

ЯРОСЛАВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ
им. П.Г. ДЕМИДОВА

На правах рукописи



Илья Сергеевич Ненахов

**Неэталонная оценка качества телевизионных
изображений на основе локальных бинарных шаблонов
и алгоритмов машинного обучения**

ДИССЕРТАЦИЯ

на соискание ученой степени кандидата технических наук
по специальности 05.12.04 Радиотехника, в том числе системы
и устройства телевидения

Научный руководитель – д.т.н., профессор Брюханов Юрий Александрович

Ярославль – 2016

ОГЛАВЛЕНИЕ

ГЛАВА 1.	ВВЕДЕНИЕ	4
	ОЦЕНКА КАЧЕСТВА ЦИФРОВЫХ ИЗОБРАЖЕНИЙ	12
1.1.	Вводные замечания	12
1.2.	Эталонные критерии оценки качества изображений	16
1.3.	Неэталонные критерии оценки качества изображений	20
1.4.	Методы машинного обучения	22
1.4.1.	Машина опорных векторов	22
1.4.2.	Бустинг (AdaBoost).....	24
1.4.3.	Градиентный бустинг	25
1.4.4.	Рандомизированные деревья	26
1.4.5.	Сверточные нейронные сети	28
1.4.5.1.	Слои сверточной нейронной сети	30
1.4.5.2.	Методы настройки параметров сверточной нейронной сети	32
1.5.	База тестовых изображений LIVE	34
1.6.	База тестовых изображений TID2013	36
1.7.	Метрики, используемые для тестирования точности алгоритмов неэталонной оценки качества изображений	39
1.8.	Краткие выводы	40
ГЛАВА 2.	РАЗРАБОТКА АЛГОРИТМОВ НЕЭТАЛОННОЙ ОЦЕНКИ КАЧЕСТВА ИЗОБРАЖЕНИЙ	42
2.1.	Вводные замечания	42
2.2.	Этапы работы алгоритма НОК ЛБШ.....	43
2.2.1.	LOG-фильтр	43
2.2.2.	Вычисление ЛБШ-шаблонов и построение гистограммы	43
2.2.3.	Классификация на основе машины опорных векторов	44
2.3.	Результаты моделирования алгоритма на тестовой базе изображений LIVE.....	44
2.4.	Локальные бинарные шаблоны	45
2.5.	Описание алгоритма неэталонной оценки качества на основе локальных бинарных шаблонов	49
2.6.	Визуализация ЛБШ-признаков для различных типов искажений ..	50
2.7.	Неэталонная оценка качества в пространственной области	56
2.7.1.	Описание алгоритма.....	56
2.7.2.	Результаты моделирования.....	58
2.8.	Краткие выводы	62
ГЛАВА 3.	РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ НЕЭТАЛОННЫХ МЕТРИК ОЦЕНКИ КАЧЕСТВА ИЗОБРАЖЕНИЙ	63
3.1.	Вводные замечания	63
3.2.	Тестирование различных видов классификаторов для неэталонного алгоритма оценки качества.....	64

3.2.1.	Исследование для одномасштабных локальных бинарных шаблонов	64
3.2.2.	Исследование для многомасштабных локальных бинарных шаблонов	66
3.2.3.	Сравнительный анализ времени обучения.....	67
3.3.	Тестирование модифицированного алгоритма неэталонной оценки качества	69
3.3.1.	Расчет ЛБШ-признаков для различных изображений	69
3.3.2.	Вычисление смеси признаков	71
3.3.3.	Итоговый модифицированный алгоритм.....	73
3.4.	Тестирование времени работы алгоритмов оценки качества	75
3.5.	Суммарное тестирование алгоритмов на базе LIVE.....	76
3.6.	Суммарное тестирование алгоритмов на базе TID2013	77
3.7.	Краткие выводы	79
ГЛАВА 4.	АНАЛИЗ ПРАКТИЧЕСКИХ ПРИМЕНЕНИЙ НЕЭТАЛОННЫХ АЛГОРИТМОВ ОЦЕНКИ КАЧЕСТВА ДЛЯ ТЕЛЕВИЗИОННЫХ ИЗОБРАЖЕНИЙ	82
4.1.	Вводные замечания	82
4.2.	Оценка качества видеопоследовательностей	82
4.2.1.	Покадровая оценка качества видео.....	83
4.2.2.	Неэталонное оценивание качества видеоданных относительно субъективных экспертных оценок	89
4.3.	Оценка качества изображений лиц в задаче идентификации при использовании алгоритма распознавания на основе сверточной нейронной сети	90
4.3.1.	Описание системы распознавания	90
4.3.2.	Детектор лиц	91
4.3.3.	Алгоритм определения особых точек	91
4.3.4.	Получение представления лица с помощью глубокой сверточной сети.....	93
4.3.5.	Классификация	95
4.3.6.	Анализ точности распознавания лиц при наличии искажений на изображении.....	95
4.4.	Выбор наилучшего лица для задач биометрической идентификации на основе алгоритмов неэталонной оценки качества	100
4.4.1.	Описание используемых метрик.....	103
4.4.2.	Результаты моделирования	105
4.5.	Краткие выводы	109
	ЗАКЛЮЧЕНИЕ	110
	СПИСОК ЛИТЕРАТУРЫ	113
	ПРИЛОЖЕНИЕ 1. ТЕСТОВАЯ БАЗА RETRIEVED.....	125
	ПРИЛОЖЕНИЕ 2. АКТЫ ВНЕДРЕНИЯ РЕЗУЛЬТАТОВ РАБОТЫ	127

ВВЕДЕНИЕ

Актуальность темы. Цифровая обработка визуальной информации в настоящее время охватывает различные виды инфокоммуникационных приложений и расширяет их спектр. Сюда относятся как традиционные приложения (вещательное, промышленное, охранное телевидение), так и относительно новые приложения (видеоконференцсвязь, техническое зрение, цифровое кино, телевидение высокой и сверхвысокой четкости, 3D-телевидение) [1–5].

В соответствии с современной классификацией, предложенной В.П. Дворковичем и А.В. Дворковичем цифровая обработка телевизионных и компьютерных изображений включает в себя следующие основные области [6]:

- коррекция изображения, их «препарирование», т. е. сознательное разделение на части цифровыми средствами, видоизменение этих частей и их обратная «сборка»;
- оценка параметров изображений с целью контроля качества их передачи и приема;
- преобразование и кодирование изображений для хранения и передачи по каналам связи;
- компьютерная графика, а также визуализация информации, т. е. представление массивов данных в виде различных изображений, что очень эффективно, так как облегчает решение многих задач, сложных именно своей абстрактностью.

Актуальность разработки алгоритмов обработки визуальной информации объясняется происходящим в настоящее время в России переходом к цифровому телевещанию на основе стандартов DVB первого и второго поколений.

Программа по переводу эфирного телевидения на цифровые технологии началась в 2009 году. С 2012 по 2016 годы федеральное государственное унитарное предприятие «Российская телевизионная и радиовещательная сеть» (РТРС) активно ведет строительство объектов телерадиовещания. Сегодня построено 4075 объектов первого мультиплекса и 2612 — второго. Первым мультиплексом охвачено 91,55% населения страны, вторым — 63,62%. Построены 83 центра формирования мультиплексов для организации регионального вещания.

Развитие цифрового эфирного вещания в РФ сделает доступными 98,1% населения не менее 20 телеканалов в хорошем цифровом качестве. При этом треть населения сегодня подключена к кабельным ТВ-сетям, более четверти россиян — к спутниковому телевидению, которое позволяет принимать большее количество цифровых каналов, в том числе и в формате высокой четкости (HDTV). На сегодняшний день более чем у 60% российских семей есть возможность бесплатно принимать 20 эфирных телевизионных каналов в цифровом формате. Телевизионное изображение высокого качества (телевидение высокой четкости) имеют возможность получать более 40% населения [7].

Значительный вклад в данную область науки и техники внесли как отечественные ученые М.И. Кривошеев, В.А. Сойфер, М.К. Чобану, А.С. Крылов, Д.С. Ватолин, А.С. Конушин, Б.А. Алпатов, Ю.С. Бехтин, Б.В. Костров, В.Ю. Волков, Е.П. Петров, И.С. Трубин, Е.В. Медведева, М.Н. Фаворская, Ю.С. Радченко, так и зарубежные – S. Mitra, R. Gonzalez, R. Woods, Z. Wang, A. Bovik, H. Sheikh, E. Simoncelli, J. Astola, K. Egiazarian, R. Szeliski, R. Lukac и др. [1–29, 37-67].

Алгоритмы обработки, сжатия и кодирования мультимедийной информации неизбежно вносят искажения различного типа, что делает актуальным вопрос об оценке параметров этих искажений и объективного оценивания качества искаженного изображения. На современном этапе

развития все методы оценки качества классифицируются на две большие категории:

1. Субъективные критерии оценки качества изображений. Такая оценка в данном случае осуществляется человеком (оценки MOS по рекомендации ITU-R BT.500-11).
2. Объективные критерии оценки качества изображений. Здесь оценка производится алгоритмически. Такой алгоритм может использовать оригинал изображения для оценки качества, то есть быть эталонным, или может определять качество без наличия неискаженной копии изображения, то есть быть неэталонным.

В связи с необходимостью к унификации исследований в данной области в последние годы были предложены две общедоступные тестовые базы изображений с экспертными оценками MOS: LIVE (LIVE Image Quality Assessment Database, University of Texas at Austin, USA) и TID2008/2013 (Tampere Image Database, Tampere, Finland). Задача неэталонной оценки качества изображений вызывает в последние годы повышенный интерес исследователей. Большая часть предлагаемых неэталонных алгоритмов разработана для конкретных типов априори известных искажений в изображении (блочность, размытие и др.), что не всегда удобно для использования в практических приложениях [11-16, 54-59].

Исследовательские лаборатории в области оценки качества в задачах цифровой обработки изображений и речевых сигналов созданы во многих крупных университетах мира. Среди них стоит выделить MIT, Stanford University, UCLA, University of Texas at Austin, Tampere University of Technology, ведущие университеты Японии и Сингапура. Практическими разработками занимаются как крупные глобальные компании: Intel, Panasonic, Samsung, так и мелкие технологичные компании-стартапы, созданные при ведущих университетах.

Обзор существующих неэталонных критериев оценки качества изображений и видеопоследовательностей показал, что универсальных

решений в данной области на настоящий момент не существует [1, 3, 8, 19, 21, 27, 28, 39–62]. Перспективным подходом представляется решение данной проблемы на основе современных алгоритмов машинного обучения. Для этого следует, во-первых, определить признаки, отвечающие за качество изображений, во-вторых, определить эффективный классификатор, способный разделять изображения разного качества на классы.

Другой важной областью, где решение задачи неэталонной оценки качества изображений представляет большой практический интерес, является биометрическая идентификация человека по изображению его лица. Существует несколько стандартов, определяющих качество изображений лиц, – ISO/IEC 19794-5, ICAO 9303 [84,85], которые содержат описание характеристик, влияющих на принятие решения о пригодности изображения для систем автоматического распознавания. Все стандартизованные характеристики можно разделить на два класса: текстурные (резкость, контрастность, освещенность, степень компрессии, наличие прочих искажений) и относящиеся непосредственно к признакам лица (симметричность, поза, поворот, видимость глаз, наличие бликов или теней на лице). Для их автоматического определения в научной литературе описаны следующие методы – определение позы [74], не фронтальности освещения и поворота [80].

