отсканированного паспорта гражданина РФ

Поставим задачу следующим образом: на имеющемся образе документа, полученном, например, с помощью сканирования, требуется определить факт наличия (детектировать) и, при наличии, местонахождение (локализовать) периодического фоновых шаблона известной структуры. Размеры документа можно считать фиксированными, границы документа должны совпадать с границами изображения. Структура потенциально присутствующего периодического шаблона заранее известна и характеризуется горизонтальным и вертикальным периодом элементов шаблона и размерами одного периодического элемента. Идеальные изображения периодических элементов могут быть недоступны, но имеется тестовая выборка, из которой могут быть получены примеры периодических элементов.

В разделе 1.3.4.2 был приведен обзор существующих подходов к поиску и удалению периодических шумов. Методы, принадлежащие первой части обзора, не могут быть эффективно использованы в детектировании периодических структур в сложных случаях, поскольку они не позволяют модифицировать

модель различий периодического шаблона от сложного фона, что требуется, например, для российского паспорта. Большинство других существующих методов являются универсальными и нацелены на значительно более широкий класс задач по сравнению с ранее поставленной задачей детектирования и локализации периодических фоновых элементов известной структуры на изображении документа. Нацеленность методов на функционирование в условиях излишней неопределенности приводит к увеличению ошибок, которых можно было бы избежать с помощью принятия во внимание всех имеющихся в задаче ограничений. При работе с удостоверяющими личность документами стоимость ошибки особенно велика.

Данный раздел посвящен разработке алгоритмов детектирования и локализации периодических фоновых элементов известной структуры на изображении документа, учитывающих всю специфику задачи и имеющуюся модель документа с целью увеличения точности и устойчивости относительно существующих алгоритмов.

#### 4.3.1 Модель сигнала изображения

Пусть  $I(x)$  – исходный сигнал изображения длины  $N$ . Он составлен из двух сигналов изображений: фонового изображения  $h(x)$  и изображения  $g(x)$ , содержащего периодический шаблон. Будем считать, что эти два изображения комбинируются аддитивным образом:

$$I(x) = h(x) + g(x). \quad (4.12)$$

Предположим, что число периодических элементов на  $g(x)$ , которое умещается в длину сигнала  $N$ , известно заранее и равно  $M$ . Пусть  $f(x)$  – сигнал единичного периодического элемента, имеющий период  $T = \frac{N}{M}$ . Тогда  $g(x)$  может быть представлен следующим образом:

$$g(x) = \sum_{m=0}^{M-1} f(x - mT). \quad (4.13)$$

Рассмотрим функцию  $c(x)$ , известную как гребень Дирака [137], которая представляет собой сумму единичных импульсов или дельта-функций  $\delta(x)$ , регулярно распределенных с интервалом  $T$  таким образом, что каждый единичный

импульс соответствует одному периодическому элементу  $f(x)$ :

$$c(x) = \sum_{m=0}^{M-1} \delta(x - mT). \quad (4.14)$$

Тогда  $g(x)$  – это свертка  $f(x)$  и  $c(x)$ :

$$g(x) = f(x) * c(x). \quad (4.15)$$

При такой модели задача детектирования формулируется, как оценка мощности  $f(x)$ , а локализации – как оценка фазы (циклического сдвига)  $c(x)$ .

#### 4.3.1.1 Преобразование Фурье сигнала периодического шаблона

По теореме о свертке, свертка в исходном пространстве эквивалентна поэлементному умножению в частотном пространстве. Тогда, при вычислении преобразования Фурье от  $g(x)$  и используя уравнение 4.15, имеем:

$$\mathcal{F}g(x) = \mathcal{F}f(x) \cdot \mathcal{F}c(x). \quad (4.16)$$

Рассмотрим подробнее структуру  $\mathcal{F}c(x)$ . Рассматривая дискретный случай преобразования Фурье, определенного (периодически) для  $k = 0, 1, \dots, N-1$ , Фурье-спектр единичной дельта-функции равен:

$$\mathcal{F}_k \delta(x - x_0) = e^{-2\pi i k x_0}. \quad (4.17)$$

Из-за линейности преобразования Фурье, спектр  $\mathcal{F}c(x)$  суммы дельта-функций равен:

$$\mathcal{F}_k c(x) = \sum_{m=0}^{M-1} \mathcal{F}_k \delta(x - mT) = \sum_{m=0}^{M-1} e^{-2\pi i k m T} = \sum_{m=0}^{M-1} e^{-2\pi i k \frac{m}{M} N}. \quad (4.18)$$

Легко увидеть, что для  $k = 0, M, 2M, \dots$  сумма 4.18 обращается в  $M$ , поскольку комплексная экспонента является  $2\pi i$ -периодичной и каждый член суммы обращается в 1. Более того, для всех других  $k$  сумма является нулевой, так как она содержит все комплексные экспоненты с частотами от 0 до  $m-1$ , которые в результате взаимоуничтожаются.

Это показывает важный факт: преобразование Фурье импульсного сигнала  $c(x)$  также является импульсным сигналом, но с периодом  $M$ :

$$\mathcal{F}_k c(x) = \sum_{m=0}^{M-1} e^{-2\pi i k \frac{m}{M}} = M \cdot \sum_m \delta(k - mM). \quad (4.19)$$

Например, на рисунке 4.12 приведено изображение амплитуды импульсного сигнала с  $T = 32$  и  $N = 256$ , который вписывается 8 раз, следовательно, его ДПФ имеет период, равный 8.

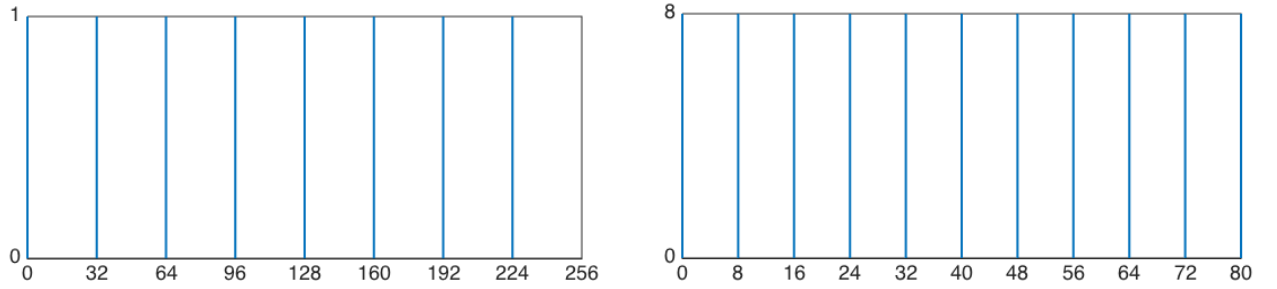


Рисунок 4.12 — Импульсный сигнал и его ДПФ амплитуда

Из уравнений 4.16, 4.19 получается окончательное выражение для  $\mathcal{F}g(x)$ :

$$\mathcal{F}_k g(x) = \begin{cases} M \cdot \mathcal{F}_k f(x), & \text{если } k = 0, M, 2M, \dots \\ 0, & \text{иначе.} \end{cases} \quad (4.20)$$

Последующий анализ ДПФ спектра основан на специфичной форме  $\mathcal{F}g(x)$  в результате данного свойства.

#### 4.3.1.2 Сдвиг сигнала периодического шаблона

Для упрощения модели ранее предполагалось, что периодический шаблон всегда начинается в начальный момент времени. Данное предположение в общем случае не выполняется, поэтому следует обобщить модель.

Сдвиг сигнала периодического шаблона  $g(x)$  равносителен сдвигу импульсной функции  $c(x)$ , который приводит к измененной форме  $\mathcal{F}c(x)$ . Спектр преобразования Фурье сдвинутого сигнала изменяется только в фазе, оставляя неизменной амплитуду, позволяя, тем самым детектировать периодические шаблоны вне независимости от их начального сдвига. Как важное следствие,



вся полезная фазовая информация сосредоточена в тех же самых ДПФ позициях (частотах), что и пики амплитуды импульсного сигнала.

Предположим, что  $c(x)$  сдвинут вправо на  $S$ , тогда, в соответствии с теоремой о ДПФ сдвинутого сигнала:

$$\begin{aligned}\mathcal{F}_k c(x - S) &= \mathcal{F}_k c(x) \cdot e^{i\Phi_k}, \\ \Phi_k &= \Phi \cdot k, \quad \Phi = -\frac{2\pi}{N}S.\end{aligned}\tag{4.21}$$

Поскольку фазовый угол  $\arg \mathcal{F}_k c(x)$  для несмещенного  $c(x)$  равен нулю, для смещенного  $c(x - S)$  он становится равным  $\Phi_k$ . Для получения фазового сдвига  $\arg \mathcal{F}_k g(x)$  нам необходимо добавить каждый член  $\Phi_k$  к соответствующему фазовому углу в  $\arg \mathcal{F}_k f(x)$  для каждого  $k$ . Заметим, что любой фазовый угол всегда берется по модулю  $2\pi$  и ограничивается интервалом  $[-\pi; \pi)$ .

Фазовый сдвиг на  $2\pi$  в частотной области представляет сдвиг  $N$  во временной области, но  $c(x)$  имеет период  $T = \frac{N}{M}$ , поэтому мы можем избежать избыточности рассмотрением лишь  $s = S \bmod T$ , обеспечив таким образом  $0 \leq s < T$ . Для удобства, пусть  $\varphi$  – это фазовый сдвиг, у которого  $2\pi$  соответствует временному сдвигу на  $s$ , а также пусть  $\varphi_m$  – это фазовый угол на  $m$ -ом амплитудном пике:

$$\varphi_m = \varphi \cdot m, \quad \varphi = -\frac{2\pi}{T}s = \Phi \cdot M.\tag{4.22}$$

Рисунок 4.13 показывает ДПФ фазу сдвинутого импульсного сигнала с рисунка 4.12 (а) со значениями  $N = 256$ ,  $T = 32$  и  $s = 7$ . Например,  $\varphi_0 = 0$  и  $\varphi_6 = (-\frac{2\pi}{32} \cdot 7) \cdot 6 \approx -1,963$ .

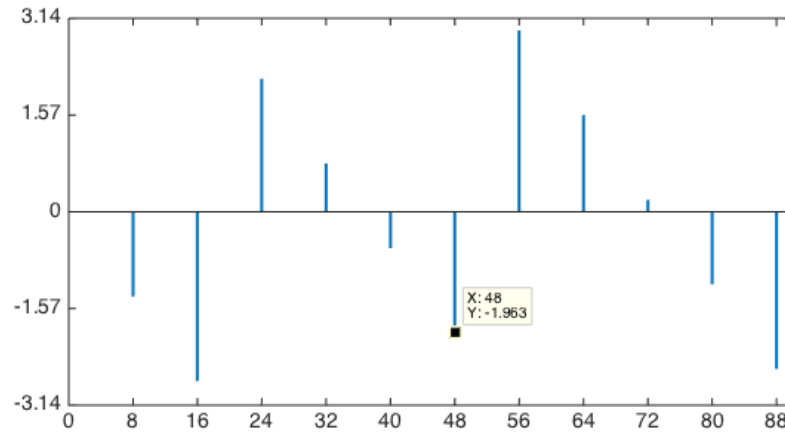


Рисунок 4.13 — Фаза ДПФ сдвинутого (на 7 вправо) импульсного сигнала

### 4.3.1.3 Модель сигнала двумерного изображения

На рисунке 4.14 представлена решетка регулярно распределенных Гауссиан (а) и сдвинутая амплитуда его ДПФ (б), увеличенная в центральной области. Аналогично одномерному случаю, здесь присутствуют различные импульсные сигналы с горизонтальным периодом 5 и вертикальным периодом 4, в соответствии с числом Гауссиан по горизонтали и вертикали.

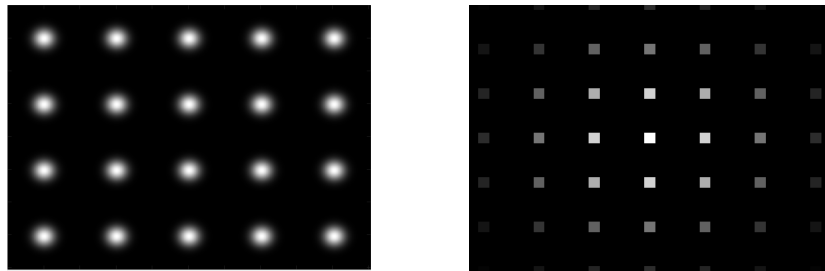


Рисунок 4.14 — Регулярно распределенные Гауссианы и их ДПФ амплитуды

Голографический периодический шаблон на паспорте РФ (рисунок 4.11) имеет несколько другую структуру. Рисунок 4.15 показывает решетку Гауссиан, подобную шахматной доске (а), ее ДПФ амплитуду (б) и фазу (в). Визуализация фазового спектра происходит по модулю  $2\pi$  и нормируется на единичный интервал. Отметим, что число Гауссиан по вертикали и горизонтали является целым и в данном случае равно в точности 2, что используется для детектирования и будет детально показано далее.

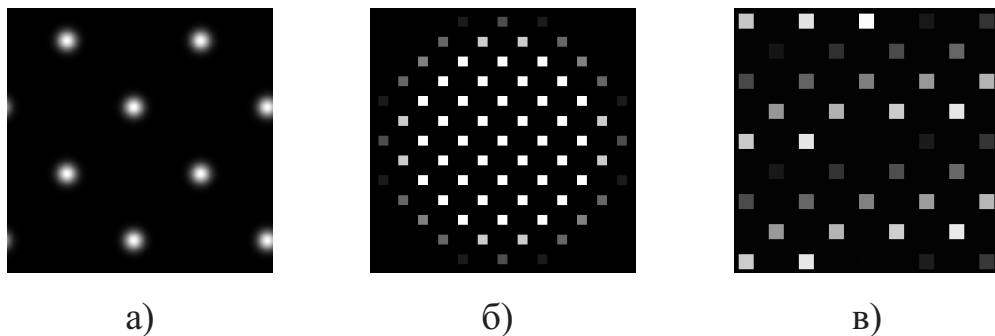


Рисунок 4.15 — Решетка Гауссиан, подобная шахматной доске (а), ее ДПФ амплитуда (б) и фаза (в)

Двумерный фазовый сдвиг  $\varphi$  теперь состоит из двух компонент  $(\varphi_x, \varphi_y)$ . Введем координатную сетку для ДПФ шахматоподобной системы пиков шаблона, показанных на рисунке 4.15 (б) и (в), такую, что центральный пик имеет

координаты  $(0, 0)$ , пики справа вверху  $(2, 0)$ ,  $(0, 2)$  и так далее; ближайший диагональный пик имеет координаты  $(1, 1)$ . В общем случае, пик  $(i, j)$  присутствует тогда и только тогда, когда  $(i + j)$  – чётно.

Применив вновь теорему о ДПФ сдвинутого сигнала, уравнение (4.22) может быть расширено и можно показать, что фазовый угол  $\varphi_{i,j}$  на  $(i, j)$  пике для двумерного импульсного сигнала равен:

$$\varphi_{i,j} = i \cdot \varphi_x + j \cdot \varphi_y. \quad (4.23)$$

Как и для одномерного случая, для получения сдвинутой фазы  $\arg \mathcal{F}_{i,j}g(x)$  необходимо добавить  $\varphi_{i,j}$  к соответствующей фазе  $\arg \mathcal{F}_{i,j}f(x)$ .

### 4.3.2 Детектирование периодического шаблона

Для детектирования (определения наличия) периодического шаблона требуется проверить, что амплитуда ДПФ сигнала исходного изображения  $|\mathcal{F}I(x)|$  содержит импульсный сигнал известного периода  $M$ . В двумерном случае вместо одного импульсного сигнала будет присутствовать определенный пиковый шаблон из импульсных сигналов. Пиковый шаблон ДПФ изображения российского паспорта будет соответствовать рисунку 4.15б. Кроме того, потребуется учесть амплитуду ДПФ фонового сигнала  $|\mathcal{F}h(x)|$ , вклад которого также присутствует в  $|\mathcal{F}I(x)|$ . Опишем основные этапы алгоритма детектирования периодического шаблона на изображении документа.

#### 4.3.2.1 Вырезание изображения

Первым этапом требуется выбрать область изображения, содержащую константное число элементов шаблона по вертикали и горизонтали, желательно, с минимальным количеством фонового шума. Это может быть сделано (с ошибкой менее 5%), поскольку физические размеры документа, размеры и период периодических элементов фиксированы и известны заранее. Используя методы поиска

границ документа [174], положение области может быть установлено комбинированием физических координат документа с ее координатами на изображении. На данной стадии предполагается, что границы документа нашлись с практически идеальной точностью. Рисунок 4.16 содержит пример области, вырезанной с изображения российского паспорта (рисунок 4.11) и содержащей 2 периодических элемента по вертикали и горизонтали (персональные данные были изменены).

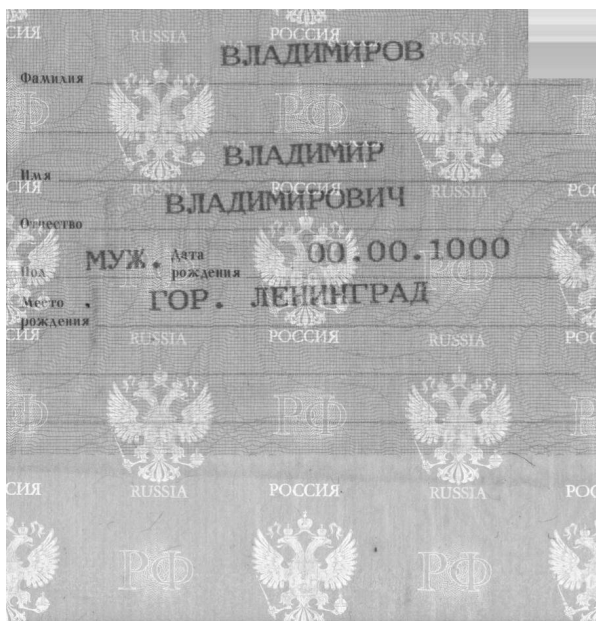


Рисунок 4.16 — Вырезанная область российского паспорта, содержащая 2x2 решетку периодического шаблона, подобную шахматной доске

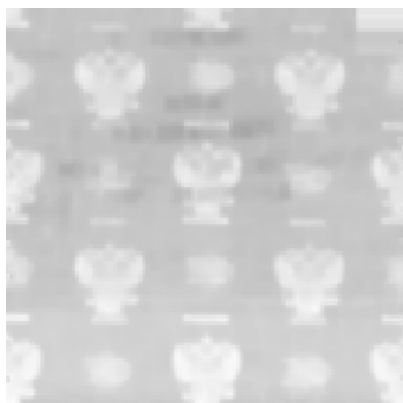
Точность вырезания области изображения крайне важна: нецелое число элементов периодического шаблона на вырезанной области может радикально изменить структуру ДПФ. По этой причине, в заключительной части главы приводится исследование устойчивости алгоритмов к ошибкам поиска границ документа.