Еще большую актуальность данному вопросу придает планируемая в РФ разработка национальной биометрической платформы и решение других перспективных задач на рынках Safenet Национальной Технологической Инициативы (НТИ).

Все вышеперечисленное доказывает, что разработка неэталонных критериев оценки качества представляет собой актуальную научно-техническую задачу как для области телевидения, так и для ряда смежных научно-технических областей.

Основной целью работы является расширение возможностей контроля качества систем вещательного и прикладного телевидения.

В соответствии с указанной целью в рамках диссертационной работы были поставлены и решены следующие **задачи**:

- проведение исследований в части выбора признаков, характеризующих качество изображения в пространственной области;
- проведение исследований по выбору алгоритма машинного обучения для классификации изображений различного качества;
- разработка и анализ алгоритма неэталонной оценки качества на основе выбранных признаков и алгоритмов классификации;
- проведение исследований по оценке корреляции между разработанным неэталонным алгоритмом и результатами субъективной экспертизы оценки качества на специализированных базах изображений LIVE и TID2013;
- проведение анализа использования предложенного алгоритма неэталонной оценки качества изображений в ряде задач вещательного и прикладного телевидения и в задачах биометрической идентификации/верификации личности.

Объектом исследования являются радиотехнические системы фильтрации и сжатия визуальной информации, биометрической идентификации и прикладного телевидения.

Предметом исследования являются неэталонные алгоритмы оценки качества телевизионных изображений на основе локальных бинарных шаблонов и методов машинного обучения.

Методы исследования. При решении поставленных задач использовались современные методы цифровой обработки изображений, машинного обучения, теории вероятностей и математической статистики. Для практической реализации алгоритмов применялись современные численные методы и язык программирования python (библиотека scikit-learn).

Достоверность полученных научных результатов обусловлена применением адекватного математического аппарата, подтверждается их

согласованностью с результатами проведенного компьютерного моделирования и сопоставлением ряда полученных результатов с научно-техническими данными, полученными на унифицированных тестовых базах изображений LIVE и TID2013.

Научная новизна. В рамках работы получены следующие новые научные результаты:

- Предложено использование локальных бинарных шаблонов в качестве признаков в задаче неэталонной оценки качества изображений.
- Предложено использование рандомизированных деревьев в качестве классификатора изображений различного качества.
- Разработан алгоритм неэталонной оценки качества телевизионных изображений на основе локальных бинарных шаблонов и методов машинного обучения.
- Разработан модифицированный алгоритм неэталонной оценки качества с использованием статистики естественных изображений.

Практическая значимость полученных результатов

1. Неэталонный характер работы предложенных алгоритмов оценки качества телевизионных изображений позволяет использовать их в широком спектре практических приложений, где, как правило, исходное (эталонное) изображение не доступно. При этом достигаемое значение коэффициента корреляции Спирмена 0,93–0,96 (база LIVE), 0,6–0,92 (база TID2013) соответствует используемым на практике эталонным критериям.
2. Разработанный алгоритм является вычислительно эффективным, так как он работает в пространственной области и использует вычислительно не затратный аппарат локальных бинарных шаблонов.
3. Проведенный анализ использования предложенного алгоритма в ряде прикладных задач показывает, что он может быть использован в системах обработки и передачи визуальной информации, прикладного телевидения, биометрической идентификации личности по

изображению лица и в других задачах цифровой обработки изображений и технического зрения.

4. На основании проведенного анализа времени работы модифицированного алгоритма можно сделать вывод о том, что он применим для задач неэталонной оценки качества изображений в офлайн-приложениях.

Основные научные положения, выносимые на защиту

1. Алгоритм неэталонной оценки качества телевизионных изображений на основе локальных бинарных шаблонов и алгоритма рандомизированных деревьев.
2. Модифицированный алгоритм неэталонной оценки качества телевизионных изображений с использованием статистики естественных изображений.
3. Результаты сопоставления предложенных объективных неэталонных оценок качества телевизионных изображений со средними экспертными оценками.
4. Результаты анализа работы неэталонного алгоритма оценки качества изображений в условиях искажений, характерных для задач биометрической идентификации/верификации личности по изображению лица.

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

- 18-я международная конференция «Цифровая обработка сигналов и ее применение», DSPA (Москва, 2016);
- 70-я и 71-я международные конференции «Радиоэлектронные устройства и системы для инфокоммуникационных технологий», РЭУС (Москва, 2015–2016);
- 4-я всероссийская конференция «Радиоэлектронные средства получения, обработки и визуализации информации», РСПОВИ (Москва, 2015);

- 13-я международная научно-техническая конференция «Телевидение: передача и обработка изображений» (Санкт-Петербург, 2016);
- 17-th conference of FRUCT association (Yaroslavl, 2015);

Публикации. По теме диссертации опубликовано 11 научных работ, из них 2 статьи в журналах, рекомендованных ВАК, и 9 докладов на научных конференциях.

Структура и объем работы. Диссертация состоит из введения, четырех глав, заключения, списка использованных источников, содержащего 111 наименование и приложения. Она изложена на 132 страницах машинописного текста, содержит 30 рисунков и 30 таблиц.

Личный вклад. Основные научные положения, выносимые на защиту, предложены и реализованы лично автором. Программная реализация разработанных алгоритмов выполнена с участием автора в составе научной группы в рамках работ, проводимых на кафедре динамики электронных систем ЯрГУ.

Благодарности. Автор выражает слова глубокой благодарности своему научному руководителю профессору Брюханову Юрию Александровичу за помощь на всех этапах выполнения данной работы. Отдельная благодарность преподавателям ЯрГУ В.В. Хрящеву, А.Л. Приорову, В.А. Волохову, работы которых оказали значительное влияние на формирование взглядов автора в данном научном направлении. Эти взгляды формировались также в совместной работе с коллегами по лаборатории «Цифровые цепи и сигналы» кафедры динамики электронных систем ЯрГУ, среди которых особенно хочется отметить Е. Павлова, Ю. Лукашевича, А. Тараканова, М. Голубева, А. Шемякова, О. Балашову, О. Степанову, А. Бажаниновского.

ГЛАВА 1. ОЦЕНКА КАЧЕСТВА ЦИФРОВЫХ ИЗОБРАЖЕНИЙ

1.1. Вводные замечания

Оценка качества - важнейшая часть процесса исследования при разработке линейных и нелинейных систем фильтрации/восстановления изображений [1, 8, 9, 18-29].

На рис. 1.1 приведена классификация современных методов оценки качества изображений.

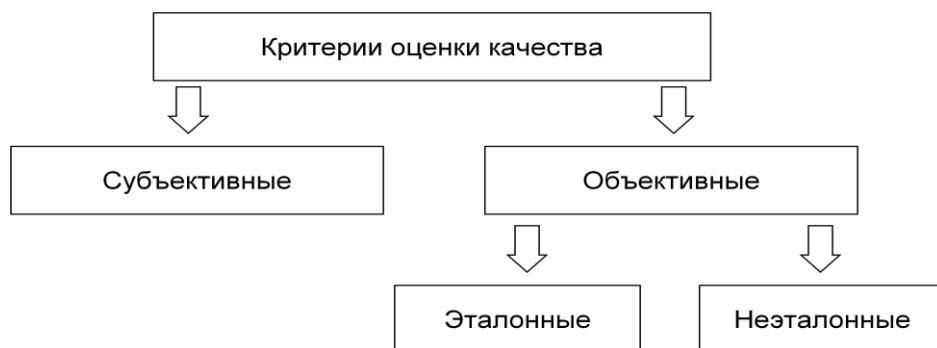


Рис. 1.1. Классификация современных методов оценки качества изображений

Субъективные критерии качества

Так как человек является приемником в большинстве приложений цифровой обработки изображений, то субъективный визуальный анализ является самым точным методом оценки качества изображений. В ходе субъективного анализа решение о качестве принимается группой экспертов, а окончательная оценка – среднее арифметическое всех оценок членов группы (MOS). Согласно [38] для проведения таких тестов необходима группа не менее 15 человек.

Экспертные оценки бывают двух видов: абсолютные и сравнительные [1, 3]. Отличаются они тем, что в случае сравнительной оценки наблюдателю вместе с тестируемым изображением предлагается

эталонное не искаженное изображение. В случае абсолютной оценки наблюдатель должен оценить качество изображения по какой-то заранее определенной шкале.

На результат субъективных экспертных оценок влияют как характер рассматриваемых изображений, так и различные внешние факторы, такие как дистанция до изображения, характеристики дисплея и освещение. Оценка качества так же может измениться в зависимости от того видел или нет наблюдатель изображение заранее. В незнакомом изображении искажения могут быть не замечены, пока наблюдателю на них не укажут. Для знакомого изображения оценки будут строже, так как у наблюдателя имеется представление о структуре изображения.

Для того что бы результаты оценки качества могли быть воспроизводимыми международный союз электросвязи принял нормативный документ ITU-R BT.500-11 [38], который содержит инструкции по проведению субъективных тестов, тестовые материалы, а также шкалы оценок качества изображений, выбор которых зависит от цели исследования. Стоит отметить, что в зависимости от метода исследования, оценка качества может принимать, как дискретные значения от 1 до 5 (1 плохое качество), так и непрерывные от 0-100.

Популярной на данный момент является метрика DMOS (difference mean opinion score), которая может быть записана в виде [37]:

$$d_{i,j} = r_{i,ref}(j) - r_{i,j},$$

$$DMOS_{i,j} = \frac{d_{i,j} - \bar{d}_i}{\sigma_i},$$

$$DMOS_j = \overline{DMOS_{i,j}},$$

где $r_{i,j}$ это оценка i -го эксперта для j изображения, $r_{i,ref}(j)$ обозначает оценку качества, которую выставил i -ый эксперт эталонному изображению, \bar{d}_i – среднее по всем изображениям значение оценок, которое выставляет i -ый эксперт, а σ_i среднее квадратическое отклонение. Метрика DMOS

принимает значения от 0-100, чем большее значение, тем хуже качество тестируемого изображения.

Несмотря на свою точность, субъективные экспертизы имеют серьезные недостатки [37]:

- дороговизна и низкая скорость работы;
- результаты зависят от точности соблюдения экспериментальных условий и эмоционального состояния экспертов.

Объективные критерии качества

Под объективными критериями оценки качества понимаются критерии, расчет которых происходит автоматически, без участия группы экспертов [1, 3, 8, 39-50]. Задача таких алгоритмов оценивать качество изображений, так как его оценивает человек, то есть выставленные автоматически оценки должны соответствовать субъективным метрикам MOS или DMOS для тех же изображений. Для установления степени соответствия между объективными и субъективными оценками используют различные метрики: коэффициент ранговой корреляции Спирмена, коэффициент линейной корреляции, коэффициент корреляции Кенделла и величину среднеквадратичной ошибки. Чем выше, значения коэффициентов корреляции и чем меньше величина среднеквадратичной ошибки, тем точнее работает рассматриваемый объективный алгоритм.

Алгоритмы объективной оценки качества можно разделить на две большие группы. Первая группа – эталонные алгоритмы. Алгоритмы из этой группы помимо искаженного изображения требуют для своей работы неискаженное эталонное изображение, то есть алгоритмы измеряют качество относительно оригинала. Вторая группа – неэталонные алгоритмы, которые работают только с искаженным изображением. Такие алгоритмы более востребованы на практике, так как зачастую в «реальных» приложениях эталонное изображение не известно. Иногда в отдельную группу выносят так называемые псевдоэталонные алгоритмы, которые для своей работы требуют

наличия не самого эталонного изображения, а некоторой априорной информации о нем (например, число потерянных пакетов в канале передаче, тип кодека) (рис 1.2).

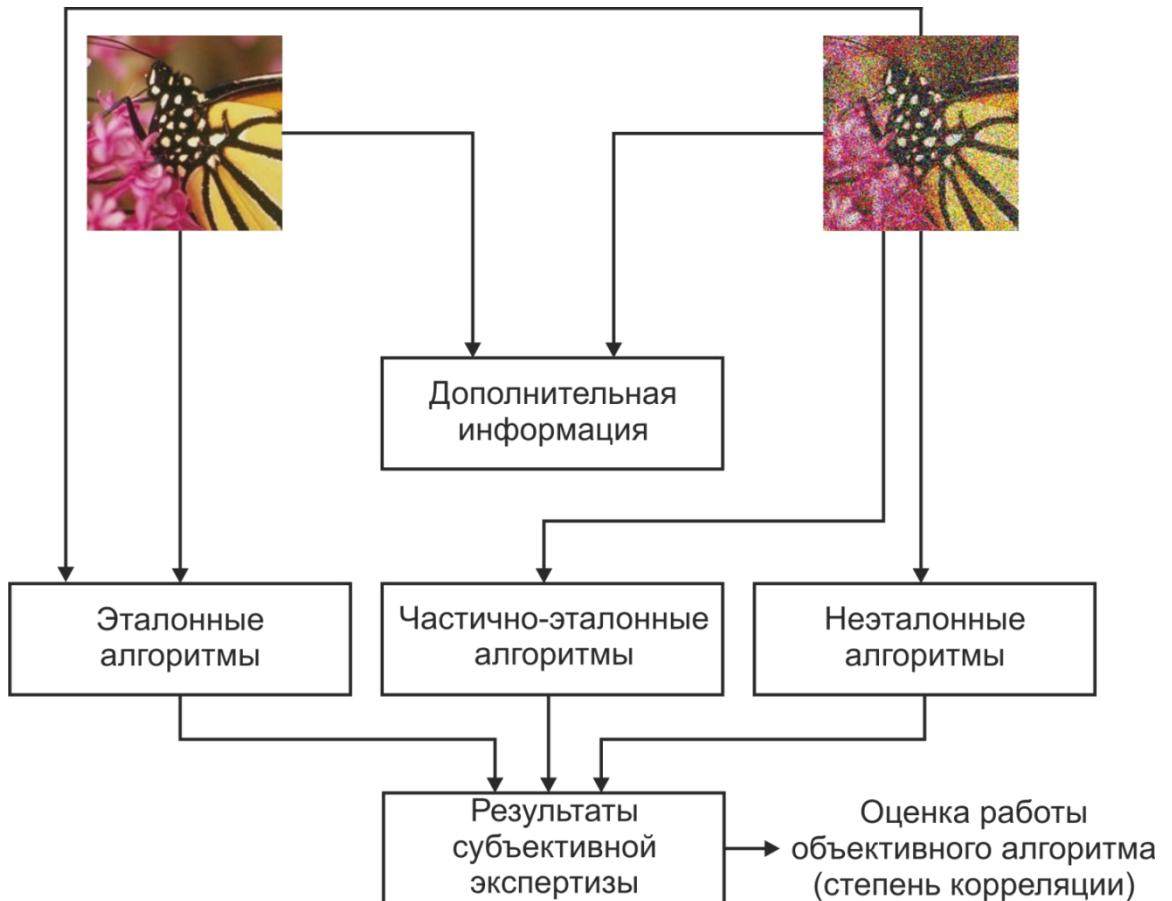


Рис. 1.2. Различные подходы к объективной оценке качества изображений

На данный момент существуют алгоритмы объективной оценки качества, основанные на статистике естественных изображений [41, 61, 62], моделирующие различные особенности зрительной системы человека [39, 40, 46], а так же осуществляющие извлечение определенных параметров изображения [42, 48, 49].

1.2. Эталонные критерии оценки качества изображений

Пиковое отношение сигнал/шум

Пиковое отношение сигнал/шум (ПОСШ) – наиболее популярная объективная метрика качества изображений. Для двух цифровых изображений с 256 возможными значениями яркости и числом пикселей N , ПОСШ можно записать как [1, 3]:

$$POS\!S\!H = 20 \log_{10} \frac{255}{\sqrt{CKO}},$$

$$CKO = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2,$$

где x_i и y_i – значения пикселей эталонного и восстановленного изображений соответственно, а СКО – среднеквадратичная ошибка.

Основными преимуществами данного критерия являются:

1. Простота расчета
2. Отсутствие параметров
3. Физический смысл – энергия сигнала ошибки
4. Возможность использования в задачах оптимизации благодаря свойствам симметричности и дифференцируемости.

При всех достоинствах данного критерия, у него можно выделить следующие недостатки:

1. Отсутствие абсолютного значения
2. Значение метрики не зависит от пространственных зависимостей между исходным и искаженным сигналом.
3. Значение метрики зависит только от величины сигнала ошибки
4. Все пиксели эталонного и входного изображения вносят одинаковый вклад в значение метрики, что не согласуется с принципами работы ЗСЧ [9,10]

Коэффициент структурного подобия

Естественные изображения структурированы, между соседними пикселями существуют сильные взаимосвязи, которые содержит важную информацию об объектах на визуальной сцене. Зрительная система человека адаптирована к извлечению такой структуры, таким образом, структурное подобие двух изображений эталона и искаженного, должно быть хорошей оценкой их взаимного качества [3].

Коэффициент структурного подобия (КСП, SSIM) это алгоритм, воплощающий идею структурного подобия изображений в пространственной области. На вход алгоритма подаются изображения x и y , одно из которых эталонное. На выходе получается индекс структурного подобия, который служит искомой оценкой качества. В ходе работы алгоритма КСП происходят сравнения яркости, контраста и структуры изображений (рис. 1.3).

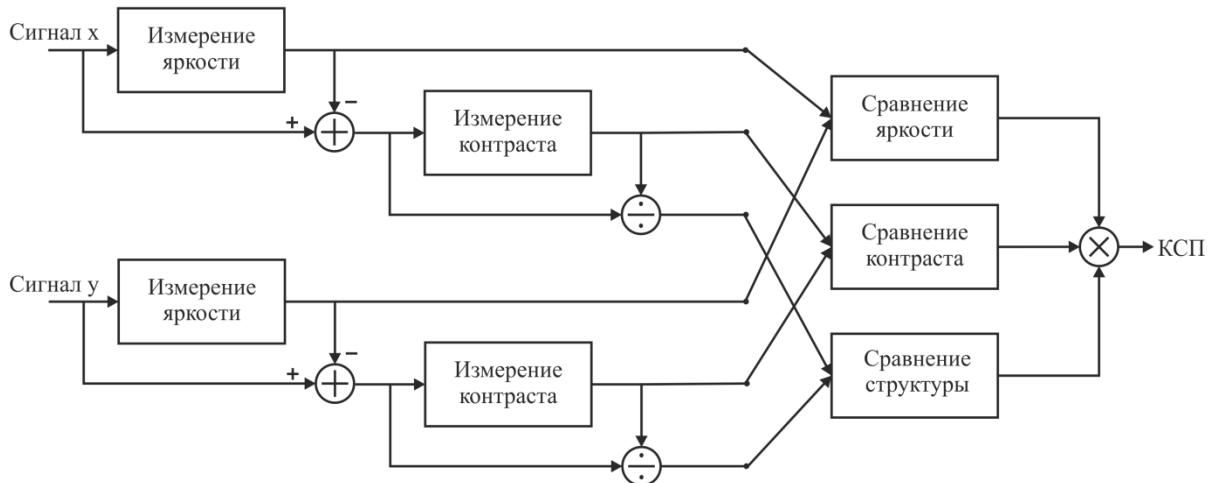


Рис. 1.3. Блок-схема вычисления КСП

Стандартный алгоритм вычисления КСП не является универсальным и требует предварительной обработки изображения. Стоит отметить, что сравниваемые параметры рассчитываются не для всего изображения, а лишь для некоторой области – окна, внутри которого изображение можно считать стационарным и вычислять статистические характеристики. Рассмотрим процесс вычисления индекса КСП внутри

одного окна, состоящего из N пикселей. Пусть x^k и y^k области изображений соответствующие окну с номером k . Для каждого изображения рассчитываются значения средней яркости:

$$\mu_x = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i^k,$$

$$\mu_y = \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i^k,$$

и задается функция сравнения яркости:

$$I(x^k, y^k) = I(\mu_x, \mu_y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1},$$

где C_1 положительная константа необходимая для того, чтобы избежать случаев деления на нуль. Для оценки контраста изображения используется стандартное отклонение, которое для областей изображений x^k и y^k записывается следующим образом:

$$\sigma_x = \left(\frac{1}{n-1} \sum_{i=1}^N (x_i^k - \mu_x)^2 \right)^{\frac{1}{2}},$$

$$\sigma_y = \left(\frac{1}{n-1} \sum_{i=1}^N (y_i^k - \mu_y)^2 \right)^{\frac{1}{2}}.$$

Функция сравнения контраста задается как:

$$c(x^k, y^k) = c(\sigma_x, \sigma_y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2},$$

где C_2 небольшая положительная константа. Функция сравнения структуры задается формулой:

$$s(x^k, y^k) = s\left(\frac{x^k - \mu_x}{\sigma_x}, \frac{y^k - \mu_y}{\sigma_y}\right) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3},$$

$$\sigma_{xy} = \frac{1}{N-1} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y),$$

где C_3 так же небольшая положительная константа.

Коэффициент структурного подобия рассчитывается, как произведение:

$$\text{КСП}^k = [I(x^k, y^k)]^\alpha [c(x^k, y^k)]^\beta [s(x^k, y^k)]^\gamma,$$

где $\alpha > 0, \beta > 0, \gamma > 0$. Для уменьшения числа параметров $\alpha = \beta = \gamma = 1$, а $C_3 = C_2/2$. Принимая во внимание формулы, описанные выше, можно записать результирующую формулу индекса структурного подобия для k -го окна:

$$\text{КСП}^k = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}.$$

Для всего изображения процесс нахождения КСП будет сводиться к следующему: «скользящее окно», начиная из верхнего левого угла, проходит пиксель за пиксели по всем строкам и столбцам этого изображения. На каждом шаге вычисляется значение КСП^k для текущего положения окна. Коэффициент структурного подобия, характеризующий качество изображения в целом, будет определяться как среднее арифметическое значений КСП^k рассчитанных на протяжении всего изображения, т.е. [17, 45]:

$$\text{КСП} = \frac{1}{K} \sum_{k=1}^K \text{КСП}^k.$$

Восприятие изображения, зависит от расстояния до наблюдателя и от особенностей его зрительной системы. Что бы учесть этот факт, предложена модификация алгоритма КСП – многомасштабный коэффициент структурного подобия МКСП (MSSIM) [46]. Модификация состоит в том, что индекс КСП считается не только для исходного изображения, но и для изображения, полученного из исходного путем децимации с коэффициентом 2. В общем случае, алгоритм МКСП предполагает, что исходное изображение подвергается M этапам децимации с

коэффициентом 2, причем для каждого такого изображения вычисляются $c(x^k, y^k), s(x^k, y^k)$.

$$\text{МКСП}^k = [I_M(x^k, y^k)]^{\alpha_M} \prod_{j=1}^M [c_j(x^k, y^k)]^{\beta_j} [s_j(x^k, y^k)]^{\gamma_j},$$

$$\text{МКСП} = \frac{1}{K} \sum_{k=1}^K \text{МКСП}^k,$$

где K – число окон, а $\beta_j > 0, \gamma_j > 0, \alpha_M > 0$ – положительные константы, регулирующие вклад параметров алгоритма КСП в результирующий индекс качества на разных масштабах.