#### 4.3.2.2 Масштабирование и предварительная обработка изображения

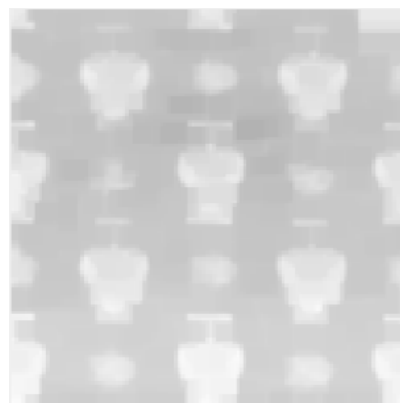
После вырезания интересующей области изображения, требуется подавить различия между экземплярами периодического шаблона, что может быть сделано путем масштабирования изображения к меньшему размеру. Наилучшие экспериментальные результаты были получены при масштабировании исходного

изображения размера 910x938 в 16 раз по каждой стороне, к размеру 56x58 (рисунок 4.17а). Итоговый размер после масштабирования в общем случае является константным и не зависит от исходного размера изображения.

Вдобавок к масштабированию используется операция морфологического закрытия [30] с целью дополнительной очистки изображения от остатков букв, а также для большего сглаживания отдельных элементов периодического шаблона.



а) Масштабирование



б) Морфологическое  
закрытие

Рисунок 4.17 — Предварительная обработка вырезанной области изображения

Кроме морфологического закрытия, проводились эксперименты и с другими операциями и их комбинациями для предварительной обработки масштабированных изображений, такими, как оператор Лапласиана Гауссиана или методом морфологической реконструкции [175]. В результате экспериментов было выяснено, что простое морфологическое закрытие способствует наилучшей точности детектирования.

#### 4.3.2.3 Маскирование и замещение областей изображения

Помимо зашумленных с точки зрения периодического шаблона областей, на неиспользованном регионе изображения все же присутствуют не подверженные искажениям периодические элементы. Они содержат полезную информацию для алгоритма, которая теряется при их игнорировании. Для того, чтобы включить в анализ интересные области документа без учета заведомо зашумленных, будем использовать маски интересных областей и замещение зашумленных.

Идея заключается в том, что для каждого заведомо искаженного региона, если его размеры не слишком велики, существует информативная область документа, которую можно скопировать в этот регион, не нарушая периодичности шаблонов. Таким образом, можно увеличить рассматриваемую область до периодической решетки  $3 \times 2$ , игнорируя вред от добавления зашумленных регионов. Описать разметку изображения можно с помощью маски, представляющей собой прямоугольную область, содержащую целое число периодов шаблона, внутри которой находится произвольное количество регионов с помехами для их последующего замещения.

Заметим, что область маски может заходить за пределы документа, что позволяет увеличить число периодов шаблона в случае большого дробного остатка периода на документе. Например, если бы на паспорте присутствовало 3.8 периодов шаблона по горизонтали, было бы оправдано периодически продолжить область документа на 0.2 периода – до 4 целых.

На рисунке 4.18 показана выбранная для паспорта РФ область маски и результат замощения шумовых регионов подходящими областями.



Рисунок 4.18 — Регион маски и результат замощения искаженных областей

Последующие этапы предобработки не отличаются от ранее введенных. На рисунке 4.19 приведен результат вырезания  $3 \times 2$  региона маски и его предобработки.

По координатным осям на амплитуде ДПФ видно, что для региона  $3 \times 2$  пики встречаются каждые 3 пикселя по горизонтали, а не 2, как для региона  $2 \times 2$ .



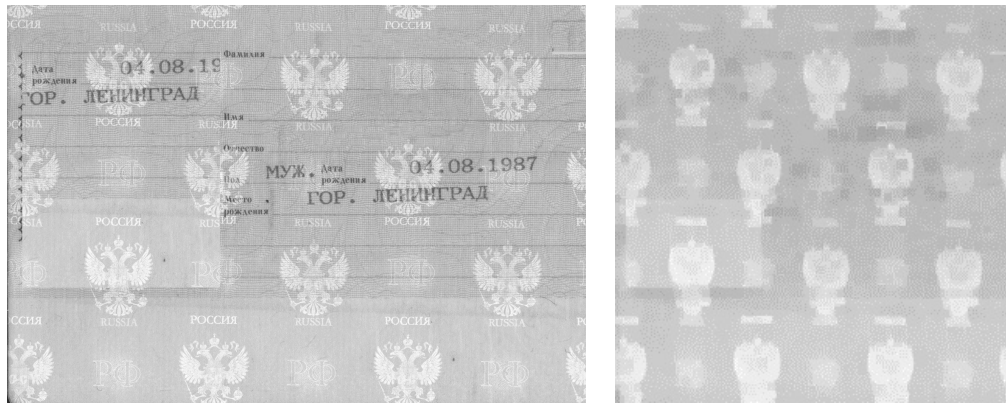


Рисунок 4.19 — Вырезанный и предобработанный 3x2 регион паспорта РФ

#### 4.3.2.4 Анализ спектра ДПФ

На рисунке 4.20 показан сдвинутый логарифмический график амплитуды ДПФ предобработанного изображения 2x2 области паспорта. Как и ожидалось, на нем присутствуют заметные периодические пики, которые должны быть сопоставлены с известной заранее структурой.

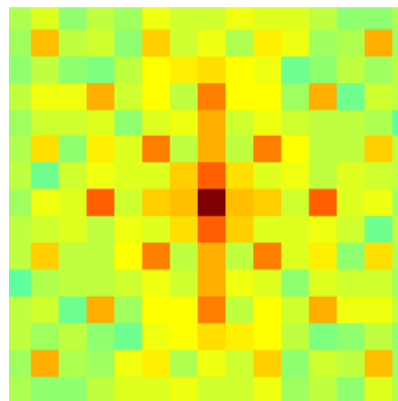


Рисунок 4.20 — Сдвинутый логарифмический график амплитуды ДПФ предобработанного изображения

Ранее была введена координатная сетка с точкой  $(0, 0)$  в центральной позиции сдвинутого двумерного спектра, демонстрация которой для случаев 2x2 и 3x2 изображена на рисунке 4.21.

Для сопоставления спектра с известной структурой пиков будем использовать следующий алгоритм. Рассмотрим первые  $K$  потенциальных позиций пиков, например, расположенных вдоль диагонального луча, выходящего из центра. В экспериментах для региона 2x2 использовалось значение  $K = 3$  и позиции  $(2; 2)$ ,  $(4; 4)$  и  $(6; 6)$ .

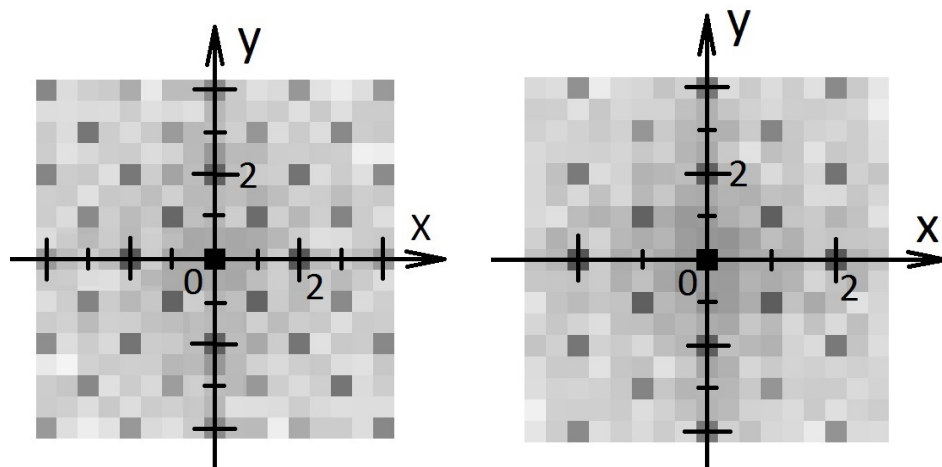


Рисунок 4.21 — Координатные сетки для амплитуды ДПФ вырезанных 2x2 и 3x2 регионов

Будем считать, что амплитуда ДПФ изображения, не содержащего периодическую структуру, имеет практически случайное распределение по сравнению со случаем присутствия периодической структуры. Тогда, высока вероятность того, что один из соседей для рассматриваемого потенциального пика будет иметь значение, большее, чем в рассматриваемом потенциальном пике. Поэтому, для каждой потенциальной позиции пика вычисляется минимальная разность  $D(x_i, y_i)$  между значением спектра в этом пике и значениями его ближайших 8 соседей. Итоговым показателем является среднее значение  $Q$  между всеми полученными значениями  $D(x, y)$ , которое сравнивается с пороговым значением  $Q^*$ . Если  $Q > Q^*$ , то принимается гипотеза о наличии периодического шаблона на изображении. Для российских паспортов использовалось значение  $Q^* = 0$ .

На рисунке 4.22 показан пример предобработанного изображения русского паспорта, не содержащего периодический шаблон, а также аналогичный график его амплитуды ДПФ. Видно, что амплитуда ДПФ не содержит пиковой структуры, как в случае наличия периодического шаблона, и потому описанный алгоритм возвращает решение об отсутствии шаблона.

**Оценка сложности алгоритма детектирования** Пусть  $N$  – количество пикселей области изображения после ее первоначального вырезания. При предварительной обработке происходит масштабирование (к заранее установленной константе  $N^*$ ) и операция морфологического закрытия; обе операции имеют сложность  $O(N)$ . Затем, вычисляется ДПФ обработанного изображения с помощью

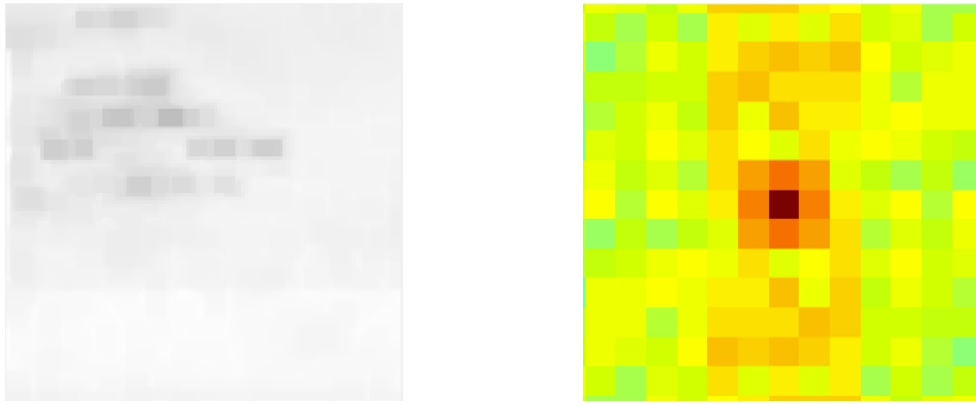


Рисунок 4.22 — Предобработанное изображение российского паспорта без периодического шаблона и его ДПФ

алгоритма быстрого преобразования Фурье, имеющего сложность  $O(N^* \log N^*)$ . Последним шагом является  $O(1)$  анализ спектра.

Итоговая вычислительная сложность довольно низкая, потому что БПФ производится на масштабированном изображении, имеющем малый константный размер, не зависящий от размера исходного изображения.

### 4.3.3 Локализация периодических шаблонов

В то время, как детектирование только определяет факт наличия периодических шаблонов, их локализация, в свою очередь, определяет их точную позицию, что может оказаться очень полезным для последующей обработки документа. Например, в зонах расположения шаблона можно динамически варьировать настройки алгоритмов распознавания для достижения наилучшего результата.

Для определения точного местоположения периодического шаблона достаточно оценить его фазовый сдвиг  $\varphi = (\varphi_x, \varphi_y)$ , поскольку его пиксельная периодическая структура заранее известна. Пиксельный сдвиг  $s = (s_x, s_y)$ , впоследствии, нетрудно восстановить по информации о фазовом сдвиге.

Опишем алгоритм поиска периодических шаблонов на изображении документа.

### 4.3.3.1 Предварительная обработка спектра

В соответствии с моделью изображения, заданной уравнениями (4.12) и (4.15), исходное изображение состоит из трех независимых сигналов: фонового изображения  $h(x)$  и единичного экземпляра периодического шаблона  $f(x)$ , который свертывается со сдвинутым импульсным сигналом  $c(x)$ , чтобы получить периодический шаблон  $g(x)$ :

$$I(x) = h(x) + f(x) * c(x). \quad (4.24)$$

После выполнения ДПФ на  $I(x)$ , имеем:

$$\mathcal{F}I(x) = \mathcal{F}h(x) + \mathcal{F}f(x) \cdot \mathcal{F}c(x). \quad (4.25)$$

Для получения фазового сдвига  $\varphi$ , информация о котором содержится в  $c(x)$ , из  $\mathcal{F}I(x)$ , вычисленного для данного изображения, необходимо подавить остальные компоненты уравнения:  $\mathcal{F}h(x)$  и  $\mathcal{F}f(x)$ .

**Подавление спектра экземпляров шаблона** Допустим, что спектр фона в уравнении 4.25  $\mathcal{F}h(x)$  был подавлен. Спектр изображения  $\mathcal{F}I(x)$  в таком случае стал равен  $\mathcal{F}f(x) \cdot \mathcal{F}c(x)$ . При перемножении двух комплексных чисел их фазы складываются. Тогда, для получения фазового угла из  $\arg \mathcal{F}c(i, j)$  в позиции пика  $(i, j)$ , соответствующий фазовый угол спектра единичного экземпляра периодического шаблона  $\arg \mathcal{F}f(i, j)$  должен быть вычтен.

Проблема состоит в том, что спектр периодического шаблона, как правило, неизвестен. В качестве первичного способа подавления спектра может быть использовано следующее решение: предположим, что фаза периодического экземпляра  $\arg \mathcal{F}f(i, j)$  везде равна нулю, а к результирующему сдвигу  $\varphi$  добавляется константный сдвиг, который может быть оценен экспериментально как систематическая ошибка на тестовом наборе данных и подавлен с помощью центрирования гистограммы ошибок.

Рассмотрим более детальную модель, в которой фазовый вклад  $\arg \mathcal{F}f(i, j)$  постоянен в рамках конкретного  $(i, j)$  пика. Данная модель согласуется с исходной моделью сигнала изображения  $f(i, j)$  для случая локализации, поскольку анализу подвергается только фазовая информация спектра. Тем не менее, для

данной модели требуется способ экспериментального вычисления фазового вклада  $\arg \mathcal{F}f(i, j)$ .

Пусть  $\theta_k(i, j)$  – идеальный фазовый сдвиг в пике  $(i, j)$  для  $k$ -го изображения из тестовой выборки, вычисленный по формуле 4.23 с использованием разметки сдвигов периодических шаблонов, а  $\varphi_k(i, j)$  – актуальное значение фазы в пике  $(i, j)$  для спектра  $k$ -го изображения. Аналогично, обозначим за  $\arg \mathcal{F}f_k(i, j)$  актуальное фазовое значение периодического элемента, которое может быть вычислено как разница между  $\theta_k(i, j)$  и  $\varphi_k(i, j)$ :

$$\arg \mathcal{F}f_k(i, j) = \theta_k(i, j) - \varphi_k(i, j) \pmod{2\pi}. \quad (4.26)$$

Введенная модель предполагает константность  $\arg \mathcal{F}f(i, j)$  в рамках пика  $(i, j)$  среди изображений, поэтому требуется способ оценки этих параметров по тестовым данным. В качестве оценки фазы периодического элемента для данного пика возьмем среднее значение для случая угловых величин [176], вычисляемое по формуле:

$$\arg \mathcal{F}f(i, j) = \arg \sum_k \exp(\arg \mathcal{F}f_k(i, j)). \quad (4.27)$$

Другими словами, для каждого угла  $\arg \mathcal{F}f_k(i, j)$  строится вектор на единичной окружности, затем все векторы складываются, после чего искомой оценкой является угол результирующего вектора. Рисунок 4.23 содержит пример гистограммы распределения  $\arg \mathcal{F}f_k(i, j)$  в рамках конкретного пика на тестовой выборке вместе с вычисленной оценкой фазы шаблона в этом пике.

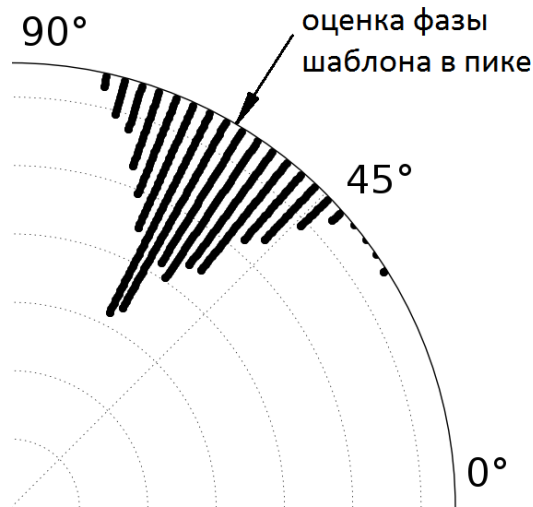


Рисунок 4.23 — Пример гистограммы распределения  $\arg \mathcal{F}f_k(i, j)$  на тестовой выборке для пика  $(i, j)$

Такой способ вычисления также можно трактовать как определение угла луча, проведенного из центра единичной окружности в направлении центра масс гистограммы.

Отметим, что при подсчете оценки фазы периодического элемента в пиках используются данные без предполагаемого подавления фона, о котором было сказано в начале. Эксперименты показали, что влияние спектра фона на распределение фаз в пиках для вычисления подобной оценки не является существенным.

**Подавление спектра фона** Если фон  $h(x)$  после предварительной обработки достаточно однороден на обрабатываемых документах, возможно простое вычисление усредненного спектра  $\mathcal{F}h(x)$  на изображениях документов, на которых отсутствует искомый периодический шаблон, с последующим его вычитанием из  $\mathcal{F}I(x)$ .

Однако, фоновое изображение  $h(x)$  в используемой модели фактически не является фоном в терминах структуры Российского паспорта: оно содержит персональные данные, которые по определению различаются на обрабатываемом наборе.

Ранее уже упоминался тот факт, что ДПФ позиции (частоты) без пиков не содержат полезной информации о периодическом шаблоне, поскольку они представляют фон. Предположив, что  $\mathcal{F}h(x)$  является гладким, возможно интерполировать его усредненный на тестовой выборке вклад в каждый пик  $\mathcal{F}I(i, j)$ , основываясь на значениях в соседних с пиком позициях. Усредненное значение  $\overline{\mathcal{F}I(x', y')}$  по  $(x, y)$  по ближайшим соседям пика  $\{\mathcal{F}I(x', y')\}$  в  $3 \times 3$  окне является базовой оценкой  $\mathcal{F}h(x, y)$  для вычитания из  $\mathcal{F}I(x, y)$  с целью подавления фона:

$$\mathcal{F}I(x, y) := \mathcal{F}I(x, y) - \overline{\mathcal{F}I(x', y')}. \quad (4.28)$$

Рисунок 4.24 иллюстрирует предобработанное изображение (а) и изображение, полученное в результате обратного ДПФ (б) после вычитания интерполированного спектра фона и обнуления ДПФ на всех непиковых частотах. Как и ожидалось, остался только периодический шаблон на простом монотонном фоне, а экземпляры периодического шаблона стали более похожими друг на друга. Отметим, что фон на рисунке не является черным, поскольку значение ДПФ на  $(0, 0)$ , которое содержит усредненное по изображению значение, не было обнулено.