1.3. Неэталонные критерии оценки качества изображений

Неэталонные алгоритмы оценки качества подразделяются на две группы: алгоритмы для работы, которых необходима априорная информация о типе искажения и универсальные алгоритмы, которые работают без использования дополнительной информации об искажениях (рис 1.4).



Рис. 1.4. Классификация неэталонных алгоритмов

В первую группу входят алгоритмы, измеряющие «силу» определенного типа искажения: размытия [39], артефактов при сжатии по стандарту JPEG [56] и JPEG2000 [55] и др. повреждений. Большинство таких

алгоритмов во время работы используют процедуру выделения границ на изображении. В случае изображений, сжатых по стандарту JPEG2000, определяется величина звона и размытия этих границ. В случае JPEG измеряется величина блочности, которая имеет периодическую структуру на границах.

В последнее время особое внимание уделяется синтезу неэталонных критериев оценки качества из второй группы, работа которых не ограничивается априорным видом искажений. На практике информация о типе искажения доступна не всегда, поэтому актуальной является задача разработки именно универсальных алгоритмов. К таким алгоритмам относятся метрики, основанные на статистике естественных изображений. Такая статистика может быть рассчитана в пространственной области [61, 62], в пространстве вейвлет-преобразования [40] и в пространстве коэффициентов дискретного косинусного преобразования [41]. Другой подход в разработке универсальных алгоритмов неэталонной оценки качества состоит в определении параметров искаженного изображения, которые непосредственно отражаются на его качестве [42, 49]. Такими параметрами могут быть энтропия искаженного изображения [60] или отдельные пиксели изображения [42].

Отличительной особенностью алгоритмов, которые не учитывают априорную информацию, является необходимость процедуры предварительного обучения. Обучение универсальных алгоритмов производится на тестовой базе, содержащей изображения с различными типами искажений и соответствующие субъективные оценки качества. Популярными тестовыми базами изображений являются LIVE [65], TID2013 [67]. В работе универсальных алгоритмов можно выделить 2 этапа: извлечение качественных признаков из изображения и сопоставление этих признаков с соответствующей субъективной оценкой. Обучение необходимо, что бы найти функцию, связывающую значения признаков и оценку качества на выходе. Отыскание такой функции – задача регрессии, решается при

помощи методов машинного обучения. Для решения этой задачи в алгоритмах неэталонной оценки качества используют машину опорных векторов, рандомизированные деревья и искусственные нейронные сети. После процедуры обучения алгоритм потенциально способен оценить качество абсолютно любого изображения при условии, что рассматриваемый тип искажений присутствовал в базе.

1.4. Методы машинного обучения

1.4.1. Машина опорных векторов

Метод машинного обучения машина опорных векторов (МОП) обладает некоторыми замечательными свойствами [93,94]. Во-первых, обучение МОП сводится к задаче квадратичного программирования, имеющей единственное решение, которое вычисляется достаточно эффективно даже на выборках в сотни тысяч объектов. Во-вторых, положение оптимальной разделяющей гиперплоскости зависит лишь от небольшой доли обучающих объектов, которые и называются опорными векторами; остальные объекты фактически не задействуются. В-третьих, метод обобщается на случай нелинейных разделяющих поверхностей, за счет введения функции ядра.

Рассмотрим задачу классификации на два непересекающихся класса, в которой объекты описываются l -мерными векторами $X = R^l$. Метки классов – $Y = \{-1, +1\}$ и дана выборка объектов размером N . Для решения задачи используется линейный классификатор

$$h(n) = \text{sign}\left(\sum_{j=1}^l w_j x^j - w_0\right) = \text{sign}(\langle w, x \rangle - w_0).$$

Введем понятие зазора $z(x_i) = y_i(\langle w, x \rangle - w_0)$. Отрицательный зазор говорит о том, что классификатор совершил ошибку, следовательно, для построения точного классификатора требуется, что бы зазор был

положительным и как можно большим. В терминах разделяющей гиперплоскости, ищется такая гиперплоскость, которая максимально далеко отстоит от ближайших к ней точек обоих классов.

Алгоритм $h(x)$ не изменится, если w и w_0 одновременно умножить на положительную константу, для удобства константы подбираются такими что бы выполнялось условие:

$$\min_{i=1,\dots,N} z(x_i) = 1.$$

Пусть x_- и x_+ два объекта из X^N , принадлежащие классам -1 и +1 соответственно и лежащие на границе полосы, разделяющей классы $\{x: -1 \leq \langle w, x \rangle - w_0 \leq 1\}$, тогда ширина полосы есть:

$$\langle (x_+ - x_-), \frac{w}{\|w\|} \rangle = \frac{2}{\|w\|}.$$

Ширина полосы максимальная, когда норма вектора $\|w\|$ минимальна, получаем задачу квадратичного программирования

$$\begin{cases} \langle w, w \rangle \rightarrow \min \\ z(x_i) \geq 1 \end{cases}.$$

Задача обобщается для случая линейно неразделимых классов:

$$\begin{cases} \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^N \xi_i \rightarrow \min_{w, w_0, \xi}, \\ z(x_i) \geq 1 - \xi_i, \quad i = 1, \dots, l, \\ \xi_i \geq 0, \quad i = 1, \dots, l, \end{cases}$$

где $\xi_i \geq 0$ величина ошибки на объектах x_i . Положительная константа C является параметром алгоритма МОВ и позволяет находить компромисс между максимизацией ширины разделяющей полосы и минимизацией суммарной ошибки. Решение данной задачи основано на множителях Лагранжа.

Другим способом, позволяющим разделять линейно-неразделимые множества, является преобразование входных данных с помощью функции ядра. Идея заключается в том, чтобы отобразить исходное пространство в пространство более высокой размерности, в котором, как может оказаться,

множества разделимы. Примером, распространенной функции ядра служит ядро Гаусса:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right),$$

Параметр σ , как и параметр C необходимо подбирать во время обучения алгоритма МОВ.

Для задач регрессии задачу оптимизации машины опорных векторов записывают в виде:

$$\begin{cases} \langle w, w \rangle + C \sum_{i=1}^N (\xi_i + \xi_i^*) \rightarrow \min, \\ (wx_i + b) - y_i \leq \varepsilon + \xi_i, \\ y_i - (wx_i + b) \leq \varepsilon + \xi_i^*, \\ \xi_i^* \geq 0, \end{cases}$$

где ε и C – заранее заданные параметры алгоритма.

1.4.2. Бустинг (AdaBoost)

Метод бустинг (AdaBoost) для задач регрессии представлен в работах [95, 96]. Рассмотрим задачу восстановления регрессии, в которой $\{x_i\}^N$, $x_i \in X$ – множество элементов обучающего множества, а $\{y_i\}^N$, $y_i \in Y$ – множество соответствующих меток, такое, что y_i метка для соответствующего объекта x_i . Изначально каждому объекту тренировочного множества в соответствие ставится вес w_i равный единице.

Работа алгоритма AdaBoost состоит в построении M регрессоров F_m . Этапы работы алгоритма:

1. Вычисляется вероятность $p_i = \frac{w_i}{\sum_{i=1}^M w_i}$ с которой элемент x_i попадет в обучающее множество для заданного простого регрессора F_m . Выбираются N_1 элементов из x .

2. Обучается регрессор F_m на N_1 элементах
3. Все N элементов из тренировочного сета, поступают на вход регрессора F_m , формируя на выходе множество элементов $F(x_i)$
4. Для каждого элемента из тренировочного набора рассчитывается функция потерь $Q_i = \frac{(F(x_i) - y_i)^2}{D^2}$, где $D = \sup |F(x_i) - y_i|$
5. Вычисляется $\bar{Q} = \sum_{i=1}^{N_1} Q_i p_i$
6. Вычисляется $\beta = \frac{\bar{Q}}{1-\bar{Q}}$
7. Происходит переназначение весов $w_i \rightarrow w_i \beta^{1-Q_i}$
8. Предсказание для каждого значения x_i формируется по формуле

$$h_f = \inf \left\{ y \in Y : \sum_{t: h_t \leq y} \log \left(\frac{1}{\beta_t} \right) \geq \frac{1}{2} \sum_t \log \left(\frac{1}{\beta_t} \right) \right\}$$

1.4.3. Градиентный бустинг

Градиентный бустинг представлен в работе [97]. Суть метода заключается в следующем – рассматривается задача классификации для многомерного пространства X с пространством меток Y . Пусть дано множество элементов обучающего множества $\{x_i\}^N$, где $x_i \in X$, и множество соответствующих меток $\{y_i\}^N$, где $y_i \in Y$. Метка y_i метка объекта x_i . Так же дано некоторое семейство базовых алгоритмов H , каждый элемент которого $h(x; a) \in H: X \rightarrow R$ определяется некоторым вектором параметров $a \in A$. Необходимо построить классификатор, как можно более точно предсказывающий метки для объектов из X не входящих в тренировочный набор. Классификатор ищется в виде композиции:

$$F_M(x) = \sum_{m=1}^M b_m h(x; a_m), \quad b_m \in R, \quad a_m \in A.$$

Поиск оптимальных параметров $\{a_m, b_m\}^M$ очень трудоемкая задача, для её приближенного решения используется жадный алгоритм – на каждом шаге добавления классификатора в сумму, классификатор выбирается наиболее оптимальным из возможных. Предположим, что уже построен

классификатор F_{m-1} длины $m - 1$. Таким образом, задача сводится к поиску пары наиболее оптимальных параметром a_m, b_m :

$$F_m(x) = F_{m-1}(x) + b_m h(x; a_m), \quad b_m \in R, a_m \in A.$$

Критерием оптимальности является минимум функции потерь $Q(y_i, F_m(x_i))$, $i = 1, \dots, N$, значения которой показывают, насколько «сильно» предсказанный класс $F_m(x_i)$ отличается от правильного класса y_i . Вводится понятие функционала ошибки:

$$\bar{Q} = \sum_{i=1}^N Q(y_i, F_m(x_i)) \rightarrow \min.$$

Градиент функционала ошибки задается формулой:

$$\nabla \bar{Q} = \left[\frac{\partial \bar{Q}}{\partial F_{m-1}}(x_i) \right]_{i=1}^N = \sum_{i=1}^N \left[\frac{\partial Q(y_i, F_{m-1})}{\partial F_{m-1}} \right]_{i=1}^N.$$

Наиболее выгодно добавить новое слагаемое в ансамбль классификаторов следующим образом:

$$F_m = F_{m-1} - b_m \nabla \bar{Q}, \quad b_m \in R,$$

Тогда оптимальный классификатор a_m задается в виде:

$$a_m = \operatorname{argmin} \sum_{i=1}^N Q(\nabla Q_i, h(x_i, a)) \equiv \text{обучить}(\{x_i\}_{i=1}^N, \{\nabla Q_i\}_{i=1}^N),$$

а параметр b_m выбирается линейным поиском по вещественным числам R :

$$b_m = \operatorname{argmin} \sum_{i=1}^N Q(F_{m-1}(x_i) - b h(x_i, a_m)).$$

1.4.4. Рандомизированные деревья

Классификатор на основе рандомизированных деревьев строит ансамбль неусеченных деревьев регрессии в соответствии с классической процедурой «сверху – вниз» [91]. Псевдокод алгоритма представлен в табл. 1.1.