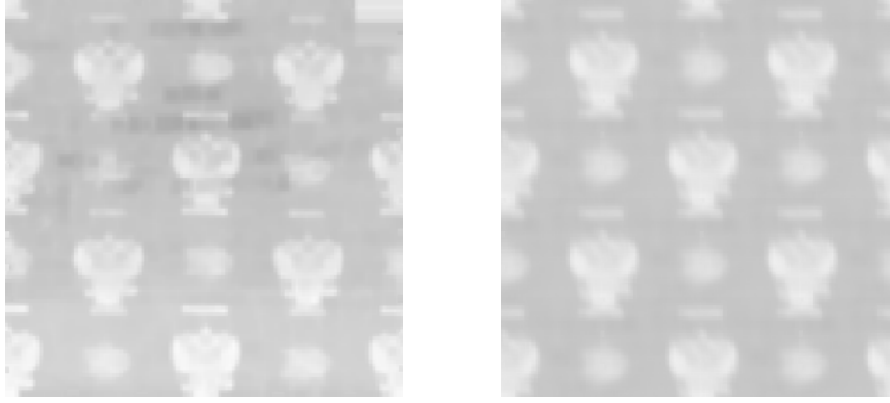


Рисунок 4.24 — Обработанное изображение паспорта и его обратное ДПФ после вычитания фона и сохранения только ДПФ пиков

Обобщим данный подход. Рассмотрим линейную модель, в которой значение спектра в пике  $(i, j)$  вычисляется, как линейная комбинация спектра самого пика и его непосредственных соседей, представляющих фон. Пусть  $\mathcal{N}(i, j)$  – множество позиций соседей пика  $(i, j)$ , включая сам пик:  $(i, j) \in \mathcal{N}(i, j)$ . Сопоставим каждому соседу  $\mathcal{N}_p(i, j)$  индекс  $p \in [0, 8]$ ,  $\mathcal{N}_0 = (i, j)$ . Пусть  $c_p$  – коэффициент перед  $p$ -м соседом пика в итоговой линейной комбинации, тогда:

$$\mathcal{FI}(i, j) := \sum_{p=0}^8 \mathcal{FI}(\mathcal{N}_p(i, j)) \cdot c_p. \quad (4.29)$$

Коэффициенты  $c_p$  вектора  $c$  для фиксированного пика  $(i, j)$  будем находить с помощью решения переопределенной системы уравнений при тестовой выборке размера  $K$ :

$$A_{2K \times 9} c_{9 \times 1} = b_{2K \times 1}. \quad (4.30)$$

Матрица  $A$  левой части системы содержит актуальные значения спектров множества соседей  $\mathcal{FI}_k(\mathcal{N}_p(i, j))$  на тестовой выборке для  $k \in [1, K]$ . Поскольку модель подразумевает действительные коэффициенты, т.е.  $c_p \in \mathbb{R}$ , а спектры соседей – комплексные, т.е.  $\mathcal{FI}_k(\mathcal{N}_p(i, j)) \in \mathbb{C}$ , разобьем каждое уравнение на два – по одному уравнению для действительной и мнимой части:

$$\begin{cases} A_{2k,p} = \text{Re } \mathcal{FI}_k(\mathcal{N}_p(i, j)) \\ A_{2k+1,p} = \text{Im } \mathcal{FI}_k(\mathcal{N}_p(i, j)) \end{cases} \quad (4.31)$$

Вектор  $b$  правой части системы содержит информацию о желаемых значениях комплексных аргументов в данном пике  $(i, j)$ . Каждый желаемый аргумент

состоит из суммы ранее введенного идеального фазового сдвига  $\theta_k(i, j)$  и оценки фазы периодического шаблона  $\arg \mathcal{F}f(i, j)$  для этого пика. Сопоставим каждому аргументу комплексное число единичной амплитуды и аналогично разобьем уравнение на два уравнения для действительной и мнимой части:

$$\begin{cases} b_{2k} = \cos(\theta_k(i, j) + \arg \mathcal{F}f(i, j)) \\ b_{2k+1} = \sin(\theta_k(i, j) + \arg \mathcal{F}f(i, j)) \end{cases} \quad (4.32)$$

Полученная переопределенная система уравнений решается методом наименьших квадратов.

#### 4.3.3.2 Восстановление фазового сдвига

После проведения предварительной обработки спектра в оставшемся сигнале  $\mathcal{F}c(x)$  содержится фазовая информация об итоговом фазовом сдвиге шаблона  $\varphi = (\varphi_x, \varphi_y)$ . Используя уравнение 4.23, построим систему уравнений по модулю  $2\pi$ , в которой каждое уравнение соответствует пику  $(i, j)$ :

$$\begin{cases} \dots \\ i \cdot \varphi_x + j \cdot \varphi_y = \varphi_{i,j} \pmod{2\pi} \\ \dots \end{cases} \quad (4.33)$$

Или, в матричной форме:

$$A\varphi = b \pmod{2\pi} \quad (4.34)$$

Правые части  $\varphi_{i,j}$  системы для каждого уравнения вычисляются на основании описанных способов предварительной обработки спектра: сначала происходит подавление фона, а затем из фазовой части спектра вычитаются оценки фазы периодического шаблона в пике.

Эта система переопределена, т.к. она имеет  $n$  уравнений (по числу рассматриваемых пиков) и только 2 переменные:  $\varphi_x$  и  $\varphi_y$ . Переопределенные системы уравнений могут быть приближенно решены методом наименьших квадратов (МНК) поиском решения, минимизирующего сумму квадратов невязок [177]. Однако, эта система является системой по модулю  $2\pi$  поэтому напрямую МНК

не может быть применен, но можно вычислить начальное решение и скорректировать его последовательным решением системы.

**Вычисление начального решения** Рассмотрим два уравнения для пиков под номерами  $(2, 0)$  и  $(0, 2)$ :

$$\begin{cases} 2 \cdot \varphi_x + 0 \cdot \varphi_y = \varphi_{2,0} \pmod{2\pi} \\ 0 \cdot \varphi_x + 2 \cdot \varphi_y = \varphi_{0,2} \pmod{2\pi} \end{cases} \quad (4.35)$$

Эта система имеет 4 допустимых решения:  $(\frac{\varphi_{2,0}}{2}, \frac{\varphi_{0,2}}{2})$ ,  $(\frac{\varphi_{2,0}}{2} + \pi, \frac{\varphi_{0,2}}{2})$ ,  $(\frac{\varphi_{2,0}}{2}, \frac{\varphi_{0,2}}{2} + \pi)$  и  $(\frac{\varphi_{2,0}}{2} + \pi, \frac{\varphi_{0,2}}{2} + \pi)$ . Из-за подобной шахматной доске пиковой структуры в данном случае, а также, поскольку  $2\pi$  представляет собой сдвиг на единичный период, первое с четвертым решения идентичны, как и второе с третьим. Остается два потенциальных решения:  $\varphi_1^* = (\frac{\varphi_{2,0}}{2}, \frac{\varphi_{0,2}}{2})^T$  и  $\varphi_2^* = (\frac{\varphi_{2,0}}{2} + \pi, \frac{\varphi_{0,2}}{2})^T$ .

Для выбора правильного решения добавим к системе оставшиеся уравнения для пиков, вновь сделав систему переопределенной. Тогда, значение невязки  $\delta_k$  вычисляется для  $k$ -го решения  $\varphi_k^*$  с учетом всех  $n$  выбранных уравнений, будет равно:

$$\delta_k = \frac{1}{n} \sum_{i,j} \left( \begin{pmatrix} i & j \end{pmatrix} \cdot \varphi_k^* - \varphi_{i,j} \right)^2 = \frac{1}{n} |A\varphi_k^* - b|^2 \quad (4.36)$$

Следует заметить, что промежуточные вычисления, затрагивающие фазовые углы (вычисление разностей), выполняются в  $[-\pi; \pi)$ .

Решение  $\varphi_k^*$ , имеющее наименьшее значение  $\delta_k$  выбирается в качестве  $\varphi^*$ , поскольку другое решение будет иметь намного большее значения невязки. Большое значение  $\delta_k$  может означать ложное детектирование периодического шаблона.

**Корректировка решения** Найденное выше начальное решение могло бы быть приемлемым, если бы оставшийся спектр содержал только информацию про  $s(x)$  и ничего более. Однако, он искажен спектральными составляющими от фона и от периодических элементов, которые не были полностью удалены во время предварительной обработки.

На данном этапе выбранное решение  $\varphi^*$  является все еще верным, но может быть уточнено с использованием фазовой информации в других пиках, а не только по двум первым. Вычтем  $A\varphi^*$  (используя  $n = 4$  уравнения) из

обеих частей системы уравнений, сдвинув таким образом пространство решений по  $\varphi^*$  и получив следующую переопределенную систему уравнений для корректирующего решения  $\theta$ :

$$A\theta = b - A\varphi^*. \quad (4.37)$$

Предположив, что текущая ошибка не превышает  $\pi$  по каждой искомой компоненте решения, можно не брать систему по модулю  $2\pi$ , т.к. после сдвига системы каждое уравнение будет внутри  $2\pi$ . Это означает, что возможно приближенно решить систему, используя метод наименьших квадратов. После того, как решение  $\theta^*$  найдено, оно должно быть добавлено к ранее найденному начальному решению  $\varphi^*$  в качестве корректирующего значения.

До сих пор рассматривались уравнения для первых 4 пиков  $(i, j)$  имеющих  $(|i| + |j|)$  сумму равную 2 (и  $i \geq 0$  из-за симметричности ДПФ), но ведь существуют другие пики с суммами, равными 4, 6 и так далее. Причина, почему они не были добавлены ранее, состоит в том, что при увеличении  $(|i| + |j|)$  увеличивается и количество подходящих решений системы (4.34), и тем ближе они становятся друг к другу по модулю  $2\pi$ .

Для использования информации о фазовом сдвиге, содержащейся в остальных пиках, следует повторить вышеописанный процесс корректировки начального решения (4.37), заменив  $\varphi^*$  скорректированным решением, полученным для суммы, равной 4 и используя все уравнения, имеющие сумму, равную 6 и так далее, до достижения требуемой точности. Аналогично, предположив, что после каждой итерации ошибка снижается настолько, что каждое уравнение будет находиться в пределах  $2\pi$ , возможно применять МНК для решения системы.

Более глубокий предварительный анализ спектра фона и фазы периодического шаблона позволяет использовать уравнения для всех имеющихся пиков, тогда как с базовыми вариантами оправдано использование только несколько ближайших к центру. В работе проводились эксперименты с двумя способами добавления пиков в систему: в порядке увеличения расстояния пика от центра и в порядке увеличения дисперсии оценки фазы периодического элемента в данном пике по формуле 4.27. Оба способа достигают одинаковой точности при добавлении всех пиков, но второй показывает более стабильные результаты на первых этапах добавления.

### 4.3.3.3 Восстановление расположения периодического шаблона

Теперь, имея фазовый сдвиг  $\varphi = (\varphi_x, \varphi_y)$ , необходимо преобразовать его обратно в пиксельный сдвиг  $s = (s_x, s_y)$ . Зная период и размер шаблона, можно восстановить положение каждого элемента периодической решетки. Размер и положение вырезанного региона, используемого в процессе обработки, также известны, благодаря чему можно также получить позиции элементов шаблона на исходном изображении документа.

Рисунок 4.25 представляет пример локализации голографического периодического шаблона на примере российского паспорта с рисунка 4.11.

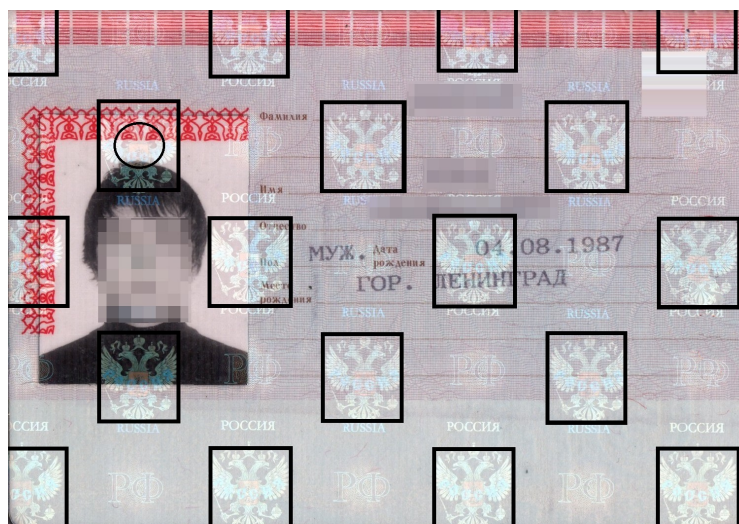


Рисунок 4.25 — Пример нахождения голографического периодического шаблона на российском паспорте

Рамкой с кругом внутри обозначен результат, найденный после решения системы уравнений. Другими рамками обозначены экстраполированные позиции остальных экземпляров шаблона.

### 4.3.3.4 Использование результатов локализации для детектирования

Ранее был описан алгоритм детектирования периодического шаблона на изображении, основанный на проверке наличия амплитудных пиков. В работе был

также опробован другой алгоритм, который показал лучшее разделение изображений тестовой выборки с периодическим шаблоном и без него (см. результаты).

Алгоритм основан на анализе ошибки решения переопределенной системы 4.33 с помощью метода наименьших квадратов. Если среднеквадратичная ошибка полученного в итоге решения относительно самой системы достаточно большая (выше фиксированного порога), значит периодический шаблон на исходном регионе изображения отсутствует.

Таким образом, если локализация проводилась после положительного результата детектирования предыдущим способом, но описанная ошибка системы оказалась выше порога, то можно говорить о ложно-положительном срабатывании алгоритма детектирования.

#### **4.3.4 Анализ устойчивости к ошибкам нахождения границ документа**

Исследование алгоритмов детектирования и локализации периодических шаблонов, проведенное в работе, подразумевает предварительное нахождение границ документа с достаточно высокой точностью. Для оценки потенциальных возможностей алгоритмов как при ошибках нахождения границ документа при его сканировании, так и при расширении алгоритма на изображения, полученные в неконтролируемых условиях съемки, проведем анализ деградации точности детектирования и локализации на искусственно сгенерированных выборках изображений, моделирующих ошибки нахождения границ документа.

##### **4.3.4.1 Моделирование ошибок нахождения границ документа**

Рассматриваются два типа искажений: ошибки в каждом из найденных углов документа и поворот документа на произвольный угол. Для каждого типа искажения независимо генерируется отдельная искусственная тестовая выборка с заданным распределением параметров ошибки, для которой алгоритмы настраиваются оптимальным образом. На рисунке 4.26 показаны примеры изображений,



подверженных искусственным искажениям, моделирующим ошибки нахождения углов (а) и поворота (б) документа.



а)

б)

Рисунок 4.26 — Пример искусственного моделирования ошибки нахождения углов (а) и поворота (б) документа

При моделировании ошибок нахождения углов документа для каждого угла изображения исходной тестовой выборки случайным образом генерируется смещение по вертикали и горизонтали. Смещение – нормально распределенная случайная величина с нулевым математическим ожиданием, параметром которой является заранее определенная  $\sigma^2$ . Для фиксированной  $\sigma^2$  и размеров изображения  $(W, H)$  актуальные дисперсии распределений горизонтальных и вертикальных смещений вычисляются, как:

$$\sigma_x^2 = \sigma^2 \cdot W, \quad \sigma_y^2 = \sigma^2 \cdot H. \quad (4.38)$$

Затем, смещенные позиции углов полагаются за найденные углы документа, после чего выполняется проективное преобразование изображения по этим углам к прямоугольному изображению исходного размера. Так как смещенные углы могут заходить за рамки изображения, предполагается, что каждое изображение помещено на белый фон.

Аналогичным образом, при моделировании поворота каждое изображение поворачивается на угол, являющийся нормально распределенной случайной величиной с нулевым математическим ожиданием, выраженной в градусах.

#### 4.3.4.2 Оценка точности локализации при моделировании ошибок

Идеальной разметкой локализации периодического шаблона для каждого из изображений документа исходной выборки является единственный горизонтальный и вертикальный сдвиг, из которого, зная размер и период элементов, периодический шаблон может быть однозначно экстраполирован. В обычных условиях оценкой точности локализации является расстояние между найденным и идеальным сдвигом, но для случая локализации на искаженных изображениях при моделировании ошибок нахождения границ документа требуется другой подход.

Пусть  $S$  – множество векторов сдвигов, найденных алгоритмом локализации с помощью простого периодического продолжения одного сдвига, а  $S^*$  – множество идеальных сдвигов для каждого элемента периодического шаблона с учетом проективного преобразования или поворота на угол для данного изображения. Тогда, ошибкой  $e$  на искаженном изображении является максимум среди расстояний от найденных сдвигов до соответствующих им ближайших идеальных сдвигов:

$$e = \max_{\vec{p} \in S} \min_{\vec{q} \in S^*} \|\vec{p} - \vec{q}\|. \quad (4.39)$$

При данном способе вычисления ошибка может быть ненулевой даже в случае полного соответствия найденного сдвига идеальному, так как текущий алгоритм локализации экстраполирует сдвиг, не обладая информацией об ошибках нахождения границ документа. По этой причине, в разделе экспериментальных результатов в качестве оценки теоретического предела точности алгоритма в его текущей реализации также приводится ошибка при экстраполяции идеального сдвига из разметки.

#### 4.3.5 Экспериментальные результаты

В экспериментах использовались 3 набора данных, каждый из которых состоял из отсканированных изображений российских паспортов. Первый из них, «scan500» – позитивный, т.е. на каждом изображении в нем присутствует периодический голографический шаблон. Другие два – негативные: «nomrz print» состоит из изображений паспортов без машиночитаемой зоны (МЧЗ) и основной

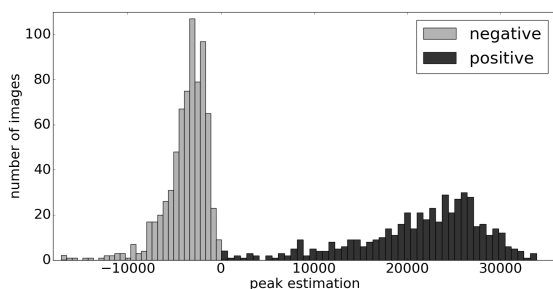
текст является обычным, т.е. печатным; «pomrz hw» содержит рукописный текст в полях документа. Суммарный размер позитивной выборки – 497 изображений, негативных – 714 изображений.

Каждое изображение позитивной выборки содержит вручную размеченный сдвиг периодического шаблона. Для тестирования точности локализации использовалась только позитивная выборка, а для детектирования – позитивная и негативная.

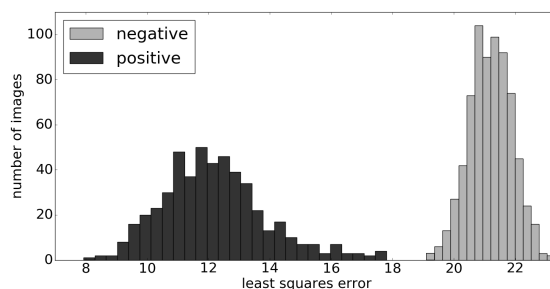
#### 4.3.5.1 Экспериментальные результаты алгоритмов детектирования

Была протестирована точность и разделяющая способность двух детекторов периодических шаблонов. Первый, базовый, детектор основан на проверке наличия амплитудных ДПФ пиков. Второй детектор проверяет ошибку при решении итоговой системы методом наименьших квадратов относительно самой системы. Оба детектора показывают абсолютную (100%) точность детектирования периодических шаблонов на позитивных и негативных выборках изображений при оптимальном подборе порога.

На рисунке 4.27 показано распределение множеств возвращаемых показателей разработанных алгоритмов детектирования: оценки соответствия амплитуды ДПФ пиковой структуре и ошибки решения системы при нахождении фазового сдвига методом наименьших квадратов соответственно.



а) Анализ амплитудных пиков



б) Анализ невязки решения системы

Рисунок 4.27 — Распределение множеств показателей алгоритмов детектирования на тестовых выборках

По графикам распределений видно, что второй способ детектирования имеет лучшую разделяющую способность, увеличивая зазор между позитивным и негативным множеством тестовой выборки.

В процессе предварительной обработки изображения сжимались до размера  $64 \times 64$ , после чего подвергались морфологическому закрытию. Фильтр Лапласиана Гауссиана обеспечивал наилучшее визуальный контраст голографических элементов, но его комбинирование с другими методами предварительной обработки ухудшали итоговую точность детектирования.

#### 4.3.5.2 Экспериментальные результаты алгоритмов локализации

Для каждого результирующего сдвига локализации вычисляется расстояние до позиции заранее размеченного идеального сдвига по горизонтали ( $x$ ) и вертикали ( $y$ ), выраженное в процентах от стороны окаймляющего прямоугольника периодического элемента. Итоговый показатель ошибки вычисляется, как корень из суммы квадратов этих расстояний.

Таблица 3 содержит экспериментальные результаты алгоритма локализации с разными параметрами, включающими подавление фона (вычетом линейной комбинации, простым вычетом среднего и без подавления), количество используемых пиков в системе, использование вычитания оценки фазы периодического элемента в пиках, размерности периодической решетки региона изображения и маскирование региона с его замещением.

Видно, что разработанные этапы увеличивают точность алгоритма. В процессе предварительной обработки изображения сжимались до размера  $64 \times 64$  и при решении итоговой системы использовалось 162 уравнения.

Были проведены дополнительные эксперименты, в которых изображения сжимались до  $256 \times 256$ , получая 2698 уравнений для пиков; ценой замедления алгоритма может быть достигнута точность 2.64. Эксперименты также показали, что добавление уравнений группами (например, по 10% от общего количества уравнений) ускоряет алгоритм без потерь точности.

Рисунок 4.28 содержит гистограмму распределения горизонтальных и вертикальных ошибок для лучшей комбинации параметров из таблицы 3.

Таблица 3 — Экспериментальные результаты алгоритма локализации

Вычитание фона	Кол-во пиков	Подавление спектра элементов шаблона	Размер решетки	Маскирова- ние и заливка	Итоговая ошибка
Лин. комб.	162	Да	3x2	Да	2.78
Лин. комб.	162	Да	2x2	Да	2.95
Лин. комб.	162	Да	2x2	Нет	3.02
Среднее	162	Да	2x2	Нет	3.36
Нет	162	Да	2x2	Нет	3.39
Лин. комб.	162	Да	3x2	Нет	3.62
Лин. комб.	4	Да	2x2	Нет	4.73
Лин. комб.	4	Нет	2x2	Нет	4.84
Среднее	4	Да	2x2	Нет	5.55
Среднее	4	Нет	2x2	Нет	5.72
Нет	4	Да	2x2	Нет	5.95
Нет	4	Нет	2x2	Нет	6.09

Для сравнения, на рисунке 4.29 показаны распределения ошибок для версий алгоритма локализации с ошибками 6.09 и 5.55 соответственно.

Видно, что распределение ошибок на рисунке 4.29а менее разбросано, чем на 4.29б, а распределение для наилучшего варианта алгоритма на рисунке 4.28 значительно сильнее сконцентрировано около нуля.

Рисунок 4.30 содержит график ошибки в зависимости от максимального количества добавленных в систему пиков при нахождении итогового фазового сдвига.

График ошибок на рисунке 4.30 подтверждает, что наилучшая точность локализации достигается при использовании уравнений для всех пиков в системе.

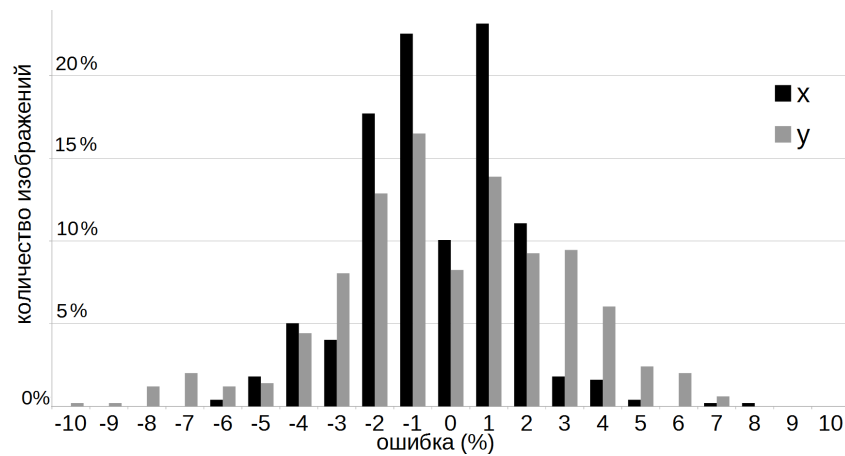


Рисунок 4.28 — Гистограмма распределений ошибки локализации для наилучшего варианта алгоритма

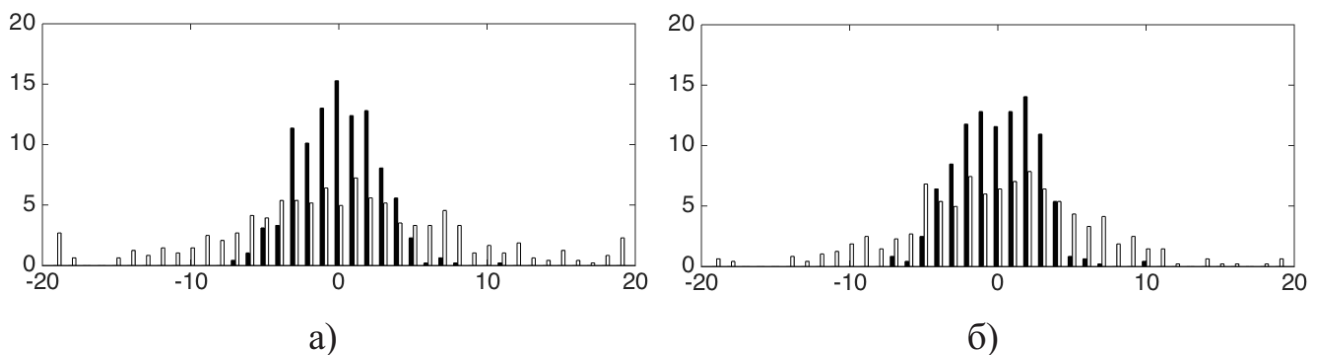


Рисунок 4.29 — Гистограмма распределений ошибок локализации для двух промежуточных вариантов алгоритма

#### 4.3.5.3 Результаты на искаженной выборке изображений

Рисунок 4.31 содержит графики деградации точности локализации в зависимости от дисперсии соответствующих распределений ошибок в нахождении углов документа (а) и углов поворота (б).

При серьезных ошибках в нахождении углов документа точность алгоритма (как разработанного, так и теоретически наилучшего) становится равносильна случайной, поскольку экстраполяция сдвигов не учитывает непараллельные осям координат смещения. Аналогичная ситуация наблюдается для поворотов документа на угол. Тем не менее, графики точности локализации идут вплотную с теоретическими максимумами при данном способе экстраполяции, что показывает потенциальную устойчивость алгоритма на этапах, предшествующих финальной экстраполяции.



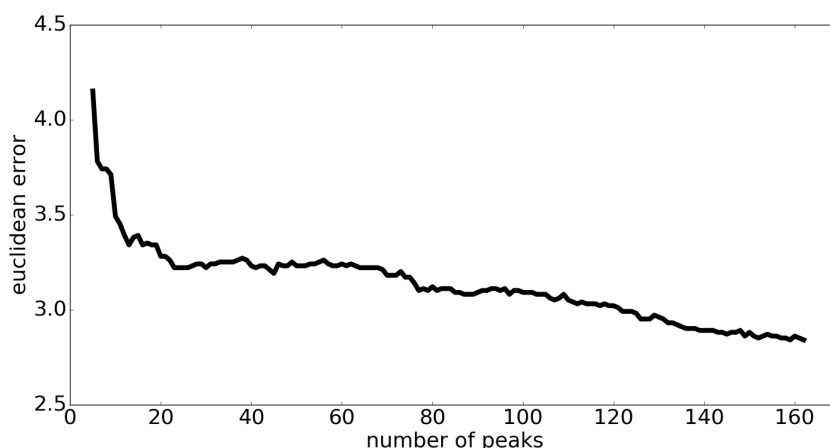
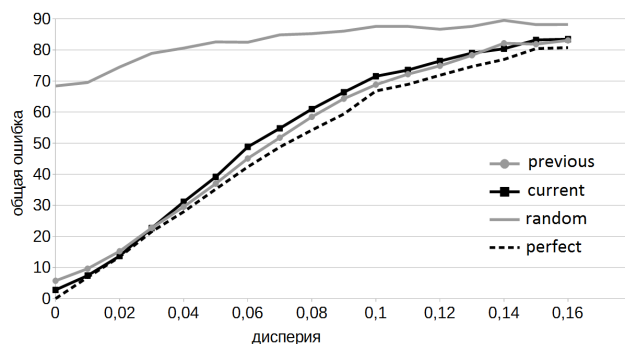
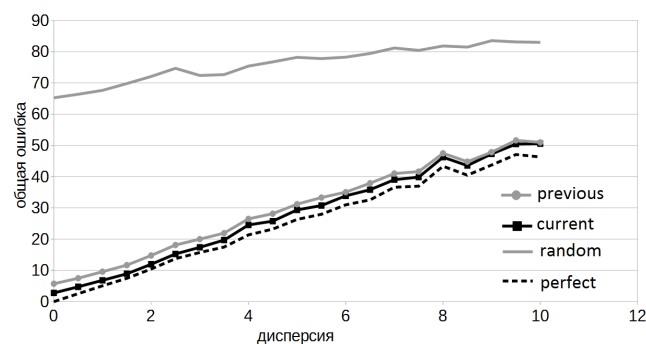


Рисунок 4.30 — График ошибки от количества пиков в системе уравнений при локализации



а) Моделирование ошибки нахождения углов



б) Моделирование ошибки угла поворота

Рисунок 4.31 — Точность локализации в зависимости от распределения при моделировании ошибок

На рисунке 4.32 показаны графики деградации точности алгоритмов детектирования при указанных искажениях для базового алгоритма, модифицированного алгоритма, случайного алгоритма и теоретически наилучшего алгоритма при используемой метрике точности.

Разделяющие позитивную и негативную выборку пороги для фиксированных параметров искажений выбирались независимо оптимальным с точки зрения точности способом. Разработанная модификация алгоритма детектирования с помощью анализа ошибки системы менее устойчива к серьезным ошибкам нахождения границ документа, чем базовый алгоритм анализа соответствия амплитуды ДПФ пиковой структуре. Таким образом, алгоритм детектирования успешно применим при небольших ошибках нахождения границ документа.

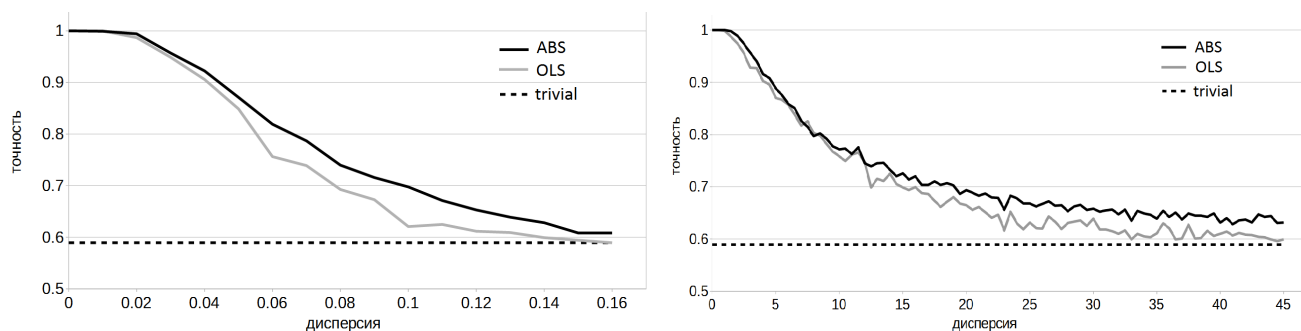


Рисунок 4.32 — Точность детектирования в зависимости от распределения ошибки нахождения углов и угла поворота

### 4.3.6 Общая схема алгоритма

На рисунке 4.33 приводится общая схема разработанного алгоритма.

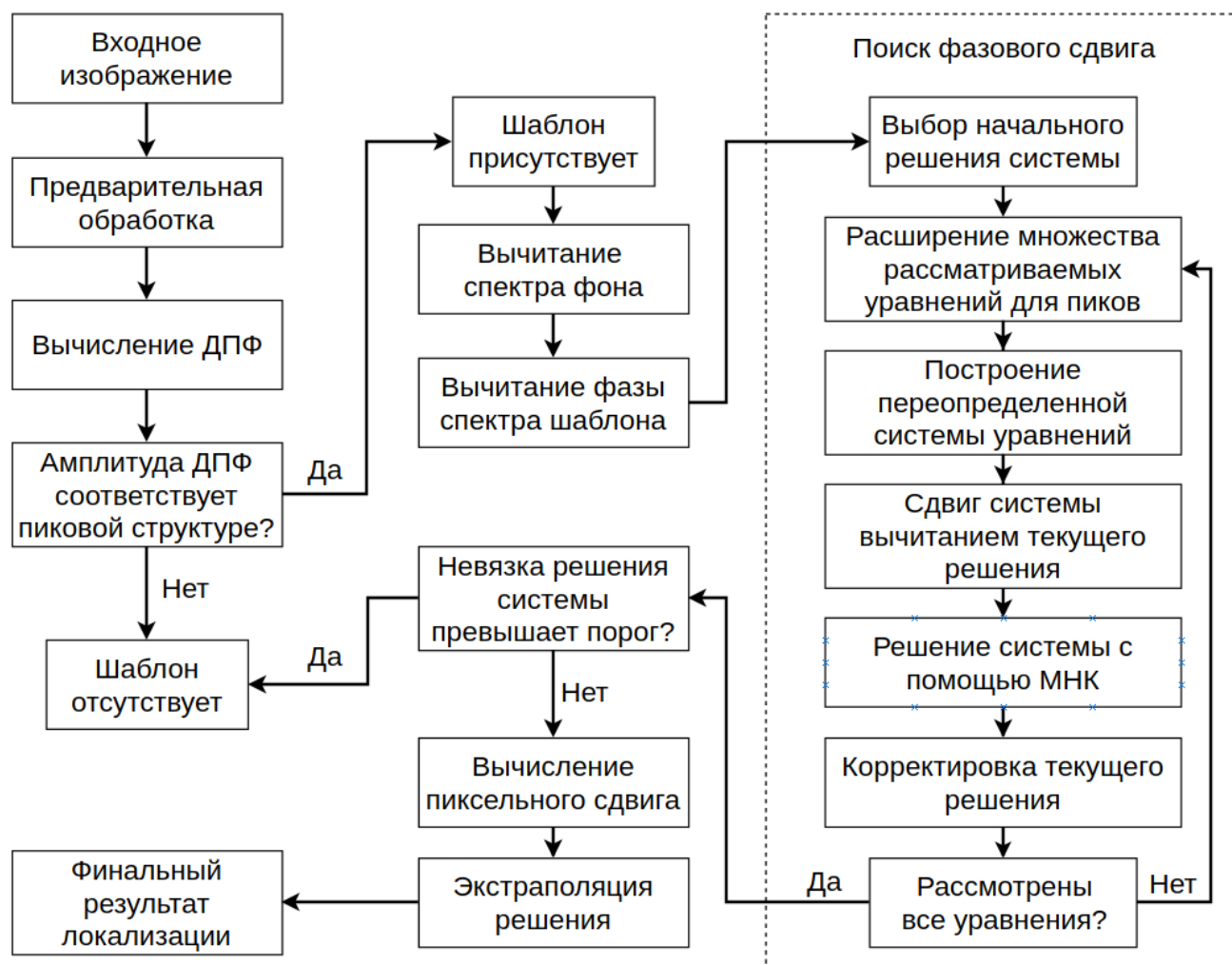


Рисунок 4.33 — Общая схема алгоритма детектирования и локализации

#### 4.4 Выводы по главе 4

В данной главе были рассмотрены приложения оценки качества изображений в системах распознавания при обеспечении контроля доступа и безопасности.

В разделе 4.1 был проведен анализ негативного влияния бликов на документах на работу системы поиска и распознавания документов. Разработаны алгоритмы детектирования и интерполяции поврежденных бликами областей, быстроедействие которых позволяет применять их в системах распознавания реального времени на мобильных устройствах. В экспериментах было показано, что фильтрация бликов благоприятно влияет как на устойчивость подсистемы поиска документа, так и на общую точность распознавания.

В разделе 4.2 был проведен анализ проблем при отслеживании движения плоских ригидных документов в видеопотоке на этапе проективного сопоставления кадров. Рассмотрен алгоритм взвешенного поиска проективного оптического потока, позволяющий учитывать извне заданную оценку достоверности значений пикселей изображений. Экспериментально показано, что для изображений сцены, удовлетворяющей линейной модели формирования цветных изображений, поиск проективного преобразования можно заметно улучшить путем сопоставления пикселей в цветовом подпространстве  $\alpha\beta$ . Полученные результаты подтверждают негативное влияние бликов на алгоритм поиска оптического потока. Кроме того, возможность сделать устойчивый к не зашкаливающим бликам оптический поток является дополнительным аргументом в пользу разработки и использования алгоритма детектирования зашкаливающих бликов, рассмотренного в разделе 4.1.

В разделе 4.3 исследовалась задача детектирования и локализации периодических фоновых элементов на изображении документа. Разработаны алгоритмы детектирования и локализации, работа и высокие точностные характеристики которых продемонстрирована для изображений паспорта гражданина Российской Федерации. Проведен экспериментальный анализ устойчивости алгоритмов к ошибкам нахождения границ документа, показавший, что алгоритмы устойчивы к небольшим ошибкам, но при серьезных ошибках их эффективное использование невозможно, поскольку нарушается основное для алгоритмов требование к периодичности фоновых элементов.

## **Глава 5. Контроль качества входного сигнала двухканального рентгенографического сепаратора**

Алмаз – это общеизвестный, редкий, но вместе с тем довольно широко распространенный минерал. С минералогической точки зрения алмаз является метастабильной аллотропной формой углерода, которая характеризуется кубической симметрией кристаллической структуры.

В процессе добычи алмазов руда, полученная из карьера, вывозится на обогатительную фабрику, где после дробления, сортировки сырья по плотности, размеру и фракциям происходит основная процедура – сепарация. Различают множество методов сепарации алмазов из руд: липкостная, магнитная, по степени шероховатости, гравитационная (тяжелосредная, ТС), рентгенолюминесцентная (РЛ).