Псевдо код алгоритма «Рандомизированные деревья»

ФУНКЦИЯ РАЗДЕЛИТЬ_УЗЕЛ(S)

ВХОД: S подмножество из тренировочного набора данных, которое мы хотим разделить в данной вершине;

ВЫХОД: Разделенное подмножество S , либо ничего;

ЕСЛИ ОСТАНОВКА_РАЗБИЕНИЯ(S)=TRUE

Заканчиваем выполнение;

ИНАЧЕ

Выбираем K признаков $\{a_1, \dots, a_K\}$ из всех не константных признаков в S ;

Вычисляем K разбиений множества $S\{s_1, \dots, s_K\}$,

$s_i = \text{СЛУЧАЙНОЕ_РАЗБИЕНИЕ}(S, a)$;

Возвращаем разбиение $s_* : Score(s_*, S) = \max_{i=1, \dots, K} (Score(s_i, S))$;

ФУНКЦИЯ СЛУЧАЙНОЕ_РАЗБИЕНИЕ(S, a)

ВХОД: Подмножество объектов S из тренировочного набора и признак a , по которому искать предикат;

ВЫХОД: Разделенное подмножество S по найденному предикату;

Пусть a_{max}^S и a_{min}^S максимальное и минимальное значение признака a на множестве S ;

Выбираем значение a_c произвольно из множества $[a_{min}^S, a_{max}^S]$;

Разбиваем выборку S на два множества

$$S_0 = \{s \in S : a^s < a_c\}$$

$$S_1 = \{s \in S : a^s \geq a_c\}$$

ФУНКЦИЯ ОСТАНОВКА_РАЗБИЕНИЯ(S)

ВХОД: подмножество S из тренировочного набора;

ВЫХОД: TRUE или FALSE;

Если $|S| < n_{min}$ то возвращаем TRUE;

Если все признаки в S принимают одинаковое значение, то возвращаем TRUE;

Если метки объектов в S совпадают, то возвращаем TRUE;

Иначе возвращаем FALSE;

У алгоритма есть два основных различия от других методов основанных на деревьях регрессии. Первое заключается в том, что предикат, по которому происходит разбиение выборки внутри вершины, выбирается случайным образом. Второе отличие в том, что в каждой вершине случайно выбирается подмножество признаков, по которым будет происходить поиск наилучшего предиката.

Алгоритм имеет следующие параметры: K – число признаков случайно выбираемых в каждой вершине, n_{min} – минимальное число объектов, которое можно разбить, M – количество деревьев в ансамбле. Параметры используются вместе с исходной обучающей выборкой для построения ансамбля деревьев регрессии. Эти параметры могут быть адаптированы к специфике задачи в ручном или автоматическом режиме.

1.4.5. Сверточные нейронные сети

Нейронные сети с успехом используются для решения задач классификации и распознавания объектов [70-72]. Широко распространенной моделью нейронной сети является многослойный персептрон, но его использование со стандартной структурой для решения реальных задач распознавания может создавать определенные трудности. Анализируемые изображения чаще всего обладают большой размерностью, в связи с чем растет количество нейронов и синаптических связей в сети. Это влечет за собой увеличение обучающей выборки, в результате чего растет время и вычислительная сложность процесса обучения. Кроме того, преобразование изображения в цепочку байт сопровождается безвозвратной потерей части информации, и от слоя к слою потери увеличиваются. В дополнение к этому, нейронная сеть должна быть устойчивой к смещениям, поворотам и масштабированию изображения.

Устранение всех вышеперечисленных недостатков привело к созданию нового вида нейронных сетей, называемых сверточными. Идея сверточных

нейронных сетей (СНС), которую предложил Ян Лекун в 1998 году, изначально была направлена на результативное распознавание изображений.

Применение СНС в настоящее время очень обширно:

- распознавание автомобильных номерных знаков;
- распознавание номеров домов на улицах, вагонов, платформ;
- распознавание лиц;
- распознавание дорожных знаков;
- сегментация и выделение различных объектов: людей, животных, внутренних частей тела (в медицине);
- распознавание речевых команд, идентификация диктора;
- распознавание рукописных символов и подписей.

Структура сверточной нейронной сети представлена на рис. 1.5.

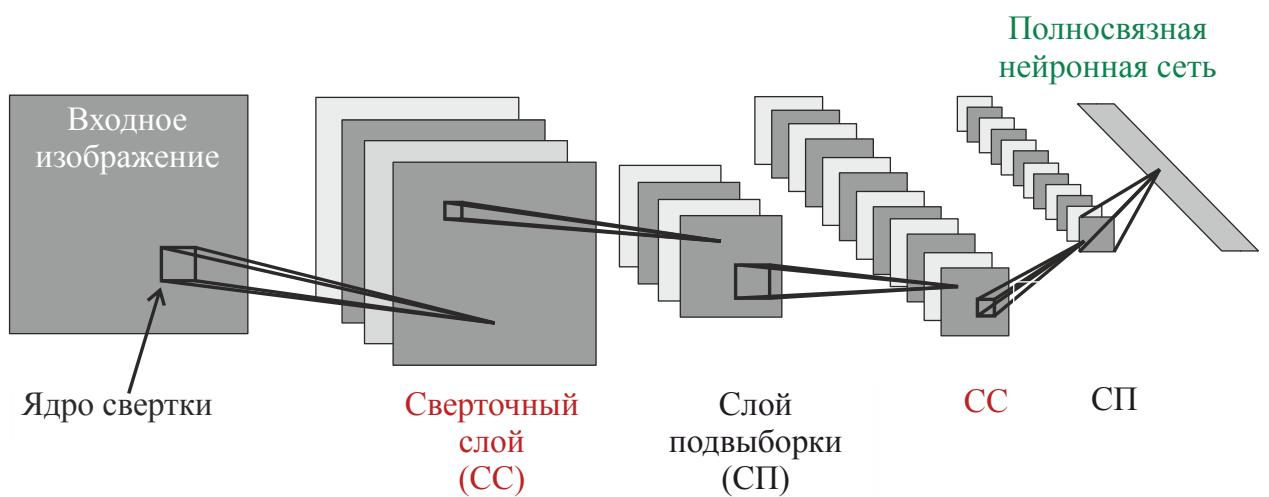


Рис. 1.5. Структура сверточной нейронной сети

СНС состоит из слоев свертки, субдискретизации (подвыборки) и слоев полносвязной нейронной сети.

Первоначально структура сверточной нейронной сети создавалась с учетом особенностей строения некоторых частей человеческого мозга, отвечающих за зрение [71].

В основу разработки таких сетей заложено три механизма:

- локальное восприятие;
- формирование слоев в виде набора карт признаков (разделяемые веса);
- субдискретизация (подвыборка).

Под локальным восприятием подразумевается, что на вход нейрона поступает не все изображение, а только некоторая его часть. Это позволяет сохранять конфигурацию изображения при переходе от слоя к слою.

Идея разделяемых весов подразумевает, что к большому числу связей применяется небольшой набор весов, т. е. каждая область изображения, на которые оно разделено, будет обработана одним и тем же набором весов. При таком искусственно созданном ограничении весов улучшается свойство сети к обобщению.

1.4.5.1. Слои сверточной нейронной сети

Большинство систем распознавания строится с помощью двумерных фильтров, которые представляют собой матрицы коэффициентов (ядра). Матрица применяется к изображению с помощью операции свертки. Суть операции заключается в том, что каждая часть изображения умножается на матрицу (ядро) поэлементно и полученные результаты суммируются. Далее эта сумма записывается в соответствующую позицию изображения на выходе системы. Основной особенностью этих фильтров является то, что их выходное значение тем больше, чем больше область изображения, к которой он применяется. В результате при сворачивании изображения с ядром получается другое изображение, каждый пиксель которого определяет степень сходства области изображения и фильтра. При каждом сдвиге ядра по изображению формируется сигнал активации для нейрона следующего слоя с соответствующей позицией. Ядро свертки строится так, чтобы кодировался один признак. В результате каждый следующий слой, который получается после свертки матрицы и области изображения, показывает наличие этого признака в обрабатываемом слое. Таким образом, получается

карта признаков, которая по своей сути осуществляет поиск признака во входных данных. Другие карты признаков обладают другими наборами весов и, следовательно, производят поиск других признаков. Признаки, обнаруживаемые такими картами, определяются на этапе обучения нейронной сети с учителем [71].

Такая форма усложняет структуру сети, но при этом позволяет значительно уменьшить число используемых параметров и, как следствие, вычислительную сложность системы. В СНС используется не один набор весов. Их количество зависит от числа признаков, которые требуется выделить.

Процесс обработки изображения в сверточном слое представлен на рис. 1.6.

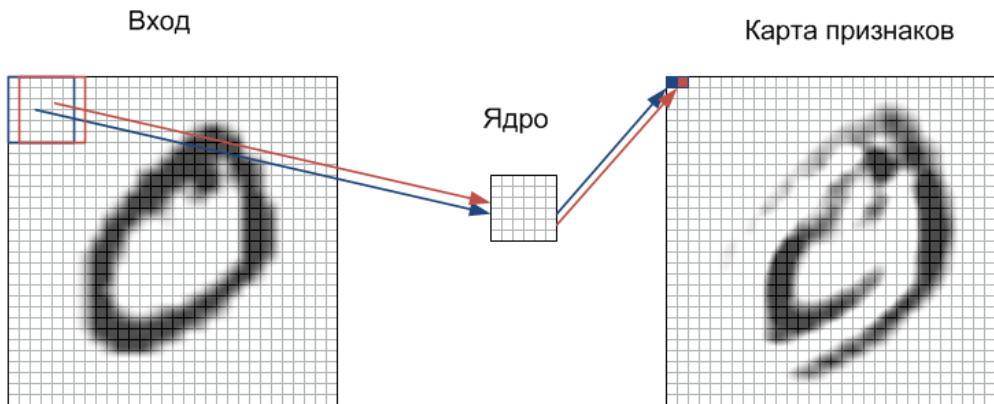


Рис. 1.6. Обработка изображения в сверточном слое

Слой субдискретизации (подвыборки)

За каждым сверточным слоем следует слой подвыборки, или вычислительный слой, который производит уменьшение размерности изображения (рис. 1.7). В архитектуре сверточной сети принято считать, что наличие признака важнее информации о его расположении. Поэтому из нескольких соседних нейронов в карте признаков выбирается максимальный и его значение считается одним нейроном в карте признаков

меньшей размерности. Отличие слоя подвыборки от слоя свертки состоит в том, что в последнем области соседних нейронов перекрываются, чего не происходит в слое субдискретизации.

Таким образом, СНС строится путем чередования слоев свертки и субдискретизации. На выходе сети обычно устанавливается несколько слоев полно связной нейронной сети, на вход которых подаются конечные карты признаков. Каждый нейрон этого слоя представляет собой персепtron, который имеет нелинейную функцию активации.