### **5.1 Рентгенографическая сепарация**

Метод рентгенографической (РГ) сепарации [178] – перспективный метод, основанный на учете зависимости ослабления проходящего рентгеновского излучения от элементного состава исследуемого материала. В случае сепарации алмазосодержащей руды элементный состав фракций различается значительно. Алмаз – форма углерода (атомный номер  $Z = 6$ ), а поглощающие свойства алмазосодержащей породы в рентгеновском диапазоне в основном обеспечиваются элементами с  $Z$  в диапазоне 12–14. Наряду с магнитным и РЛ методами, РГ сепарация относится к неразрушающим методам, не требующим специальных рабочих сред. При этом в отличие от них РГ сепарация теоретически устойчива как к вариациям примесей в алмазах, так и к составу руды, поскольку основана на эффекте первого порядка относительно элементного состава.

### 5.1.1 Линейная модель рентгенографического изображения

В работе рассматривается следующая упрощенная модель работы рентгенографического сепаратора: на вход подается один объект произвольной формы и варьирующейся плотности, но постоянного элементного состава. Заранее известно, что объект принадлежит к одному из нескольких классов с фиксированным элементным составом. Требуется по интенсивности прошедшего через объект рентгеновского излучения классифицировать объект.

Будем считать рентгеновский пучок монохроматическим с энергией квантов  $E$ . Тогда в приближении бесконечно малого пикселя (элемента детектора) интенсивность  $I_d(E)$  излучения, прошедшего через объект и попавшего на элемент детектора, будет равна:

$$I_d(E) = I_0(E) \cdot \exp \left\{ - \int_0^L \mu_i(E) \cdot \rho(l) dl \right\}, \quad (5.1)$$

где  $I_0(E)$  – интенсивность излучения, падающего на объект;  $\mu_i(E)$  – массовый коэффициент ослабления излучения энергии  $E$  для материала элементного состава  $i$  единичной плотности;  $\rho(l)$  – плотность объекта в точке  $l$ ;  $L$  – толщина облучаемого слоя, содержащего объект. Здесь считается, что за пределами объекта  $\rho(l) = 0$ , поскольку поглощением в воздухе можно пренебречь.

Можно показать, что одноканальная рентгенографическая сепарация возможна только при известной эффективной толщине объектов [135]. В противном случае, сепарация алмазов и однородной породы может быть выполнена рентгенографическим сепаратором с числом каналов не менее двух при правильном подборе энергий рентгеновских источников. Далее будет рассматриваться двухканальный сепаратор. В предположении, что эффективные толщины алмаза и пустой породы не могут быть одновременно не равны нулю (на алмазах нет вкраплений породы), ожидаемое распределение наблюдаемых сигналов в пространстве интенсивностей  $(a_1, a_2)$  будет иметь характерный V-образный вид, что показано на рисунке 5.1.

Цифрой 1 помечен кластер, в который проецируются сигналы от кусков породы, цифрой 2 – от алмазов, а цифрой 3 – от пустого транспортера. Заметим, что ожидаются и физически бессмысленные отрицательные значения параметра поглощения как результат влияния различных шумов.

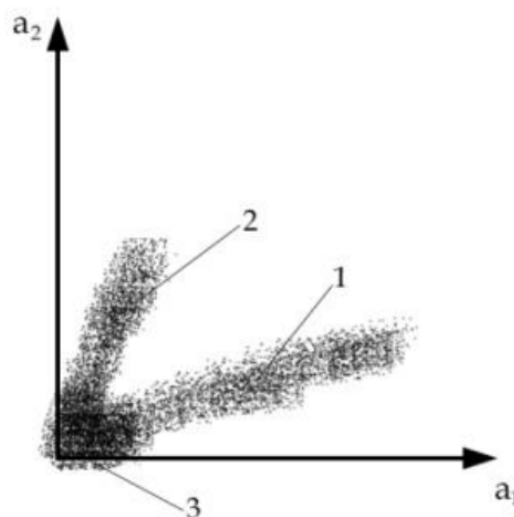


Рисунок 5.1 — Ожидаемое распределение входных сигналов РГ сепаратора

### 5.1.2 Устройство рентгенографического сепаратора

Рассмотрим теперь внутреннее устройство двухканального рентгенографического сепаратора (рисунок 5.2). Основой сепаратора служит конвейерный транспортер, на который дозированно подается раздробленная алмазосодержащая руда (4). Двигаясь вместе с конвейерной лентой, материал попадает в зону регистрации, где облучается сначала излучением рентгеновской трубки (РТ) с напряжением около 50 кВ («высокая энергия»), а затем излучением РТ с напряжением около 25 кВ («низкая энергия») (1), разделенных непрозрачным для рентгеновского излучения экраном (2).

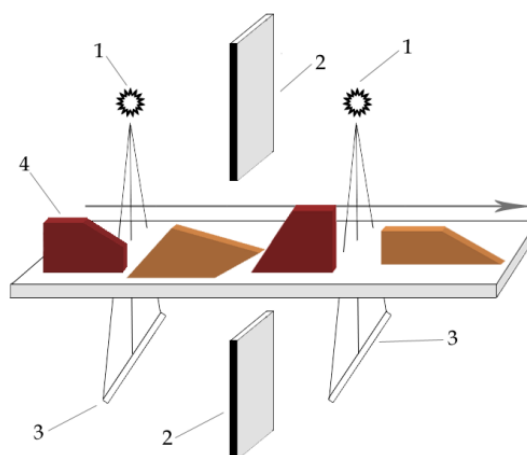


Рисунок 5.2 — Условная схема рентгенографического сепаратора



Под конвейерной лентой в зоне регистрации расположены две группы детекторов (3). Детектор представляет собой фотодиодную линейку с расположенным над ней слоем люминофора. При попадании на люминофор рентгеновского излучения возникает световой сигнал, пропорциональный интенсивности рентгеновского излучения. Фотодиодная ячейка преобразует принимаемый световой сигнал, поступающий на АЦП.

Руда, прошедшая зону регистрации, сбрасывается с транспортера в контейнер для пустой породы. В случае, если с вычислителя, исполняющего распознающий алгоритм, поступает сигнал о присутствии алмаза на ленте, в расчетный момент он отбрасывается импульсом сжатого воздуха в отдельный контейнер.

## 5.2 Контроль качества входного сигнала

Принципы рентгенографической сепарации, изложенные ранее, могут быть использованы в распознающем алгоритме только при условии калибровки излучения рентгеновских источников, чувствительности детекторов, оптической плотности ленты транспортера и скорости его движения. Однако даже в этом случае в процессе эксплуатации параметры могут постепенно «уходить» от калиброванных значений; кроме того, могут возникать другие нештатные ситуации, искажающие сигнал. Во всех таких случаях требуется либо остановка сепаратора, либо динамическая корректировка параметров, иначе качество сепарации может существенно упасть, что может привести к значительным убыткам. Следовательно, необходим способ контроля качества входного сигнала, позволяющий передавать сигнал о раскалибровке параметров или возникновении другой нештатной ситуации.

Поскольку многоканальный РГ-метод был предложен относительно недавно [179], непосредственно для него не удалось найти способов контроля качества сигнала. Патент [179] ссылается на одноканальный РГ-метод [180], где для первичного контроля сигнала предлагается использовать простое пороговое отсечение.

Для контроля сигнала в смежной области рентгенолюминесцентных (РЛ) методов [181] сепараторы оборудуются системами индикаторного контроля [182;

183]. Современный принцип их работы состоит в использовании автоматически подаваемых в руду специальных имитаторов алмаза, например, шариков из феррита бария, покрытых люминофором. При нормальном функционировании всех систем аппарата сепаратор произведет положительное срабатывание. В противном случае, от специального датчика, настроенного на регистрацию магнитных материалов, поступит специальный сигнал и сепаратор будет остановлен. Согласно [183], задача повышения достоверности контроля за счет увеличения соответствия физических характеристик имитатора и алмаза, а также снижения стоимости его изготовления для удовлетворения потребности алмазодобывающей промышленности, остается актуальной.

Важным требованием к методам контроля качества (как и к методам классификации) является их производительность по причине высокой скорости движения конвейерной ленты. По той же причине в качестве программных платформ для алмазных сепараторов выбирают операционные системы реального времени [184]. При низкой производительности программное обеспечение сепаратора может не успевать обрабатывать входной поток данных, что приведет к необходимости замедления конвейерной ленты и, следовательно, снижению пропускной способности сепаратора.

Основной проблемой, свойственной существующим решениям, является отсутствие диагностики нештатной ситуации: система лишь сообщает о поломке, но не уточняет, в какой именно ее части она произошла. Отсутствие диагностики ограничивает способность мониторинга системы, замедляет поиск неисправности, а также делает невозможным ее автоматическое исправление. К другим проблемам можно отнести дополнительные затраты на подготовку имитаторов алмаза, усложнение структуры сепаратора и то, что некоторые из способов контроля работают только в режиме предварительной калибровки.

Таким образом, для решения описанных проблем необходима разработка алгоритмов контроля качества сигнала, входными данными которых являлись бы только рентгенографические изображения, поступающие с сенсора, благодаря чему пропадает потребность в имитаторах алмаза и других внешних способах контроля. Алгоритмы должны быть способны работать в реальном времени в процессе непосредственной эксплуатации сепаратора, что накладывает дополнительные требования к их производительности, а также позволять производить диагностику возникших нештатных ситуаций. Заметим, что применение таких

алгоритмов возможно и вместе с использованием имитаторов алмаза – например, в целях диагностики.

Рассмотрим задачу контроля качества входного сигнала двухканального рентгенографического сепаратора для трех возможных некорректируемых нештатных ситуаций: раскалибровки пары источник/детектор, заслонения детектора и загрязнения конвейерной ленты. Задача состоит в разработке независимых алгоритмов детектирования соответствующих нештатных ситуаций, тем самым обеспечивающих их диагностику, удовлетворяющих сформулированным выше требованиям.

Далее будет произведен детальный анализ каждой из трех возможных ситуаций разладки, будут разработаны алгоритмы их детектирования, приведены экспериментальные результаты.

### 5.3 Динамический контроль чувствительности канала

Изменение чувствительности детектора (например, ее уменьшение за счет «выцветания» люминофора), а также интенсивности входного излучения (за счет изменения напряжения на трубке) может приводить к плавному изменению диапазона наблюдаемого сигнала. Пример такого эффекта показан на рисунке 5.3.

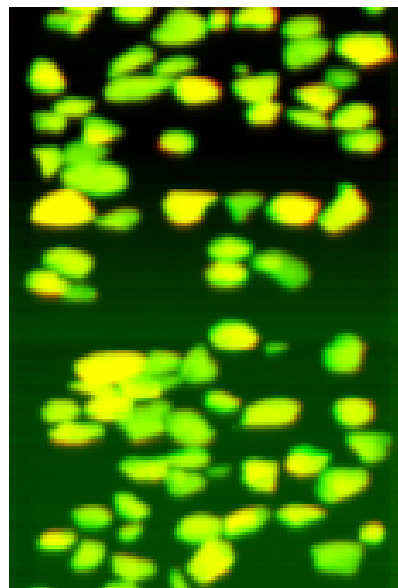


Рисунок 5.3 — Сигнал, полученный при изменении интенсивности излучения

Здесь (и далее на аналогичных иллюстрациях) ось времени направлена сверху вниз, а оба канала совмещены на одном цветном изображении. Яркость изображения визуализирует поглощение  $\vec{a}$ , то есть черный цвет соответствует пустому транспортеру, при этом зеленый цветовой канал визуализирует сигнал низкой, а красный – высокой энергии. На рисунке показан случай понижения напряжения на трубке низкой энергии, что приводит к видимому «позеленению» нижней части изображения.

Рассмотрим алгоритм, способный устойчиво детектировать подобные ситуации. Будем использовать тот факт, что заполнение ленты никогда не бывает близким к 100%, иначе зерна начнут накладываться друг на друга, затрудняя классификацию. Как правило, коэффициент заполнения не превышает 70%, причем расчетное значение известно заранее. Если возможно устойчиво оценивать текущую среднюю яркость сигнала в области пустой ленты, то задача контроля чувствительности канала сводится к пороговому детектированию отклонения этого параметра от начального значения.

Для оценки яркости фона будем поддерживать гистограмму  $H$  яркости канала в экспоненциально затухающем окне. Первым этапом обработки очередной строки изображения является обновление гистограммы, которое происходит следующим образом:

$$H_i^t = \alpha \cdot H_i^{t-1} + (1 - \alpha) \cdot h_i^t, \quad 0 < \alpha < 1, \quad (5.2)$$

где  $\alpha$  – коэффициент экспоненциального затухания, а  $h$  – гистограмма яркости новой строки.

Такой способ обновления называется экспоненциальным сглаживанием или методом экспоненциально взвешенного скользящего среднего (GMA) [185]. Отметим, что для подобного класса задач также существуют другие методы, такие, как метод кумулятивного накопления дрейфа (CUSUM), но нет оснований считать, что GMA менее эффективен; к тому же, он понятнее и легче поддается настройке [186].

В получившейся гистограмме найдем наименьший номер  $e$  ячейки гистограммы, для которой сумма долей значений (квантиль) с нулевой ячейки до данной не меньше, чем параметр  $q_e$ , задающий предполагаемую долю фона, не занятого объектами. Заметим, что  $q_e$  следует выбирать меньше расчетного заполнения, поскольку за счет неидеальной работы дозатора реальное заполнение может испытывать существенные колебания. В этом случае  $e$  будет несколько

смещенной, но зато устойчивой оценкой яркости фона. Для уменьшения числа ложных срабатываний вводится значение накопленного штрафа  $p$  (в начале работы системы равное нулю), которое наращивается в сторону единицы в случае, если отклонение  $e$  от исходного значения превышает порог. Пересчет накопленного штрафа производится по формуле:

$$p_t = \beta \cdot p_{t-1} + (1 - \beta) \cdot f, \quad 0 < \beta < 1, \quad (5.3)$$

где параметр  $\beta$  задает требуемую протяженность вертикальной серии превышающих порог пикселей, а  $f$  принимает значение единицы при превышении отклонения и нуля в противном случае. Сигнал о раскалибровке соответствующего канала передается при превышении значением накопленного штрафа порогового значения.

На рисунке 5.4 представлены графики зависимости доли черных пикселей  $q(t)$  и накопленного штрафа  $p(t)$  одного из каналов от времени с моментом раскалибровки данного канала  $t_{\text{start}} = 7000$ . Пунктиром показан порог накопленного штрафа, равный 0,75, при превышении которого происходит передача сигнала о раскалибровке. Видно, что процесс накопления штрафа  $p(t)$  начался практически сразу же после раскалибровки, сопровождаемой падением  $q(t)$  в ноль, а передача сигнала произошла спустя  $\Delta t \approx 500$ .

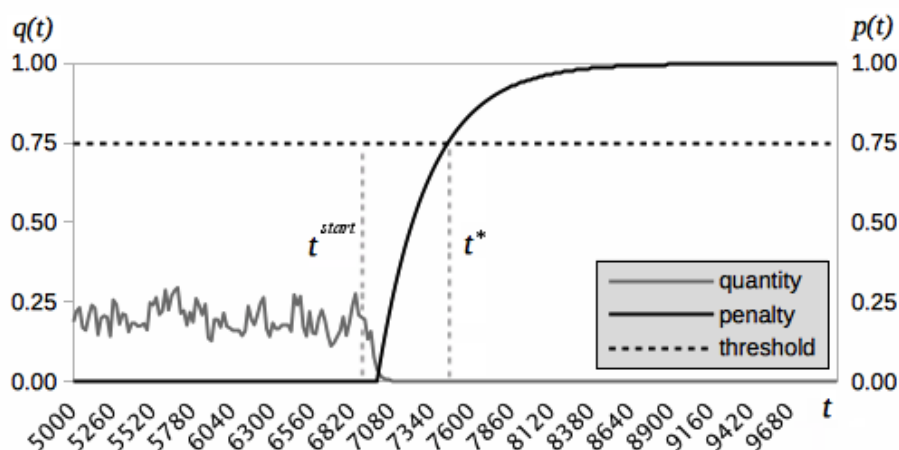


Рисунок 5.4 — Динамика изменения доли черных пикселей и штрафа при раскалибровке

Алгоритм был протестирован на 100 тестовых пакетах данных, содержащих записанные в виде двумерных массивов сигналы каналов «высокая» и «низкая» энергия рентгенографического сепаратора (каждый массив имеет ширину 256 пикселей и длину 60 000 строк). Алгоритм обнаружил изменение интенсивности

сигнала на 7 пакетах из 8 и не дал ни одного ложноположительного срабатывания на остальных 92 пакетах. Следует заметить, что на пакете с ложноотрицательным результатом изменение интенсивности оказалось подпороговым и для алгоритма классификации, что делает этот случай спорным.

#### 5.4 Обнаружение заслонения детектора

Из-за конструкционных особенностей сепаратора теоретически возможно попадание зерен породы непосредственно на детектор, где они будут создавать постоянную мультипликативную помеху в нескольких ячейках детектирующей линейки. Пример изображения, приходящего с детектора в таком случае, представлен на рисунке 5.5. Видно, что помеха присутствует в обоих каналах, с некоторым смещением друг относительно друга.

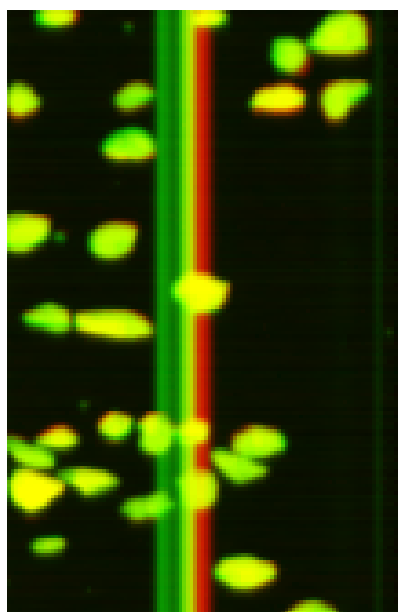


Рисунок 5.5 — Пример заслонения зерном породы

Для детектирования заслонения детектора требуется найти продолжительный по времени («вертикальный») участок, яркость сигнала на котором выше порогового значения. Отсечение ложных срабатываний на длинных участках подряд лежащих камней можно производить за счет ограничения на максимальную длину таких участков, но возникает проблема больших затрат памяти на хранение предыдущих строк сигнала.



С целью уменьшения использования памяти для каждой позиции  $i$  пикселя строки сигнала будем независимо поддерживать накопленный штраф, при превышении порогового значения которого (хотя бы в одной позиции) передается сигнал о заслоняющем детектор объекте. Одним из способов задания подобной функции штрафа является модификация экспоненциального сглаживания сигнала: штраф будет постепенно расти вверх к фиксированному значению, но не сможет превысить текущее значение пикселя сигнала. Обновление и анализ штрафа для каждого пикселя при обработке новой строки происходит следующим образом.

Первым шагом, штраф ограничивается яркостью пикселя, что в случае наличия блокирующего объекта не опустит накопленный штраф ниже порогового значения, а в случае отсутствия предотвратит ложное срабатывание:

$$p_i^{t-1} = \min(p_i^{t-1}, s_i^t). \quad (5.4)$$

Затем, штраф наращивается в сторону заранее известного значения яркости темнового тока  $d_i$  в соответствии с формулой:

$$p_i^t = \alpha \cdot p_i^{t-1} + (1 - \alpha) \cdot d_i, \quad 0 < \alpha < 1. \quad (5.5)$$

Значение  $\alpha$  задает требуемую протяженность вертикальной серии превышающих порог пикселей; слишком маленькое значение чревато ложными срабатываниями, а слишком большое – длительным ожиданием сигнала о заслоняющем объекте.

Наконец, обновленное значение штрафа сравнивается с пороговым значением. Заметим, что пороговое значение удобно задавать отклонением от заранее известного значения фона.

На рисунке 5.6 представлены графики зависимости накопленных штрафов от времени для заслоненной с начала работы системы при  $t = 0$  (blocked, черный) и не заслоненной (normal, серый) позиций пикселей. Пунктиром показано пороговое значение штрафа. Из-за очень близкого к единице значения сигнала о заслоняющем объекте был передан спустя  $\Delta t \approx 7500$ . Видно, что минимизация по значению сигнала не позволяет штрафу для пикселей, не принадлежащим заслонению, приблизиться к пороговому значению.

Алгоритм был протестирован на 92 тестовых пакетах, один из которых содержал сигнал с заслоненного детектора; все пакеты были обработаны без ошибок. При тестировании были исключены пакеты, содержащие сигналы, полученные при изменяющейся интенсивности излучения.

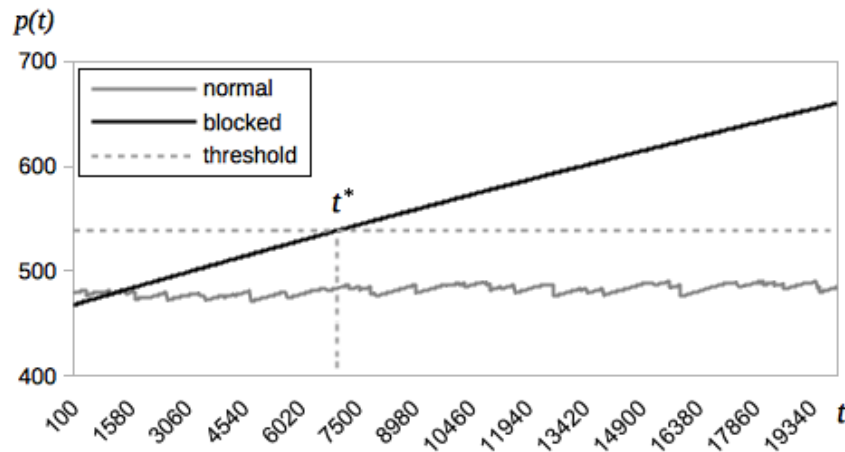


Рисунок 5.6 — Накопленные штрафы детектора заслонения

### 5.5 Обнаружение оптического загрязнения ленты транспортера

Еще одним неприятным эффектом является постепенное загрязнение ленты транспортера пылью и чешуйками породы, налипающими под действием статического электричества. Подобная помеха может существенно сместить оптимальные пороги детекции в соответствующей области, при этом она имеет нерегулярную микроструктуру, затрудняющую детектирование (рисунок 5.7).

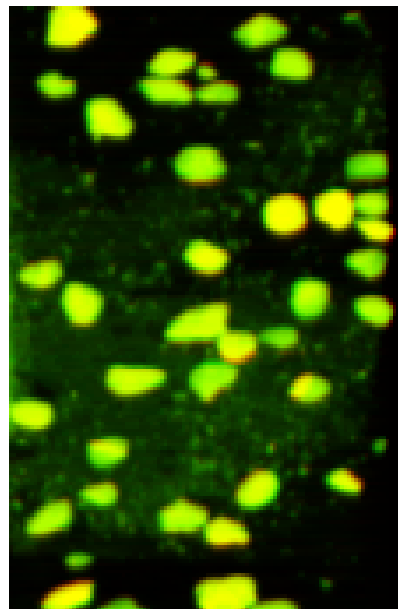


Рисунок 5.7 — Пример оптического загрязнения ленты

Поскольку данный вид помехи связан с лентой транспортера, помеха должна повторяться на каждом обороте ленты, что можно использовать при известных длине ленты и скорости движения транспортера. Для локальной детекции помехи

будем использовать тот факт, что поглощение детектируемых чешуек находится в диапазоне между шумами пустого фона и характерным поглощением зерен породы, и практически не детектируется в канале высокой энергии, обладающей большей пенетрантностью.

Таким образом, пиксель  $s_i$  входного сигнала, представленный парой компонент низкого и высокого каналов  $(l_i, h_i)$ , потенциально содержит «пыль», если значения компонент лежат в некотором заданном прямоугольнике, т.е. одновременно выполняется  $L_{min}^d \leq l_i \leq L_{max}^d$  и  $H_{min}^d \leq h_i \leq H_{max}^d$ . Однако, в результате неизбежных артефактов оцифровки сигнала, данное правило выполняется и на границах «камней», плавно переходящих в фон. Требуется отфильтровать такие случаи.

Пусть пиксель содержит «камень», если его компоненты лежат в другом заданном прямоугольнике:  $L_{min}^s \leq l_i \leq L_{max}^s$  и  $H_{min}^s \leq h_i \leq H_{max}^s$ . На основании этой информации можно отсечь ранее найденные пиксели «пыли», находящиеся на маленьком расстоянии от пикселей «камней». В частности, если представить вычисленные правила принадлежности в виде битовых масок (массивов), то близкие пиксели можно найти расширением соответствующей маски, реализуемым морфологической операцией наращивания (дилатации) [30]. Итоговая маска принадлежности к «пыли» получается в результате поэлементной конъюнкции с отрицанием маски «камней» после дилатации.

Аналогично детектору заслоняющих объектов, для каждого индекса  $i$  пикселя строки  $s_i$  удобно поддерживать свой штраф  $p_i$ . Штраф накапливается с помощью экспоненциального сглаживания ступенчатой функции  $f_i$  классификации пикселей как «пыли» (после отсечения пикселей, близких к «камням»):

$$p_i^t = \alpha \cdot p_i^{t-1} + (1 - \alpha) \cdot f_i, \quad 0 < \alpha < 1, \quad (5.6)$$

$$f_i^t = \begin{cases} 1, & s_i^t - \text{«пыль»}, \\ 0, & \text{иначе.} \end{cases} \quad (5.7)$$

Значение  $\alpha$  регулирует требуемую продолжительность вертикальной полосы «пыли» и в данном случае не должно быть очень большим.

Оптическое загрязнение обычно занимает обширную часть линейки, поэтому вместо независимого порогового сравнения штрафов в каждой позиции сравнение с порогом будет происходить для среднего значения накопленного штрафа.

С целью сокращения ложных срабатываний важно учитывать периодичность появления загрязнений на ленте транспортера. Введем понятие внутреннего события превышения порога штрафа  $e_t$ , которое кодируется временем его регистрации (положением на ленте сепаратора, т.к. лента движется равномерно с постоянным периодом обращения). Также нам потребуется список  $Q$ , в котором хранятся события для разных участков ленты и их число для каждого участка в виде пар  $(t', n_t)$ , где  $t'$  – время первой регистрации события на данном участке.

После генерации нового события превышения порога в списке ищется запись, соответствующая данному участку ленты (в окрестности шириной 10% от ее длины). Очевидно, что положение на участке ленты вычисляется, как  $t' = t \bmod B$ , где  $B$  – длина ленты транспортера в пикселях после оцифровки. Если такая запись находится, то число событий для данного участка увеличивается на единицу, если нет – то добавляется новая запись. Если число событий для данного участка ленты стало кратным заданному порогу, то генерируется сигнал о наличии оптического загрязнения.

На рисунке 5.8 представлен график зависимости накопленных штрафов от времени для случая отсутствия (а) и присутствия (б) оптического загрязнения на ленте транспортера. Пунктиром показано пороговое значение штрафа.

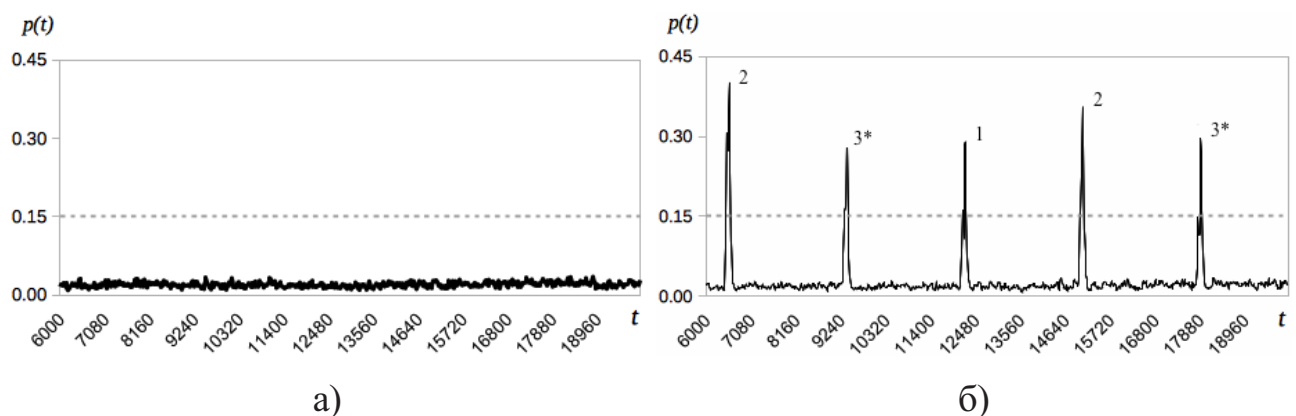


Рисунок 5.8 — Накопленные штрафы детектора оптического загрязнения для случаев отсутствия и присутствия загрязнения

Как и ожидалось, на втором графике наблюдаются резкие периодические скачки штрафа в местах оптического загрязнения, а на участках отсутствия загрязнения оба графика имеют схожее поведение. Нумерация пиков на рисунке соответствует позиции в списке зарегистрированных событий, а звездочка отмечает момент отправления сигнала о загрязнении.

Алгоритм был протестирован на 92 тестовых пакетах, в 3 из них на ленте присутствовало загрязнение; все пакеты были обработаны без ошибок.

## **5.6 Выводы по главе 5**

В данной главе были рассмотрены основные проблемы, возникающие при формировании входного сигнала рентгенографического сепаратора. Несмотря на различия в моделях помех, все алгоритмы детекции и коррекции нештатных ситуаций удалось разработать на одних и тех же простейших элементах обработки сигналов – гистограммном анализе, пороговых отсечениях и рекуррентных морфологических фильтрах, что позволяет реализовать эти алгоритмы в реальных промышленных решениях, где вычислительные ресурсы существенно ограничены, а обработка сигналов идет в режиме реального времени.

## Заключение

В ходе выполнения работы были получены следующие результаты:

1. Построена модель системы распознавания изображений и видеопотока с модулями оценки качества и обратной связью на каждом этапе обработки. Данная модель позволяет увеличить точность, устойчивость и скорость работы систем распознавания в неконтролируемых условиях съемки за счет возможности контроля уверенности результата и отказа в дальнейшей обработке в случае низкого качества входного изображения. В рамках модели рассмотрен пример задачи выбора наилучших кадров для распознавания в видеопотоке в условиях автоматической фокусировки камеры, приводящих к размытию кадра. Экспериментально показано значительное увеличение точности работы системы распознавания удостоверяющих личность документов с различными ограничениями на продолжительность видеопотока.

2. Разработан алгоритм автоматической оценки качества цветовой сегментации изображений печатных документов, основанный на независимой оценке текстового слоя через точность распознавания и графического слоя с помощью вычисления разницы плотности принадлежащих слою пикселей на размеченной карте насыщенности изображения. Экспериментально подтверждена пригодность алгоритма для автоматической оптимизации параметров цветовой сегментации.

3. Выполнено исследование негативного влияния бликов на точность поиска и распознавания документа на изображении, а также на устойчивость алгоритма вычисления взвешенного проективного оптического потока. Разработаны алгоритмы детектирования бликов и интерполяции поврежденного ими фона, экспериментально показано увеличение точности поиска и распознавания документов после фильтрации бликов. Экспериментально подтверждена неустойчивость алгоритма вычисления взвешенного проективного оптического потока к бликам в  $RGB$  пространстве и возможность его заметного улучшения при переходе в  $\alpha\beta$  пространство и исключении зашкаливающих областей.

4. Разработаны алгоритмы детектирования и локализации периодических фоновых элементов известной структуры на изображении документа, основанные на введенной математической модели изображения документа, предварительной обработке изображения и спектра двумерного дискретного преобразования Фурье, анализе спектра и итерационном решении нелинейной системы уравнений для



получения итогового результата локализации. Экспериментальные результаты демонстрируют высокую точность алгоритмов детектирования и локализации. Проведен анализ устойчивости разработанных алгоритмов к ошибкам нахождения границ документа с использованием искусственного моделирования, показавший устойчивость к небольшим ошибкам.

5. Разработаны алгоритмы детектирования нештатных ситуаций двухканального рентгенографического сепаратора для случаев раскалибровки пары источник/детектор, заслонения детектора и загрязнения конвейерной ленты, экспериментально показавшие высокую точность и позволяющие их применение в промышленных решениях для работы в реальном времени за счет высокой производительности и простоты реализации.

6. Практическая ценность полученных результатов подтверждается их использованием при проектировании и реализации промышленных систем распознавания. Реализации алгоритмов выбора наилучшего кадра для распознавания в видеопотоке, автоматической оценки качества цветовой сегментации изображений печатных документов, детектирования и фильтрации бликов на изображениях документов и детектирования и локализации периодических фоновых элементов известной структуры на изображении документа внедрены в программное обеспечение «Smart PDF/A», «Smart 3D OCR MRZ», «Smart PassportReader» и «Smart IDReader» компании ООО «Смарт Энджинс Сервис». Данные продукты интегрированы в информационные решения крупнейших российских банков, страховых компаний и организаций различных государственных структур, в том числе – международных. Реализация алгоритма контроля качества входного сигнала двухканального рентгенографического сепаратора внедрена в программный комплекс сепарации алмазов в рамках коммерческого проекта компании ООО «Визиллект Сервис». Построенная модель системы распознавания изображений и видеопотока с модулями оценки качества и обратной связью на каждом этапе обработки использовалась при проектировании данных коммерческих программных систем распознавания.

Основные результаты диссертационной работы представлены в журналах, рецензируемых ВАК, журналах, входящих в международную базу цитирования Scopus, а также сборниках трудов международных и всероссийских конференций, в том числе входящих в международные базы цитирования Scopus и Web of Science [21; 48; 135; 168; 187–206].

### Список литературы

1. *Bishop, C. M.* Pattern Recognition and Machine Learning (Information Science and Statistics) / C. M. Bishop. — Secaucus, NJ, USA : Springer-Verlag New York, Inc., 2006.
2. *Liu, J.* Pattern recognition: An overview / J. Liu, J. Sun, S. Wang // IJCSNS International Journal of Computer Science and Network Security. — 2006. — Vol. 6, no. 6. — P. 57—61.
3. *Арлазаров, В.* Документооборот как информационная база накопления знаний / В. Арлазаров, Н. Емельянов // Труды Института системного анализа Российской академии наук. — 2007. — Т. 29. — С. 6—48.
4. *Першиков, В. И.* Толковый словарь по информатике: более 10,000 терминов / В. И. Першиков, В. М. Савинков. — Финансы и статистика, 1991.
5. *Арлазаров, В.* Документооборот или управление знаниями? / В. Арлазаров, Н. Емельянов // Труды Института системного анализа Российской академии наук. — 2003. — С. 6—29.
6. *Бобылева, М.* Управленческий документооборот. От бумажного к электронному / М. Бобылева. — Litres, 2017.
7. *Волхонский, В.* Системы контроля и управления доступом / В. Волхонский. — 2015.
8. *Ворона, В.* Системы контроля и управления доступом / В. Ворона, В. Тихонов // М.: Горячая линия–Телеком. — 2010. — Т. 272.
9. *Волеводз, А. Г.* Удостоверяющие личность документы с биометрическими персональными данными: международные стандарты, иностранный опыт, отечественное правовое регулирование и его проблемы / А. Г. Волеводз // Ученые записки Санкт-Петербургского имени В.Б. Бобкова филиала Российской таможенной академии. — 2015. — № 1. — С. 87—117.
10. Анализ особенностей использования стационарных и мобильных малоразмерных цифровых видео камер для распознавания документов / В. Арлазаров [и др.] // Информационные технологии и вычислительные системы. — 2014. — № 3.

11. *Спиридонов, И. Н.* Биометрические технологии в интегрированных автоматизированных системах безопасности государства / И. Н. Спиридонов // Вестник Московского государственного технического университета им. НЭ Баумана. Серия «Приборостроение». — 2011. — С.
12. Software and hardware user authentication methods in the information and control systems based on biometrics / Z. Boriev [et al.] // IOP Conference Series: Materials Science and Engineering. Vol. 124. — IOP Publishing. 2016. — P. 012006.
13. *Грибачев, К.* Биометрические технологии в СКУД / К. Грибачев // Алгоритм безопасности. — 2011. — № 4. — С. 44.
14. *Дятлов, Е.* Машинное зрение (аналитический обзор) / Е. Дятлов // Математические машины и системы. — 2013. — № 2.
15. *Golnabi, H.* Design and application of industrial machine vision systems / H. Golnabi, A. Asadpour // Robotics and Computer-Integrated Manufacturing. — 2007. — Vol. 23, no. 6. — P. 630—637.
16. Методы и алгоритмы обработки изображений в системах технического зрения. Машинное зрение / С. Савицкий [и др.] // Российская наука в современном мире. Т. 10. — 2017. — С. 61.
17. *Awcock, G. J.* Applied image processing / G. J. Awcock, R. Thomas. — McGraw-Hill, Inc., 1995.
18. *Bow, S. T.* Pattern recognition and image preprocessing / S. T. Bow. — CRC press, 2002.
19. *Povolotskiy, M. A.* Russian License Plate Segmentation Based On Dynamic Time Warping / M. A. Povolotskiy, E. G. Kuznetsova, T. M. Khanipov // European Conference on Modelling and Simulation, 2017, Proceedings. — 2017. — P. 285—291.
20. *Gürel, C.* Design Of A Face Recognition System / C. Gürel, A. Erden // Proc. The 15th Int. Conference on Machine Design and Production (UMTIK 2012). — 2012.
21. Smart IDReader: Document Recognition in Video Stream / K. Bulatov [et al.] // 14th IAPR International Conference on Document Analysis and Recognition. Vol. 6. — IEEE, 2017. — P. 39—44.