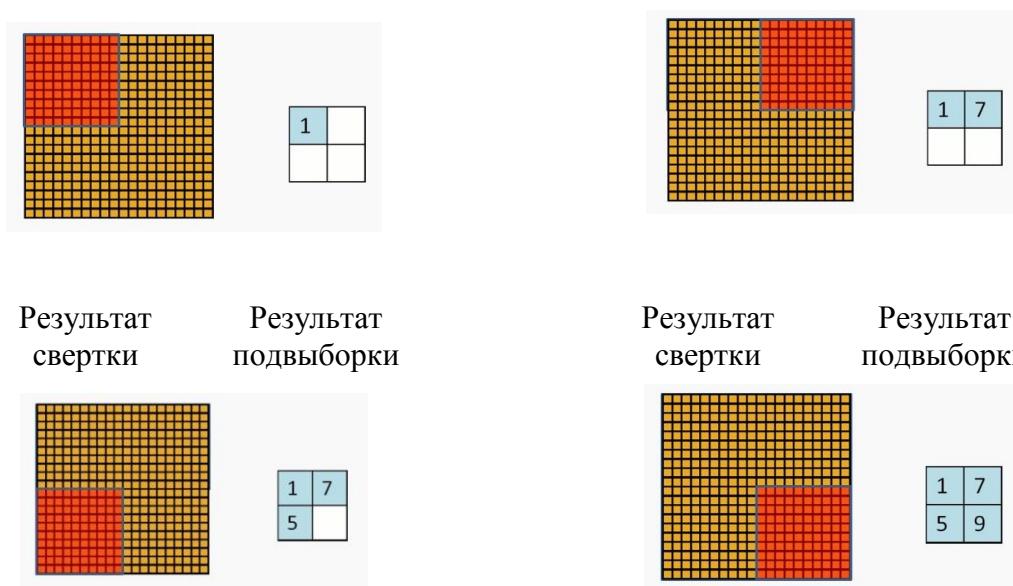


Рис. 1.7. Обработка изображения в слое субдискретизации

В процессе обучения сверточной сети также используется метод обратного распространения ошибки для вычисления ошибки в скрытых слоях, но применяется его модификация, учитывающая то, что веса у нейронов общие. В данном случае коррекция веса при рассогласовании нейронов проводится с учетом информации об ошибках всех нейронов, использующих данный вес.

1.4.5.2. Методы настройки параметров сверточной нейронной сети

Пакетные методы оптимизации, как, например, алгоритм BFGS (Бройдена-Флетчера-Гольдфарба-Шанно) с ограниченным использованием

памяти, которые используют набор обучения для вычисления следующего обновления параметров на каждой итерации, как правило, очень хорошо сходятся к локальным оптимумам (максимумам / минимумам). Тем не менее, часто на практике вычисления стоимостной функции и градиента для всей обучающей выборки полносвязной нейронной сети могут быть очень медленными, а иногда и неразрешимыми на одной машине, если набор данных слишком велик. Стохастический градиентный спуск позволяет решить проблему. В этом случае примеры из обучающей выборки поступают на вход нейронной сети друг за другом. После каждого примера происходит обновление весов. Использование стохастического градиентного спуска мотивировано высокой стоимостью запуска обратного распространения ошибки по всей обучающей выборке. Такой метод позволяет преодолеть эту стоимость и по-прежнему быстро сходится.

Стандартный алгоритм градиентного спуска обновляет параметры θ исходя из функции потерь $Q(\theta)$, как:

$$\theta = \theta - \alpha \nabla_{\theta} [Q(\theta)],$$

где математическое ожидание аппроксимируется оценкой функции потерь и градиента на полной обучающей выборке. Стохастический градиентный спуск избавляется от математического ожидания в обновлении и вычисляет градиент параметров, используя только один или несколько примеров обучения, выбранных случайным образом. Новое обновление задается следующим выражением:

$$\theta = \theta - \alpha \nabla_{\theta} Q(\theta; x^i, y^i),$$

с парой (x^i, y^i) из обучающего набора. Обычно каждое обновление параметра вычисляется в отношении нескольких обучающих примеров или мини-пакета в отличие от одного примера. Причина этому двоякая: во-первых, это уменьшает дисперсию в обновлении параметров и может привести к более устойчивой сходимости, во-вторых, позволяет при вычислении воспользоваться высоко оптимизированными матричными

операциями, которые должны использоваться в хорошо векторизованном вычислении значений функции стоимости и градиента. Типичный размер мини-пакета составляет 256 элементов. Оптимальный размер мини-пакета может изменяться для различных приложений и архитектур.

В стохастическом градиентном спуске скорость обучения α , как правило, значительно меньше, чем соответствующая скорость обучения в пакетном градиентном спуске, из-за значительных различий в обновлении. Выбор правильной скорости обучения и стратегии изменения скорости в процессе обучения может быть довольно трудным. Одним из стандартных методов, который хорошо работает на практике, является использование достаточно малой постоянной скорости обучения, которая обеспечивает устойчивую сходимость в начальную эпоху (при полном проходе через обучающий набор данных), а затем уменьшение значения скорости обучения вдвое, когда сходимость замедляется.

Еще одним важным моментом, касающимся стохастического градиентного спуска, является порядок, в котором предоставляются данные для алгоритма. Если данные представлены в некотором значимом порядке, то может возникнуть смещение градиента, приводящее к плохой сходимости. Обычно, хорошим способом избежать этого является перестановка данных случайным образом перед каждой эпохой обучения. В таком наборе данных вероятность содержания похожей информации в подряд идущих примерах резко сокращается.

1.5. База тестовых изображений LIVE

База данных LIVE получена из набора исходных изображений с различным содержанием. Двадцать девять цветных изображений высокого качества были собраны из Интернета [68]. Эти изображения включают в себя фотографии лиц, людей, животных, широкоугольные снимки природы,

искусственных объектов и изображения без какого-либо конкретного интересующего объекта (рис. 1.8).

Так же в рассматриваемой базе содержатся изображения искаженные различными типами помех (рис. 1.9):

- сжатие JPEG2000: искаженные изображения были получены путем сжатия эталонных изображений по стандарту JPEG2000 со значением параметра сжатия в диапазоне от 0,028 бит на пиксель (BPP) до 3,15 BPP;
- сжатие JPEG: искаженные изображения были получены путем сжатия эталонных изображений по стандарту JPEG с параметром сжатия в диапазоне от 0,15 BPP до 3,34 BPP;
- белый гауссовский шум: аддитивный белый гауссовский шум со значением стандартного отклонения σ был добавлен ко всем компонентам RGB изображения после их масштабирования. Значения стандартного отклонения σ изменялись от 0,012 до 2,0;
- размытие по Гауссу: три цветовые компоненты RGB изображения размывались путем свертки с двухмерным симметричным ядром Гаусса. Параметр ядра стандартное отклонение σ изменялся в интервале от 0,42 до 15 пикселей;
- битовые ошибки в канале связи: передавался поток данных сжатый по стандарту JPEG2000 при этом ОСШ приемника варьировалось от 15,5 до 26,1 дБ.

Качество каждого искаженного изображения оценивалось группой экспертов из 30 человек по шкале от одного до пяти (пять наивысшее качество), результаты оценок приводились к метрике DMOS.

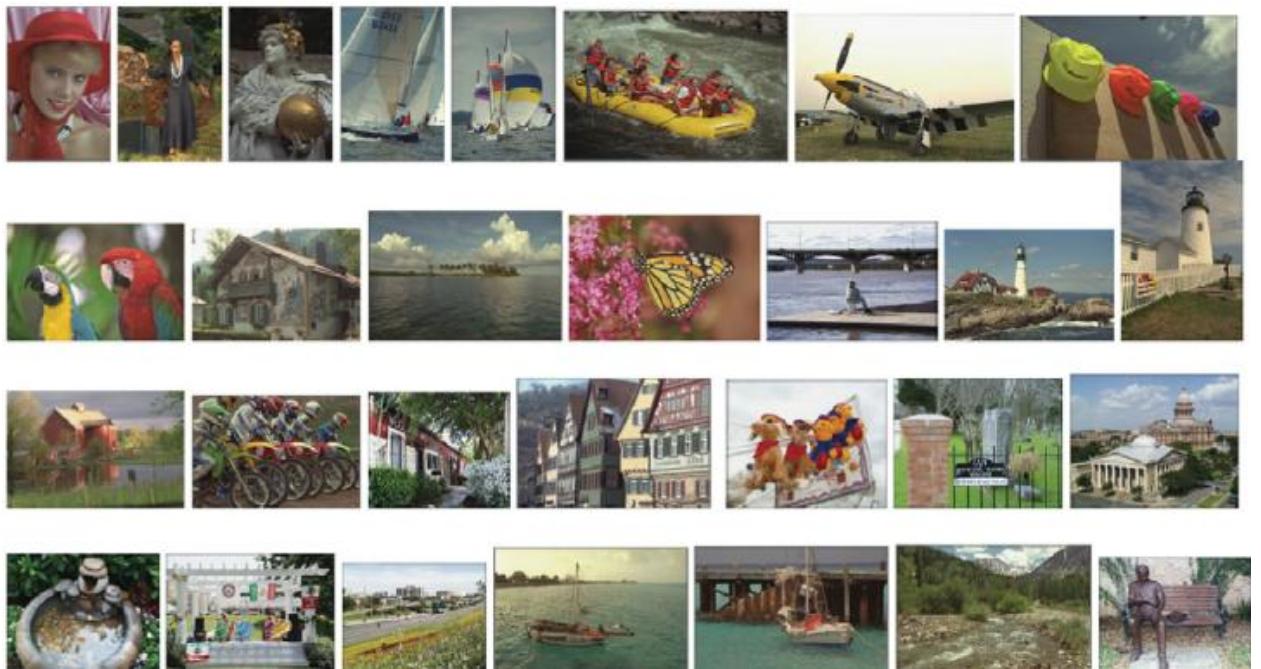


Рис. 1.8. Эталонные изображения из базы LIVE

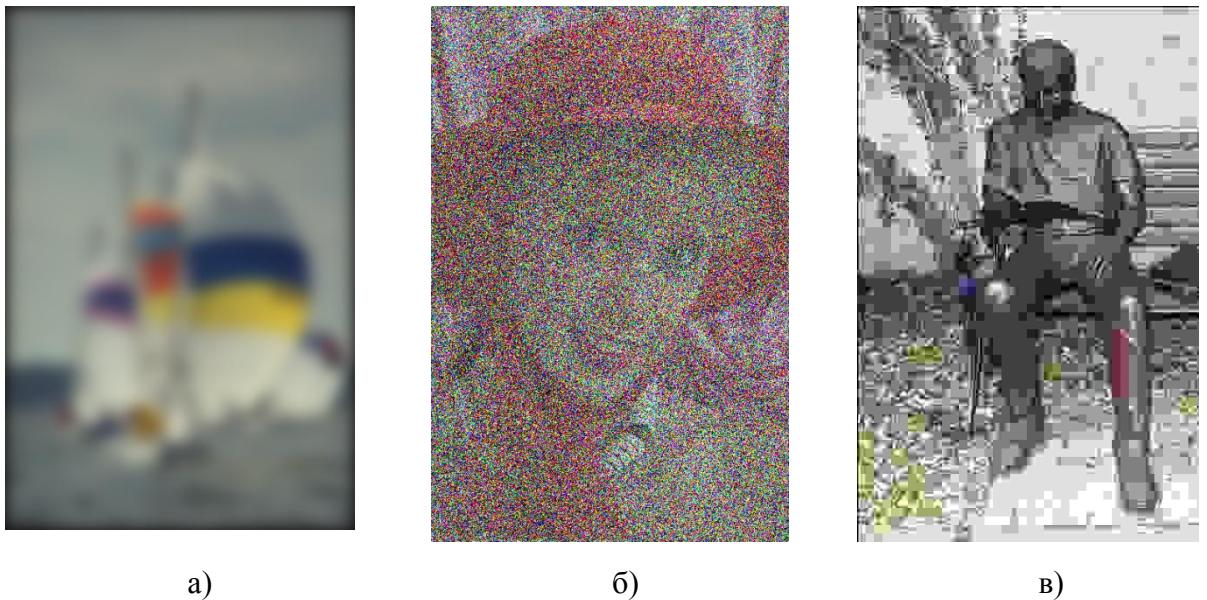


Рис. 1.9. Примеры искаженных изображений из базы LIVE:

а) размытие по Гауссу; б) белый шум; в) JPEG

1.6. База тестовых изображений TID2013

База изображений TID2013 содержит 25 эталонных неискаженных изображения, 24 из которых это изображения из базы Kodak [68], а 25 эталонное изображение создано искусственным образом (рис 1.10-1.11).



Рис. 1.10. Примеры изображений из базы TID2013

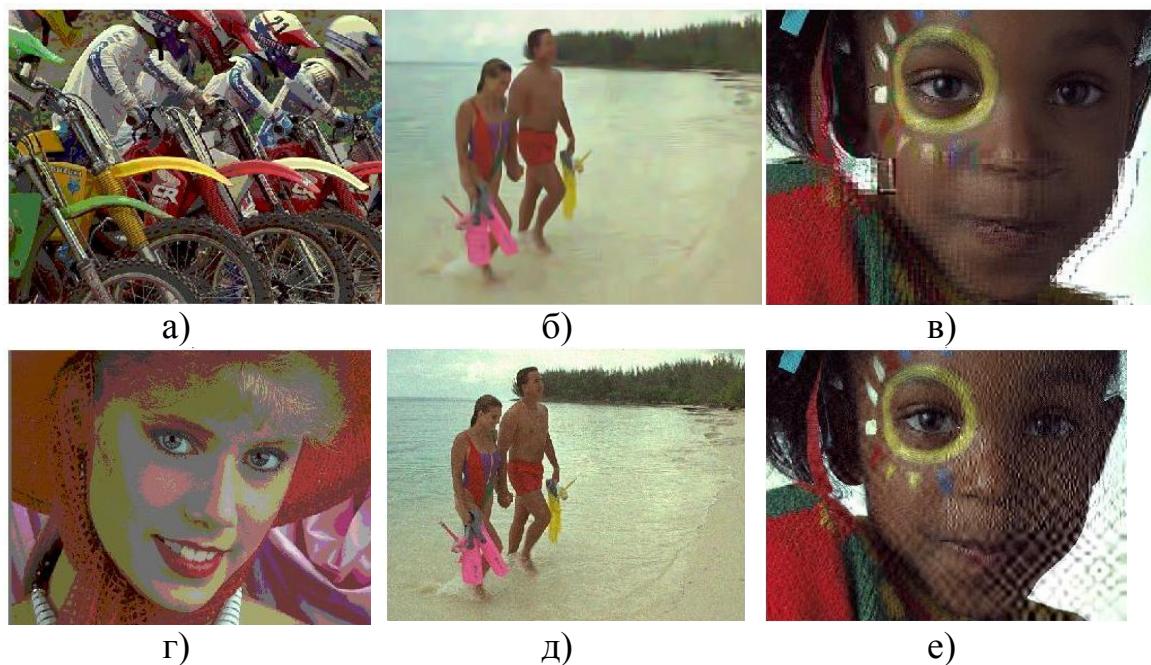


Рис. 1.11. Примеры искаженных изображений из базы TID2013:

- а) шум квантования ПОСШ 24,35 дБ; б) восстановленное от импульсного шума изображение ПОСШ 28,19 дБ; в) ошибки декодирования кодеком JPEG ПОСШ 24,05 дБ; г) шум квантования ПОСШ 24,21 дБ;
- д) импульсный шум ПОСШ 26,99 дБ; е) Ошибки декодирования кодеком JPEG2000 ПОСШ 23,98 дБ.

Все изображения из базы TID2013 подвергались 23-м типам различных искажений. Параметры искажений, выбирались такими, что бы после их воздействия качество изображения соответствовало одному из пяти уровней ПОСШ: 21 дБ, 24 дБ, 27дБ, 30 дБ, 33 дБ. В результате в базу TID2013 вошли 3000 изображений (рис 1.11).

Типы искажений, которым подвергались изображения:

1. Аддитивный гауссовский шум
2. Аддитивный шум в цветовых компонентах
3. Шум с пространственной корреляцией
4. Маскирующий шум (Masked noise)
5. Высокочастотный шум
6. Импульсный шум
7. Шум квантования
8. Размытие по Гауссу
9. Искажения, возникающие в результате применения алгоритмов шумоподавления
10. Сжатие JPEG
11. Сжатие JPEG2000
12. Битовые ошибки в JPEG потоке
13. Битовые ошибки в JPEG2000 потоке
14. Специальная модель шума (Non eccentricity pattern noise)
15. Локальные внутри блоковые искажения разной интенсивности
16. Сдвиг среднего
17. Изменение контрастности
18. Мультипликативный шум Гаусса
19. Комфортный шум (Comfort noise)
20. Сжатие с потерей информации
21. Квантование цветных изображений (Image color quantization with dither)
22. Хроматические aberrации (Chromatic aberrations)

23. Пространственное сэмплирование и реконструкция (Sparse sampling and reconstruction

1.7. Метрики, используемые для тестирования точности алгоритмов неэталонной оценки качества изображений

Пусть $\{x_i\}^N = \{x_1, x_2, \dots, x_N\}$ оценки качества на выходе алгоритма неэталонной оценки качества изображений для N тестовых изображениях, а $\{y_i\}^N = \{y_1, y_2, \dots, y_N\}$ – соответствующие экспертные оценки качества (MOS или DMOS). Тогда коэффициент корреляции Пирсона рассчитывается по формуле [69]:

$$r = \frac{\sum_{i=1}^N (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^N (y_i - \bar{y})^2 (x_i - \bar{x})^2}},$$

где \bar{y} – среднее значение экспертной оценки, а \bar{x} – среднее значение оценки на выходе алгоритма.

Коэффициент корреляции Спирмена рассчитывается по формуле [69]:

$$\rho = 1 - \frac{6}{N(N-1)(N+1)} \sum_{i=1}^N (R_i - S_i)^2,$$

где R_i – ранг наблюдения x_i в x , S_i – ранг наблюдения y_i в ряду y .

Коэффициент корреляции Кендалла рассчитывается по формуле [69]:

$$\tau = 1 - \frac{4}{N(N-1)} R,$$

$$R = \sum_{i=1}^{N-1} \sum_{j=i+1}^N \left[[x_i < x_j] \neq [y_i < y_j] \right].$$

Коэффициент τ принимает значения от $[-1; 1]$, равенство $\tau=1$ указывает на строгую линейную зависимость между x и y , $\tau=-1$ на обратную.

1.8. Краткие выводы

1. Актуальной проблемой в области цифровой обработки изображений является вопрос оценки качества изображений. На сегодняшний день наиболее точной мерой оценки качества являются субъективные критерии. В то же время проведение субъективных экспертиз не всегда удобно, поскольку занимает достаточно много времени и требует привлечения большого количества людей, поэтому в настоящее время актуальной является разработка численных алгоритмов определения качества.
2. На современном этапе развития систем обработки изображений особый практический интерес представляет разработка неэталонных критериев оценки качества изображений на основе современных алгоритмов машинного обучения.
3. Для тестирования различных метрик оценки качества изображений на международном уровне были созданы унифицированные тестовые базы изображений с известными субъективными оценками – LIVE и TID2013. Также разработана процедура сравнения алгоритмов на основе расчета коэффициентов корреляции Пирсона, Спирмена и Кендалла.
4. Проведенный анализ позволил сформулировать следующие основные задачи диссертации:
 - проведение исследований в части выбора признаков, характеризующих качество изображения в пространственной области;
 - проведение исследований по выбору алгоритма машинного обучения для классификации изображений различного качества;
 - разработка и анализ алгоритма неэталонной оценки качества на основе выбранных признаков и алгоритмов классификации;

- проведение исследований по оценке корреляции между разработанным неэталонным алгоритмом и результатами субъективной экспертизы оценки качества на специализированных базах изображений LIVE и TID2013;
- проведение анализа использования предложенного алгоритма неэталонной оценки качества в ряде задач вещательного и прикладного телевидения и в задачах биометрической идентификации личности.

ГЛАВА 2. РАЗРАБОТКА АЛГОРИТМОВ НЕЭТАЛОННОЙ ОЦЕНКИ КАЧЕСТВА ИЗОБРАЖЕНИЙ

2.1. Вводные замечания

Алгоритм неэталонной оценки качества на основе локальных бинарных шаблонов (НОК ЛБШ), разработанный группой японских ученых в 2013 году, для своей работы не требует информации о типах возможных искажений, присутствующих в изображении. Перед непосредственной работой алгоритма, его необходимо обучить на специально размеченных для этого данных. В оригинальной работе обучение и тестирование алгоритма происходит на тестовой базе данных LIVE. Отдельные функциональные блоки алгоритма моделируют различные особенности зрительной системы человека: многомасштабная декомпозиция входных данных, маскирующий эффект, формирование первоначального эскиза [10, 49]. Всего в алгоритме можно выделить 4 блока: фильтрация входящего изображения с помощью LOG-фильтра, вычисление локальных бинарных шаблонов, построение гистограммы и сопоставление столбцов гистограммы с экспертной оценкой DMOS при помощи классификатора на основе машины опорных векторов (рис. 2.1). Рассмотрим этапы работы НОК ЛБШ более подробно.



Рис. 2.1. Схема алгоритма НОК ЛБШ

2.2. Этапы работы алгоритма НОК ЛБШ

2.2.1. LOG-фильтр (лапласиан гауссиан фильтр)

Доказано, что нейроны в первичной зрительной коре (V1) производят многомасштабную декомпозицию входных данных [49]. Такая декомпозиция может быть смоделирована серией фильтров лапласиан гауссиан, которые определяются как:

$$\nabla^2 G(r, \sigma) = -\frac{1}{2\sigma^4} \left(1 - \frac{r^2}{2\sigma^2}\right) \exp\left(-\frac{r^2}{2\sigma^2}\right),$$

где $\nabla^2 G(r, \sigma) = -\frac{1}{2\pi\sigma^2} \left(1 - \frac{r^2}{2\sigma^2}\right)$ – ядро Гаусса, r это l_2 норма X , а X – позиция в тестовом изображении I .

Отклик фильтра может быть получен путем сверки ядра фильтра $\nabla^2 G(r, \sigma)$ с изображением I , формально процесс свертки с заданным ядром представлен ниже:

$$F(X, \sigma) = \nabla^2 G(r, \sigma) \otimes I(X)$$

где \otimes оператор свертки. В ходе экспериментов каждое тестовое изображение обрабатывалось 4-мя LOG-фильтрами с соответствующими значениями $\sigma = 0.5; 1.3; 2.6; 5.2$.