22. *Schroff, F.* Facenet: A unified embedding for face recognition and clustering / F. Schroff, D. Kalenichenko, J. Philbin // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2015. — P. 815—823.
23. Алгоритм применения N-грамм для корректировки результатов распознавания / Т. Манжиков [и др.] // Вестник Российского фонда фундаментальных исследований. — 2016. — Т. 92, № 4. — С. 116—123.
24. *Fiscus, J. G.* A post-processing system to yield reduced word error rates: Recognizer output voting error reduction (ROVER) / J. G. Fiscus // Automatic Speech Recognition and Understanding, 1997. Proceedings. — IEEE. 1997. — P. 347—354.
25. Generalization of the Viola-Jones method as a decision tree of strong classifiers for real-time object recognition in video stream / A. Minkina [et al.] // ICMV 2014, Proceedings. — 2015.
26. *Wallace, G. K.* The JPEG still picture compression standard / G. K. Wallace // Consumer Electronics, IEEE Transactions on. — 1992. — Vol. 38, no. 1. — P. 18—34.
27. *Queiroz, R. L. de.* Mixed raster content (MRC) model for compound image compression / R. L. de Queiroz, R. R. Buckley, M. Xu // Electronic Imaging'99. — International Society for Optics, Photonics. 1998. — P. 1106—1117.
28. A general segmentation scheme for DjVu document compression / P. Haffner [et al.] // ISMM. Vol. 2. — 2002.
29. *Усилин, С.* Cognitive PDF/A—технология оцифровки текстовых документов для публикации в Интернете и долговременного архивного хранения / С. Усилин, Д. Николаев, В. Постников // Труды Института системного анализа РАН. Технологии программирования и хранения данных. — 2009. — С. 159—173.
30. *Gonzalez, R. C.* Digital Image Processing (3rd Edition) / R. C. Gonzalez, R. E. Woods. — Upper Saddle River, NJ, USA : Prentice-Hall, Inc., 2006.
31. *Казанов, М. Д.* Сегментация цветных изображений печатных документов : дис. ... канд. тех. наук : 05.13.01 / М. Д. Казанов. — М., 2008. — 115 с.
32. *Otsu, N.* A threshold selection method from gray-level histograms / N. Otsu // Automatica. — 1975. — Vol. 11, no. 285—296. — P. 23—27.

33. *Russell, S. J.* Artificial Intelligence: A Modern Approach / S. J. Russell, P. Norvig. — 2nd ed. — Pearson Education, 2003.
34. *Арлазаров, В. В.* Структурирование визуальных представлений информационной среды и методы определения надежности распознавания : дис. ... канд. техн. наук : 05.13.01 / В. В. Арлазаров. — М., 2004. — 122 с.
35. Методы интеграции результатов распознавания текстовых полей документов в видеопотоке мобильного устройства / К. Булатов [и др.] // Вестник Российского фонда фундаментальных исследований. — 2016. — Т. 92, № 4. — С. 109—115.
36. *Samarasinghe, S.* Neural networks for applied sciences and engineering: from fundamentals to complex pattern recognition / S. Samarasinghe. — CRC Press, 2016.
37. *Мерков, А. Б.* Распознавание образов. Введение в методы статистического обучения / А. Б. Мерков. — Едиториал УРСС, 2011. — 256 с.
38. *Denker, J. S.* transforming neural-net output levels to probability distributions / J. S. Denker, Y. LeCun // Advances in Neural Information Processing Systems (NIPS 1990). Vol. 3 / ed. by R. Lippmann, J. Moody, D. Touretzky. — Denver, CO : Morgan Kaufman, 04/1991.
39. Multiclassification: reject criteria for the Bayesian combiner / P. Foggia [et al.] // Pattern Recognition. — 1999. — Vol. 32, no. 8. — P. 1435—1447.
40. *Tax, D. M. J.* Growing a Multi-class Classifier with a Reject Option / D. M. J. Tax, R. P. W. Duin // Pattern Recogn. Lett. — New York, NY, USA, 2008. — July. — Vol. 29, no. 10. — P. 1565—1570.
41. *Арлазаров, В.* Метод определения надежности распознавания в задаче распознавания тисненых символов / В. Арлазаров, К. Булатов, С. Карпенко // Труды ИСА РАН. — 2013. — Т. 63, № 3. — С. 117—122.
42. A global optimization technique for statistical classifier design / D. Miller [et al.] // IEEE Transactions on Signal Processing. — 1996. — Dec. — Vol. 44, no. 12. — P. 3108—3122.
43. *Rose, K.* Deterministic annealing for clustering, compression, classification, regression, and related optimization problems / K. Rose // Proceedings of the IEEE. — 1998. — Nov. — Vol. 86, no. 11. — P. 2210—2239.

44. *Bulatov, K. B.* Reducing Overconfidence In Neural Networks By Dynamic Variation of Recognizer Relevance. / K. B. Bulatov, D. V. Polevoy // ECMS. — 2015. — P. 488—491.
45. Regularizing neural networks by penalizing confident output distributions / G. Pereyra [et al.] // arXiv preprint arXiv:1701.06548. — 2017.
46. Meta-recognition: The theory and practice of recognition score analysis / W. J. Scheirer [et al.] // IEEE transactions on pattern analysis and machine intelligence. — 2011. — Vol. 33, no. 8. — P. 1689—1695.
47. *Cruz, J. A.* Applications of machine learning in cancer prediction and prognosis / J. A. Cruz, D. S. Wishart // Cancer informatics. — 2006. — Vol. 2.
48. Ключевые аспекты распознавания документов с использованием малоразмерных цифровых камер / Д. Полевой [и др.] // Вестник Российского фонда фундаментальных исследований. — 2016. — Т. 92, № 4. — С. 97—108.
49. *Русинов, М. М.* Композиция оптических систем / М. М. Русинов. — Кн. дом «ЛИБРОКОМ», 2010.
50. *Тропченко, А.* Методы сжатия изображений, аудиосигналов и видео / А. Тропченко, А. Тропченко // Учебное пособие. — 2009.
51. *Tekalp, A. M.* Digital video processing / A. M. Tekalp. — Prentice Hall Press, 2015.
52. Image interpolation using Shearlet based iterative refinement / H. Lakshman [et al.] // Signal Processing: Image Communication. — 2015. — Vol. 36. — P. 83—94.
53. *Hecht, E.* Optics / E. Hecht. — 5th ed. — Pearson Education, 2016.
54. *Zhuo, S.* Defocus map estimation from a single image / S. Zhuo, T. Sim // Pattern Recognition. — 2011. — Vol. 44, no. 9. — P. 1852—1858.
55. Проблемы распознавания машиночитаемых зон с использованием малоформатных цифровых камер мобильных устройств / К. Булатов [и др.] // Труды Института Системного Анализа Российской Академии Наук. — 2015. — Т. 65, № 3.
56. *Dutta, A.* Predicting Face Recognition Performance Using Image Quality / A. Dutta, R. Veldhuis, L. Spreeuwiers // ArXiv e-prints. — 2015. — Oct. — arXiv: 1510.07119 [cs.CV].



57. Quantifying how lighting and focus affect face recognition performance / J. R. Beveridge [et al.] // Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on. — IEEE. 2010. — P. 74—81.
58. Has my algorithm succeeded? an evaluator for human pose estimators / N. Jammalamadaka [et al.] // European Conference on Computer Vision. — Springer. 2012. — P. 114—128.
59. *Vasiljevic, I.* Examining the Impact of Blur on Recognition by Convolutional Networks / I. Vasiljevic, A. Chakrabarti, G. Shakhnarovich // CoRR. — 2016. — Vol. 1611.05760.
60. *Dodge, S.* Understanding How Image Quality Affects Deep Neural Networks / S. Dodge, L. Karam // ArXiv e-prints. — 2016. — Apr. — arXiv: 1604.04004 [cs.CV].
61. Atoms of recognition in human and computer vision / S. Ullman [et al.] // Proceedings of the National Academy of Sciences. — 2016. — Vol. 113, no. 10. — P. 2744—2749.
62. *Nguyen, A.* Deep neural networks are easily fooled: High confidence predictions for unrecognizable images / A. Nguyen, J. Yosinski, J. Clune // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. — 2015. — P. 427—436.
63. *Kurakin, A.* Adversarial examples in the physical world / A. Kurakin, I. Goodfellow, S. Bengio // arXiv preprint arXiv:1607.02533. — 2016.
64. *Su, J.* One pixel attack for fooling deep neural networks / J. Su, D. Vasconcellos Vargas, S. Kouichi // ArXiv e-prints. — 2017. — Oct. — arXiv: 1710.08864 [cs.LG].
65. Towards deep learning models resistant to adversarial attacks / A. Madry [et al.] // arXiv preprint arXiv:1706.06083. — 2017.
66. *Bovik, A. C.* Handbook of Image and Video Processing / A. C. Bovik. — 2000. — P. 974.
67. *Mrak, M.* Picture quality measures in image compression systems / M. Mrak, S. Grgic, M. Grgic // EUROCON 2003. Computer as a Tool. The IEEE Region 8. Vol. 1. — 09/2003. — P. 233—236.

68. *Wang, Z.* Why is image quality assessment so difficult? / Z. Wang, A. C. Bovik, L. Lu // IEEE International Conference on Acoustics Speech and Signal Processing. — IEEE, 05/2002. — P. IV–3313–IV–3316.
69. Subjective image quality assessment: A method based on signal detection theory / Y. He [et al.] // 2009 IEEE International Conference on Systems, Man and Cybernetics. — IEEE, 10/2009. — P. 4915–4919.
70. *Mohammadi, P.* Subjective and Objective Quality Assessment of Image: A Survey / P. Mohammadi // arXiv preprint arXiv:1406.7799. — 2014. — June. — P. 1–50.
71. *Wajid, R.* A study of human perception similarity for image quality assessment / R. Wajid, A. Bin Mansoor, M. Pedersen // Colour and Visual Computing Symposium (CVCS), 2013. — 2013. — P. 1–6.
72. *Wang, Z.* Applications of objective image quality assessment methods [Applications Corner] / Z. Wang // IEEE Signal Processing Magazine. — 2011. — Vol. 28, no. 6. — P. 137–142.
73. *Wang, Z.* Modern Image Quality Assessment. Vol. 2 / Z. Wang, A. C. Bovik. — 2006. — P. 1–156.
74. *Xiao, F.* Psychophysical thresholds and digital camera sensitivity: the thousand-photon limit / F. Xiao, J. E. Farrell, B. A. Wandell. — 2005.
75. Perceptually-tuned multiscale color-texture segmentation / J. Chen [et al.] // International Conference on Image Processing. Vol. 2. — 10/2004. — P. 921–924.
76. Image quality assessment based on a degradation model / N. Damera-Venkata [et al.] // Image Processing, IEEE Transactions on. — 2000. — Vol. 9, no. 4. — P. 636–650.
77. *Piella, G.* A new quality metric for image fusion / G. Piella, H. Heijmans // Image Processing, 2003. ICIP 2003. Proceedings. 2003 International Conference on. Vol. 3. — 09/2003. — P. 173–176.
78. *Rusinol, M.* Combining focus measure operators to predict OCR accuracy in mobile-captured document images / M. Rusinol, J. Chazalon, J. M. Ogier // Proceedings - 11th IAPR International Workshop on Document Analysis Systems, DAS 2014. — 2014. — P. 181–185.

79. *Peng, X.* Document image OCR accuracy prediction via latent Dirichlet allocation / X. Peng, H. Cao, P. Natarajan // Proceedings of the International Conference on Document Analysis and Recognition, ICDAR. — 2015. — Vol. 2015—Novem. — P. 771—775.
80. Incorporating image quality in multi-algorithm fingerprint verification / J. Fierrez-Aguilar [et al.] // ICB. — Springer. 2006. — P. 213—220.
81. Improving iris recognition accuracy via cascaded classifiers / Z. Sun [et al.] // IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). — 2005. — Vol. 35, no. 3. — P. 435—441.
82. *Mantiuk, R. K.* Comparison of four subjective methods for image quality assessment / R. K. Mantiuk, A. Tomaszewska, R. Mantiuk // Computer Graphics Forum. — 2012. — Vol. 31, no. 8. — P. 2478—2491.
83. Comparing subjective image quality measurement methods for the creation of public databases / J. Redi [et al.] // Spie-Is&T. — 2010. — Vol. 7529. — P. 752903-1—11.
84. Study of subjective and objective quality assessment of video. / K. Seshadri-nathan [et al.] // IEEE transactions on image processing : a publication of the IEEE Signal Processing Society. — 2010. — June. — Vol. 19, no. 6. — P. 1427—41.
85. *Dimova, R.* Performance Analysis of QoS parameters for Voice over IP Applications in a LAN Segment / R. Dimova, G. Georgiev, Z. Stanchev // International Scientific Conference Computer Science. — 2008. — P. 10.
86. Rank learning on training set selection and image quality assessment / L. Xu [и др.]. — 2014.
87. *Silverstein, D. A.* An Efficient Method for Paired-Comparison / D. A. Silverstein, J. E. Farrell // Journal of Electronic Imaging. — 2001. — Vol. 10, no. 2. — P. 394—398.
88. *Lewandowska (Tomaszewska), A.* Scene reduction for subjective image quality assessment / A. Lewandowska (Tomaszewska) // Journal of Electronic Imaging. — 2016. — Jan. — Vol. 25, no. 1. — P. 013015.
89. *Ye, P.* Active Sampling for Subjective Image Quality Assessment / P. Ye, D. Doermann // 2014 IEEE Conference on Computer Vision and Pattern Recognition. — IEEE, 06/2014. — P. 4249—4256.

90. *Dijk, A. M. van.* Quality assessment of coded images using numerical category scaling / A. M. van Dijk, J.-B. Martens, A. B. Watson // SPIE 2451, Advanced Image and Video Communications and Storage Technologies, 90 / ed. by N. Ohta, H. U. Lemke, J. C. Lehoureau. — 02/1995. — P. 90—101.
91. Video Quality Experts Group: Current Results and Future Directions / A. M. Rohaly [et al.] // Visual Communications and Image Processing 2000. — 2000. — P. 742—753.
92. *Chandler, D. M.* Seven Challenges in Image Quality Assessment: Past, Present, and Future Research / D. M. Chandler // ISRN Signal Processing. — 2013. — Vol. 2013. — P. 1—53.
93. *Rodgers, J. L.* Thirteen Ways to Look at the Correlation Coefficient / J. L. Rodgers, W. A. Nicewander // The American Statistician. — 1988. — Vol. 42, no. 1. — P. 59—66.
94. Image quality assessment: from error visibility to structural similarity / Z. Wang [et al.] // Image Processing, IEEE Transactions on. — 2004. — Apr. — Vol. 13, no. 4. — P. 600—612.
95. *Goldmark, P.* Quality in Television Pictures / P. Goldmark, J. Dyer // Journal of the Society of Motion Picture Engineers. — 1940. — Sept. — Vol. 35, no. 9. — P. 234—253.
96. *Grgić, S.* Reliability of objective picture quality measures / S. Grgić, M. Grgić, M. Mrak // Journal of Electrical Engineering. — 2004. — Vol. 55. — P. 3—10.
97. *Wang, Z.* Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures / Z. Wang, A. Bovik // Signal Processing Magazine, IEEE. — 2009. — Jan. — Vol. 26, no. 1. — P. 98—117.
98. *Hore, A.* Image Quality Metrics: PSNR vs. SSIM / A. Hore, D. Ziou // Pattern Recognition (ICPR), 2010 20th International Conference on. — 08/2010. — P. 2366—2369.
99. *Huynh-Thu, Q.* Scope of validity of PSNR in image/video quality assessment / Q. Huynh-Thu, M. Ghanbari // Electronics Letters. — 2008. — June. — Vol. 44, no. 13. — P. 800—801.
100. *Thomos, N.* Optimized transmission of JPEG2000 streams over wireless channels / N. Thomos, N. V. Boulgouris, M. G. Strintzis // Image Processing, IEEE Transactions on. — 2006. — Vol. 15, no. 1. — P. 54—67.

101. *Xu, S.* No-reference/blind image quality assessment: a survey / S. Xu, S. Jiang, W. Min // IETE Technical Review. — 2017. — Vol. 34, no. 3. — P. 223—245.
102. *Yuan, Y.* Image quality assessment: a sparse learning way / Y. Yuan, Q. Guo, X. Lu // Neurocomputing. — 2015. — Vol. 159. — P. 227—241.
103. *Mittal, A.* No-reference image quality assessment in the spatial domain / A. Mittal, A. K. Moorthy, A. C. Bovik // IEEE Transactions on Image Processing. — 2012. — Vol. 21, no. 12. — P. 4695—4708.
104. Blind image quality assessment using joint statistics of gradient magnitude and Laplacian features / W. Xue [et al.] // IEEE Transactions on Image Processing. — 2014. — Vol. 23, no. 11. — P. 4850—4862.
105. *Zhang, L.* A feature-enriched completely blind image quality evaluator / L. Zhang, L. Zhang, A. C. Bovik // IEEE Transactions on Image Processing. — 2015. — Vol. 24, no. 8. — P. 2579—2591.
106. On the use of deep learning for blind image quality assessment / S. Bianco [et al.] // arXiv preprint arXiv:1602.05531. — 2016.
107. Specular Reflection Separation Using Dark Channel Prior / H. Kim [et al.] // 2013 IEEE Conference on Computer Vision and Pattern Recognition. — 06/2013. — P. 1460—1467.
108. *Yang, Q.* Real-time Specular Highlight Removal Using Bilateral Filtering / Q. Yang, S. Wang, N. Ahuja // Proceedings of the 11th European Conference on Computer Vision: Part IV. — Heraklion, Crete, Greece : Springer-Verlag, 2010. — P. 87—100.
109. A novel and effective method for specular detection and removal by tensor voting / T. Nguyen [et al.] // 2014 IEEE International Conference on Image Processing (ICIP). — 10/2014. — P. 1061—1065.
110. *Kayabol, K.* Image Source Separation Using Color Channel Dependencies / K. Kayabol, E. E. Kuruoglu, B. Sankur // Independent Component Analysis and Signal Separation: 8th International Conference, ICA 2009, Paraty, Brazil, March 15-18, 2009. Proceedings / ed. by T. Adali [et al.]. — Berlin, Heidelberg : Springer Berlin Heidelberg, 2009. — P. 499—506.
111. *Artusi, A.* A Survey of Specularity Removal Methods / A. Artusi, F. Banterle, D. Chetverikov // Computer Graphics Forum. — 2011.

112. *Kokaram, A. C.* On missing data treatment for degraded video and film archives: a survey and a new Bayesian approach / A. C. Kokaram // IEEE Transactions on Image Processing. — 2004. — Mar. — Vol. 13, no. 3. — P. 397—415.
113. Glare Aware Photography: 4D Ray Sampling for Reducing Glare Effects of Camera Lenses / R. Raskar [et al.] // ACM Transactions on Graphics. — New York, NY, USA, 2008. — Aug. — Vol. 27, no. 3. — 56:1—56:10.
114. *Cho, C.-S.* Glare region detection in night scene using multi-layering / C.-S. Cho, J. Song, J.-I. Park // The Third International Conference on Digital Information Processing and Communications. — The Society of Digital Information, Wireless Communication. 2013. — P. 467—469.
115. Real Time Static Glare Identification in ITS / M. Shehata [et al.] // Practical Real World Technologies for Communications and Embedded Platforms. — 2006.
116. Toward Accurate and Fast Iris Segmentation for Iris Biometrics / Z. He [et al.] // IEEE Trans. Pattern Anal. Mach. Intell. — Washington, DC, USA, 2009. — Sept. — Vol. 31, no. 9. — P. 1670—1684.
117. *Madooei, A.* Detecting specular highlights in dermatological images / A. Madooei, M. S. Drew // 2015 IEEE International Conference on Image Processing (ICIP). — 09/2015. — P. 4357—4360.
118. *Vasconcelos, M. J. M.* Automatic Reflection Detection on Dermatological Images Acquired via Mobile Devices / M. J. M. Vasconcelos, L. Rosado // MIUA. — 2014. — P. 85—90.
119. *Stoyanov, D.* Removing specular reflection components for robotic assisted laparoscopic surgery. / D. Stoyanov, G.-Z. Yang // ICIP (3). — IEEE, 2005. — P. 632—635.
120. *Tchoulack, S.* A video stream processor for real-time detection and correction of specular reflections in endoscopic images / S. Tchoulack, J. M. P. Langlois, F. Cheriet // 2008 Joint 6th International IEEE Northeast Workshop on Circuits and Systems and TAISA Conference. — 06/2008. — P. 49—52.
121. *Stehle, T.* Removal of Specular Reflections in Endoscopic Images / T. Stehle // Acta Polytechnica: Journal of Advanced Engineering. — 2006. — Vol. 46, no. 4. — P. 32—36.



122. *Karapetyan, G.* Automatic detection and concealment of specular reflections for endoscopic images / G. Karapetyan, H. Sarukhanyan // Ninth International Conference on Computer Science and Information Technologies Revised Selected Papers. — 09/2013. — P. 1—8.
123. Automatic Segmentation and Inpainting of Specular Highlights for Endoscopic Imaging / M. Arnold [et al.] // Journal on Image and Video Processing. — New York, NY, United States, 2010. — Jan. — 9:1—9:12.
124. A Specular Reflection Suppression Method for Endoscopic Images / J. J. Guo [et al.] // 2016 IEEE Second International Conference on Multimedia Big Data (BigMM). — 04/2016. — P. 125—128.
125. *Das, A.* Elimination of specular reflection and identification of ROI: The first step in automated detection of Cervical Cancer using Digital Colposcopy / A. Das, A. Kar, D. Bhattacharyya // 2011 IEEE International Conference on Imaging Systems and Techniques. — 05/2011. — P. 237—241.
126. *Lange, H.* Automatic glare removal in reflectance imagery of the uterine cervix / H. Lange. — 2005.
127. Informative frame classification for endoscopy video / J. Oh [et al.] // Medical Image Analysis. — 2007. — Vol. 11, no. 2. — P. 110—127.
128. *Reitan, R. C.* Automated image quality control / R. C. Reitan. — 1997. — US Patent 5,600,574.
129. *Proudfoot, R. A.* Systems and methods for glare removal using polarized filtering in document scanning / R. A. Proudfoot, M. Levoy. — 2012. — US Patent 8,174,739.
130. *Bocharov, K.* Detecting glare in a frame of image data / K. Bocharov, M. Kostyukov. — 2014. — US Patent App. 14/564,424.
131. *Farahmand, A.* Document Image Noises and Removal Methods / A. Farahmand, A. Sarrafzadeh, J. Shanbehzadeh // IMECS. — Newswood Limited, 2013. — P. 436—440.
132. *Ali, M.* Background noise detection and cleaning in document images / M. Ali // Proceedings of the 13th International Conference on Pattern Recognition. Vol. 3. — 08/1996. — P. 758—762.

133. *Agrawal, M.* Clutter Noise Removal in Binary Document Images / M. Agrawal, D. Doermann // Proceedings of the 2009 10th International Conference on Document Analysis and Recognition. — Washington, DC, USA : IEEE Computer Society, 2009. — P. 556—560.
134. *Liang, S.* A Morphological Approach to Text String Extraction from Regular Periodic Overlapping Text/Background Images / S. Liang, M. Ahmadi, M. Shridhar // CVGIP: Graph. Models Image Process. — Orlando, FL, USA, 1994. — Sept. — Vol. 56, no. 5. — P. 402—413.
135. Diamond recognition algorithm using two-channel x-ray radiographic separator / D. P. Nikolaev [et al.] // Seventh International Conference on Machine Vision. Vol. 9445. — 2015. — P. 944507.
136. *Kuroptev, A.* Precise localisation of supporting grids for handwritten filling in document forms using dynamic programming and morphological filtering / A. Kuroptev, D. Nikolaev, V. Postnikov // Proceedings of Institute for Systems Analysis of Russian Academy of Sciences. — 2013.
137. *Brigham, E. O.* The Fast Fourier Transform and Its Applications / E. O. Brigham. — Prentice-Hall, Inc., 1988.
138. *Starovoitov, V. V.* Texture periodicity detection: features, properties, and comparisons / V. V. Starovoitov, S.-Y. Jeong, R.-H. Park // IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans. — 1998. — Nov. — Vol. 28, no. 6. — P. 839—849.
139. *Gruzman, I.* Detection of quasiperiodic textures with the use of two-dimensional power spectrum characteristics / I. Gruzman, K. Y. Petrova // Optoelectronics, Instrumentation and Data Processing. — 2015. — Vol. 51, no. 2. — P. 103—109.
140. *Gruzman, I.* Estimation of quasiperiodic texture lattice node coordinates with the use of two-dimensional local spectra characteristics / I. Gruzman, K. Y. Petrova // Optoelectronics, Instrumentation and Data Processing. — 2016. — Vol. 52, no. 3. — P. 217—222.
141. *Puglisi, G.* Periodic pattern detection for real-time application / G. Puglisi, S. Battiato // International Workshop on Fuzzy Logic and Applications. — Springer. 2009. — P. 271—278.

142. *Kortylewski, A.* Unsupervised footwear impression analysis and retrieval from crime scene data / A. Kortylewski, T. Albrecht, T. Vetter // Asian Conference on Computer Vision. — Springer. 2014. — P. 644—658.
143. Detection of Irregularities in Regular Patterns / J. Vartiainen [et al.] // Mach. Vision Appl. — Secaucus, NJ, USA, 2008. — May. — Vol. 19, no. 4. — P. 249—259.
144. *Hui, S.* Discrete Fourier transform based pattern classifiers / S. Hui, S. Żak // Bulletin of the Polish Academy of Sciences: Technical Sciences. — 2014. — Vol. 62, no. 1. — P. 15—22.
145. *Tai, C.-J.* Effects on Fourier Peaks Used for Periodic Pattern Detection / C.-J. Tai, R. Ulichney, J. P. Allebach // Electronic Imaging. — 2016. — Vol. 2016, no. 13. — P. 1—8.
146. A no-reference quality metric for measuring image blur / E. Ong [et al.] // Proceedings - 7th International Symposium on Signal Processing and Its Applications, ISSPA 2003. — 2003. — Vol. 1, no. 4. — P. 469—472.
147. The blur effect: perception and estimation with a new no-reference perceptual blur metric. / F. Crete [et al.] // Human vision and electronic imaging. Vol. 12. — 2007. — P. 64920.
148. *Ji, H.* Motion blur identification from image gradients / H. Ji, C. Liu // Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. — IEEE. 2008. — P. 1—8.
149. *Oliveira, J. P.* Parametric blur estimation for blind restoration of natural images: Linear motion and out-of-focus / J. P. Oliveira, M. A. Figueiredo, J. M. Bioucas-Dias // IEEE Transactions on Image Processing. — 2014. — Vol. 23, no. 1. — P. 466—477.
150. A No Reference Quality Metric for Measuring Image Blur In Wavelet Domain / F. Kerouh [et al.] // International Journal of Digital Information and Wireless Communications. — 2011. — Vol. 1, no. 4. — P. 767—776.
151. No-Reference Image Blur Assessment Based on Discrete Orthogonal Moments / L. Li [et al.] // IEEE Transactions on Cybernetics. — 2015. — Vol. 46. — P. 39—50.

152. *Dai, S.* Motion from blur / S. Dai, Y. Wu // Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on. — IEEE. 2008. — P. 1—8.
153. *Ye, P.* Document Image Quality Assessment: A Brief Survey / P. Ye, D. Doermann // 2013 12th International Conference on Document Analysis and Recognition. — 2013. — P. 723—727.
154. *Sharma, P.* An analysis of vision based techniques for quality assessment and enhancement of camera captured document images / P. Sharma, S. Sharma // Cloud System and Big Data Engineering (Confluence), 2016 6th International Conference. — IEEE. 2016. — P. 425—428.
155. *Peng, X.* Document Image Quality Assessment Using Discriminative Sparse Representation / X. Peng, H. Cao, P. Natarajan // Document Analysis Systems (DAS), 2016 12th IAPR Workshop on. — IEEE. 2016. — P. 227—232.
156. Blind image quality assessment based on high order statistics aggregation / J. Xu [et al.] // IEEE Transactions on Image Processing. — 2016. — Vol. 25, no. 9. — P. 4444—4457.
157. Client side filtering of card OCR images / X. Wang [et al.]. — 12 2/2014. — US Patent 8,903,136.
158. *Shvets, E. A.* Complex approach to long-term multi-agent mapping in low dynamic environments / E. A. Shvets, D. P. Nikolaev // 8'th International Conference on Machine Vision. — International Society for Optics, Photonics. 2015.
159. Robust automatic focus algorithm for low contrast images using a new contrast measure / X. Xu [et al.] // Sensors. — 2011. — Vol. 11, no. 9. — P. 8281—8294.
160. *Rangoni, Y.* OCR Based Thresholding. / Y. Rangoni, F. Shafait, T. M. Breuel // MVA. — 2009. — P. 98—101.
161. *Smith, R.* An Overview of the Tesseract OCR Engine / R. Smith // Proceedings of the Ninth International Conference on Document Analysis and Recognition. Vol. 2. — Washington, DC, USA : IEEE Computer Society, 2007. — P. 629—633.
162. *Audet, C.* Mesh adaptive direct search algorithms for constrained optimization / C. Audet, J. E. Dennis Jr // SIAM Journal on optimization. — 2006. — Vol. 17, no. 1. — P. 188—217.

163. *Viola, P.* Rapid object detection using a boosted cascade of simple features / P. Viola, M. Jones // Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001. Vol. 1. — 2001. — P. 511—518.
164. *Buades, A.* Patch-Based Video Denoising With Optical Flow Estimation / A. Buades, J.-L. Lisani, M. Miladinović // IEEE Transactions on Image Processing. — 2016. — Vol. 25, no. 6. — P. 2573—2586.
165. UAV navigation on the basis of the feature points detection on underlying surface / I. Konovalenko [et al.]. — 2015.
166. Optical benchmarking of security document readers for automated border control / K. Valentín [et al.] // SPIE Security+ Defence. — International Society for Optics, Photonics. 2016. — P. 999502—999502.
167. *Baker, S.* Lucas-Kanade 20 Years On: A Unifying Framework / S. Baker, I. Matthews // Int. J. Comput. Vision. — 2004. — Feb. — Vol. 56, no. 3. — P. 221—255.
168. Алгоритм взвешенного поиска проективного оптического потока, устойчивый к бликам / Д. Шепелев [и др.] // Информационные технологии и системы, сборник трудов. — 2017.
169. *Николаев, П. П.* Некоторые алгоритмы узнавания окраски поверхностей / П. П. Николаев // Моделирование обучения и поведения. — 1975. — С. 121—151.
170. О проблемах сопоставления пикселей стереопары с точки зрения линейной модели формирования цветного изображения / Д. Шепелев [и др.] // Сенсорные системы. — 2017. — Т. 31, № 2. — С. 150—160.
171. *Zhou, W.* Binocular Stereo Dense Matching in the Presence of Specular Reflections / W. Zhou, C. Kambhamettu // 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). Vol. 2. — 2006. — P. 2363—2370.
172. A Uniform Framework for Estimating Illumination Chromaticity, Correspondence, and Specular Reflection / Q. Yang [et al.] // IEEE Transactions on Image Processing. — 2011. — Jan. — Vol. 20, no. 1. — P. 53—63.
173. Visual appearance based document image classification / S. Usilin [et al.] // ICIP. — IEEE, 2010. — P. 2133—2136.

174. Real time rectangular document detection on mobile devices / N. Skoryukina [et al.] // Seventh International Conference on Machine Vision. Vol. 9445. — 2015. — 94452A.
175. *Soille, P.* Morphological Image Analysis: Principles and Applications / P. Soille // — Springer-Verlag New York, Inc., 1999. — P. 170—171.
176. *Watson, G. S.* Statistics on spheres. Vol. 6 / G. S. Watson. — Wiley-Interscience, 1983.
177. *Lawson, C. L.* Solving least squares problems. Vol. 15 / C. L. Lawson, R. J. Hanson. — Philadelphia, PA : Society for Industrial, Applied Mathematics (SIAM), 1995. — (Classics in Applied Mathematics).
178. Белоцерковский, С. Р. Рентгенографическая сепарация алмазов / С. Р. Белоцерковский, В. Н. Яковлев // Горный журнал. — 2011. — № 1. — С. 78—80.
179. Способ сепарации алмазов / Е. Потрахов [и др.]. — 2012. — Патент РФ №2470714.
180. Миронов, В. Способ сепарации минералов / В. Миронов. — 2010. — Патент РФ №2379130.
181. Миронов, В. П. Повышение эффективности рентгенолюминесцентной сепарации алмазов путем фильтрации спектра / В. П. Миронов // Обогащение руд. — 2017. — № 2.
182. Люминесцентный сепаратор минералов и способ контроля работы сепаратора / Е. Владимиров [и др.]. — 2005. — Патент РФ №2249490.
183. Миронов, В. П. Сепарация алмазосодержащего сырья на основе рентгенолюминесцентной абсорбции и использование имитаторов для контроля процесса / В. П. Миронов // Горный журнал. — 2012. — № 2.
184. Владимиров, Е. Сепаратор для обогащения алмазосодержащих материалов с программным обеспечением в среде ОС РВ QNX / Е. Владимиров // Современные технологии автоматизации. — 2011. — № 4.
185. Грешилов, А. А. Математические методы построения прогнозов / А. А. Грешилов, В. А. Стакун, А. А. Стакун. — Радио и связь М., 1997.
186. *Gustafsson, F.* Adaptive filtering and change detection. Vol. 1 / F. Gustafsson. — Citeseer, 2000.



187. *Chernov, T. S.* An algorithm for detection and phase estimation of protective elements periodic lattice on document image / T. S. Chernov, S. I. Kolmakov, D. P. Nikolaev // *Pattern Recognition and Image Analysis*. — 2017. — Vol. 27, no. 1. — P. 53—65.
188. Оценка качества входных изображений в системах распознавания видеопотока / Т. Чернов [и др.] // *Информационные технологии и вычислительные системы*. — 2017. — № 4. — С. 71—82.
189. *Чернов, Т.* Детектирование и фильтрация бликов в задачах распознавания документов с мобильных устройств / Т. Чернов // *Труды Института системного анализа Российской академии наук*. — 2017. — Т. 67, № 1. — С. 66—74.
190. Алгоритм взвешенного поиска проективного оптического потока, устойчивый к бликам / Д. Шепелев [и др.] // *Сенсорные системы*. — 2018. — в печати.
191. Исследование методов сегментации изображений текстовых блоков документов с помощью алгоритмов структурного анализа и машинного обучения / Т. Чернов [и др.] // *Вестник Российского фонда фундаментальных исследований*. — 2016. — Т. 92, № 4. — С. 55—71.
192. *Арлазаров, В.* Метод нечеткого поиска изображений в больших объемах видеоданных / В. Арлазаров, К. Булатов, Т. Чернов // *Системы высокой доступности*. — 2016. — Т. 12, № 1. — С. 53—58.
193. *Николаев, Д.* Метод автоматической оценки качества цветовой сегментации в задаче упаковки изображений печатных документов / Д. Николаев, Д. Полевой, Т. Чернов // *Труды Института системного анализа Российской академии наук*. — 2013. — Т. 63, № 3. — С. 78—84.
194. Image quality assessment for video stream recognition systems / T. Chernov [et al.] // *Tenth International Conference on Machine Vision*. — International Society for Optics, Photonics. 2018. — in press.
195. *Chernov, T. S.* A method of periodic pattern localization on document images / T. S. Chernov, D. P. Nikolaev, V. M. Kliatskine // *Eighth International Conference on Machine Vision*. Vol. 9875. — International Society for Optics, Photonics. 2015. — P. 987504.

196. *Chernov, T. S.* A Method Of Periodic Pattern Detection On Document Images / T. S. Chernov, D. P. Nikolaev, V. M. Kliatskine // 29th European Conference on Modelling and Simulation. — 2015. — P. 506—510.
197. Snapscreen: TV-stream frame search with projectively distorted and noisy query / N. Skoryukina [et al.] // Ninth International Conference on Machine Vision. Vol. 10341. — International Society for Optics, Photonics. 2017. — P. 103410.
198. Контроль и коррекция входного сигнала двухканального рентгенографического сепаратора / Д. Николаев [и др.] // Информационные технологии и системы, сборник трудов. — 2014. — С. 202—209.
199. *Чернов, Т.* Метод поиска периодических фоновых элементов на изображении документа / Т. Чернов, Д. Николаев, В. Кляцкин // Информационные технологии и системы, сборник трудов. — 2015. — С. 400—412.
200. Segments Graph-Based Approach for Document Capture in a Smartphone Video Stream / A. Zhukovsky [et al.] // 14th IAPR International Conference on Document Analysis and Recognition. Vol. 1. — IEEE, 2017. — P. 337—342.
201. Система распознавания изображений символов на основе обучающей выборки: патент РФ на полезную модель №161580 / В.Л. Арлазаров, К.Б. Булатов, Д.А. Ильин, Д.П. Николаев, Т.С. Чернов, А.В. Шешкус, опубл. 27.04.2016 по заявке №2015148233/08 от 10.11.2015.
202. Способ детектирования голографических элементов в видеопотоке: патент РФ на изобретение №2644513 / В.В. Арлазаров, Д.П. Николаев, Н.С. Скорюкина, Т.С. Чернов, опубл. 12.02.2018 по заявке №2017106048 от 27.02.2017.
203. Программа для распознавания идентификационных карт личности «Smart IDReader»: свидетельство о государственной регистрации программы для ЭВМ №2016616961 / В.В. Арлазаров, Д.П. Николаев, С.А. Усилин, К.Б. Булатов, Т.С. Чернов, Д.Г. Слугин, Д.А. Ильин, П.В. Безматерных, А.А. Муковозов, Е.Е. Лимонова, опубл. 22.06.2016 по заявке №2016612014 от 10.03.2016.

204. Библиотека для распознавания в видеопотоке паспорта гражданина Российской Федерации «Smart PassportReader»: свидетельство о государственной регистрации программы для ЭВМ №2015616071 / В.В. Арлазаров, К.Б. Булатов, Д.А. Ильин, А.В. Куроптев, Д.П. Николаев, Д.В. Полевой, С.А. Усилин, И.А. Фараджев, Т.С. Чернов, опубли. 29.05.2015 по заявке №2015612880 от 10.04.2015.
205. Библиотека для создания компактных цифровых копий документа «Smart PDF/A»: свидетельство о государственной регистрации программы для ЭВМ №2015615714 / В.В. Арлазаров, А.Е. Жуковский, А.В. Куроптев, Д.П. Николаев, С.А. Усилин, Т.С. Чернов, опубли. 22.05.2015 по заявке №2015612889 от 10.04.2015.
206. Библиотека для распознавания машиночитаемых строк в видеопотоке «Smart 3D OCR MRZ»: свидетельство о государственной регистрации программы для ЭВМ №2015615712 / В.В. Арлазаров, К.Б. Булатов, А.Г. Волков, Д.А. Ильин, А.В. Куроптев, А.Е. Марченко, Д.П. Николаев, Д.В. Полевой, Т.С. Чернов, Ю.С. Чернышова, опубли. 22.05.2015 по заявке №201561288 от 10.04.2015.