2.2.2. Вычисление ЛБШ-шаблонов и построение гистограммы

В алгоритме НОК ЛБШ качественными признаками являются элементы гистограммы многомасштабных локальных бинарных шаблонов. Стоить отметить, что гистограмма строится для равномерных инвариантных к повороту шаблонов. Радиус окрестности R принимает значения – 1, 2, 3. Для всех значений R параметр P выбирается равным 4-м. Всего в рассматриваемом алгоритме для тестового изображения рассчитывается 72 признака.

2.2.3. Классификация на основе машины опорных векторов

В качестве классификатора для сопоставления вектора качественных признаков с оценкой DMOS в данном алгоритме используется классификатор на основе машины опорных векторов с использованием ядра Гаусса. Параметры машины опорных векторов C и ядра σ выбирались с помощью техники сеточного поиска в логарифмическом пространстве.

2.3. Результаты моделирования алгоритма на тестовой базе изображений LIVE

В ходе процедуры обучения база LIVE делилась произвольным образом в отношении 80/20 на тренировочный и тестовый наборы данных, так что бы они не имели общих изображений. В качестве метрик точности алгоритма, использовались значения коэффициента корреляции Спирмена и Пирсона. Процедуры обучения и тестирования повторялись 1000 раз, в качестве итогового значения метрики выступает медиана. Полученные результаты относительно коэффициентов корреляции Пирсона и Спирмена приведены в табл. 2.1 и табл. 2.2.

Таблица 2.1

Значения коэффициента корреляции Пирсона для алгоритма НОК ЛБШ на базе LIVE

JPEG	JPEG2000	Белый шум	Размытие	Битовые ошибки	Все типы
0,949	0,93	0,96	0,94	0,91	0,93

Таблица 2.2

Значения коэффициента корреляции Спирмена для алгоритма НОК ЛБШ на базе LIVE

JPEG	JPEG2000	Белый шум	Размытие	Битовые ошибки	Все типы
0,93	0,92	0,94	0,94	0,88	0,93

Несмотря на достаточно высокую точность рассматриваемого алгоритма, у него существуют следующие недостатки:

1. Авторами не описана процедура выбора параметра σ для LOG-фильтра;
2. Теоретически или экспериментально не показано, почему к изображению применяется именно 4 фильтра;
3. LOG-фильтр делает алгоритм вычислительно достаточно сложным. Среднее значение времени работы алгоритма на изображении из базы LIVE 1c.

В работах [31, 33-36] показано, что локальные бинарные шаблоны достаточно точно работают в пространстве пикселей изображения без дополнительных преобразований. Расчет локальных бинарных шаблонов непосредственно в пространственной области, позволит сократить время извлечения качественных признаков и тем самым ускорить алгоритм неэталонной оценки качества, построенный на их основе.

2.4. Локальные бинарные шаблоны

Локальные бинарные шаблоны – простой оператор, используемый для классификации текстур в техническом зрении. Имеет сложность вычисления $O(n)$, n – где число пикселей в изображении. Впервые был описан в 1996 году. ЛБШ представляет собой описание окрестности пикселя изображения в двоичной форме [34]. Первоначальный вариант оператора ЛБШ работал в окрестности пикселя размером 3×3 , используя значение яркости центрального пикселя в качестве порога. Пиксели из окрестности, имеющие значение яркости больше порогового (или равное ему), принимали значения «1», остальные «0». Таким образом, формировался бинарный код, описывающий окрестность заданного пикселя (рис 2.2).



Рис. 2.2. Пример вычисления первоначального варианта оператора локального бинарного шаблона в окрестности 3×3

Позже была представлена более общая форма оператора [31]. В отличие от первоначального варианта обобщенный вариант применяется не только в окрестности 3×3 , но и в любой другой окрестности размера, которой определяются двумя параметрами P и r . Первый параметр отвечает за то, сколько точек из окрестности будет участвовать в расчетах, а второй параметр радиус, определяющий размер окрестности непосредственно. Так же в отличие от первоначального варианта в обобщенном варианте окрестность имеет форму окружности (рис 2.3).

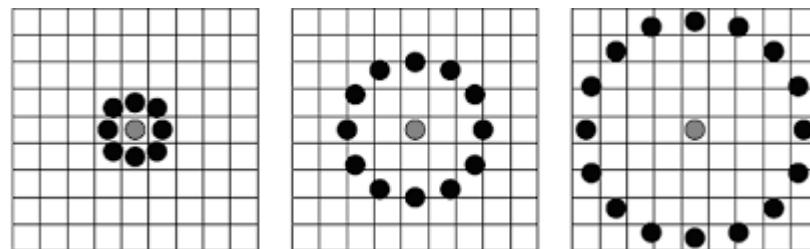


Рис. 2.3. Примеры выбора точек из окрестности оператором ЛБШ с параметрами P, r :
 $(8, 1), (12, 2), (16, 4)$

Пусть $I(x, y)$ – черно-белое изображение, g_c – значение яркости пикселя в точке с координатами (x, y) . Точки окрестности выбираются следующим образом:

$$g_p = I(x_p, y_p), \quad p = 0, \dots, P - 1$$

$$x_p = x + R \cos\left(\frac{2\pi p}{P}\right),$$

$$y_p = y - R \sin\left(\frac{2\pi p}{P}\right),$$

где g_p - значение яркости пикселя из окрестности, x_p и y_p - координаты пикселя из окрестности. Значения яркости пикселей из окрестности, которые имеют нецелые координаты, вычисляются при помощи билинейной интерполяции.

Формально обобщенный оператор ЛБШ можно представить в виде:

$$\text{ЛБШ}_{P,r}(x, y) = \sum_{p=0}^{P-1} S(g_p - g_c) 2^p,$$

$$S(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases}.$$

Для описания изображения в пространстве признаков ЛБШ используют гистограмму, столбцы которой соответствуют десятичным значениям бинарных кодов шаблонов. Формально гистограмма ЛБШ может быть представлена в виде:

$$H(k) = \sum_{m=1}^M \sum_{n=1}^N f(\text{ЛБШ}_{P,r}(m, n), k), k \in [0, K],$$

$$f(x, y) = \begin{cases} 1, & x = y \\ 0, & x \neq y \end{cases},$$

где K число шаблонов.

Локальные бинарные шаблоны определенного вида несут в себе больше информации, чем остальные. Так ЛБШ называется равномерным (РЛБШ), если он содержит не более двух переходов между «0» и «1» (например, 00000011, 00111000, 11111111). Равномерные локальные бинарные шаблоны отражают важные структурные особенности изображения, такие как концы линий, углы и пятна (рис. 2.4), поэтому могут

рассматриваться, как модель «первоначального эскиза», использующегося на ранних этапах зрительной системы человека [10].

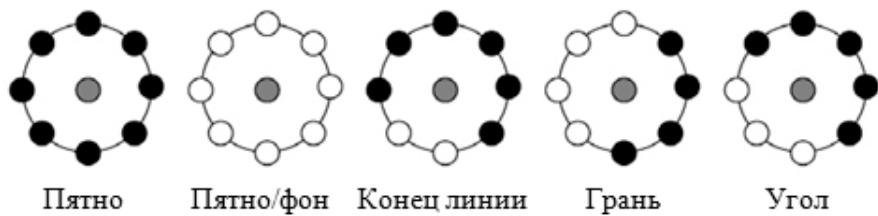


Рис. 2.4. Типы окрестностей

Исследователями обнаружено, что в среднем 90 % всех вычисленных ЛБШ для изображения являются равномерными [33]. В работе [30] отмечается, что использование только равномерных шаблонов в качестве признаков, приводит к увеличению точности систем распознавания изображений. Кроме того, работа с РЛБШ позволяет получать более компактное представление изображения. Число РЛБШ шаблонов $P(P - 1) + 2$, а число ЛБШ-шаблонов произвольного вида – 2^P . Стоит так же выделить определенную особенность при формировании бинарного кода при работе с равномерными шаблонами – всем неравномерным шаблонам присваивается одинаковое значение равное $P(P - 1) + 3$. Такая особенность нумерации РЛБШ, приводит к тому, что структурным особенностям изображения (части которые описываются РЛБШ) уделяется большее внимание, чем остальным частям, которые описываются обычными шаблонами. Данный факт может служить воплощением особенности присущей зрительной системе человека – человек отчетливее видит искажения, когда они появляются на структурированных частях изображения [32].

Поворот изображения приводит к повороту пикселей относительно центрального внутри окрестности, что ведет к изменению бинарных кодов шаблонов. Для того чтобы набор ЛБШ признаков для изображения не изменялся с поворотом, вводится специальная процедура присваивания

бинарных кодов соответствующим ЛБШ шаблонам – каждый бинарный код приводится к своему минимальному значению путем битового циклического сдвига:

$$\text{ЛБШ}_{P,r}^{ri}(x,y) = \min_i \text{СДВИГ}(\text{ЛБШ}_{P,r}(x,y), i),$$

где функция СДВИГ(код, i) представляет собой функцию циклического битового сдвига на i шагов. Таким образом, шаблонам 00110000, 11000000, 00001100 будет присвоен бинарный код 00000011. Использование инвариантных к повороту шаблонов позволяет получать еще более компактное представление изображения в пространстве ЛБШ признаков, так как число таких шаблонов $P + 1$. Код $P + 2$ зарезервирован для всех не равномерных шаблонов.

В работе [30] показано, что ЛБШ при $r=1$ чувствительны к изменению направления освещения и угла обзора. Более устойчивые признаки можно получить, используя многомасштабный оператор ЛБШ. Под многомасштабностью понимается одновременное вычисление ЛБШ для заданного пикселя в окрестностях разного радиуса. Стоит отметить, что максимальный радиус окрестности r необходимо выбирать меньшим 3, так как при $r > 3$ эффективность ЛБШ резко снижается. Описание изображения в пространстве ЛБШ признаков получается путем склейки гистограмм соответствующих одномасштабных операторов. Многомасштабные ЛБШ несмотря на повышенную точность, не теряют своей вычислительной эффективности, сложность вычисления многомасштабных шаблонов также $O(n)$ – где n число пикселей в изображении.

2.5. Описание алгоритма неэталонной оценки качества на основе локальных бинарных шаблонов

Предлагаемый алгоритм состоит из двух частей. Первая часть – извлечение из изображения признаков, содержащих в себе информацию о его качестве. В алгоритме НОК ЛБШ-1 в качестве признаков выступают

многомасштабные локальные бинарные шаблоны с радиусом 1, 2, 3 пикселя. Вторая часть – получение из вектора признаков оценки качества, которая должна максимально коррелировать с результатом субъективной оценки качества изображений (DMOS). Для вычисления оценки будет использован классификатор на основе рандомизированных деревьев. Так как предложенный алгоритм пытается предсказать субъективную оценку DMOS, исходя из изображения, то результатом его работы будет число в диапазоне от 0 до 100, причем, чем больше это число, тем хуже оцениваемое изображение. Схема алгоритма представлена на рис. 2.5.

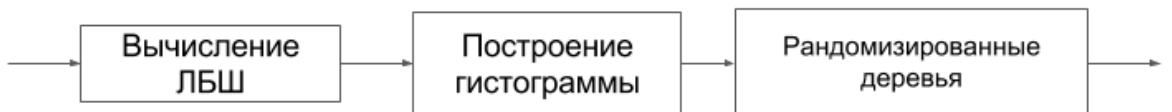


Рис. 2.5. Схема работы алгоритма НОК ЛБШ-1

В качестве вектора признаков для предложенного алгоритма выступают столбцы гистограмм многомасштабных равномерных локальных бинарных шаблонов инвариантных к сдвигу с $r = 1,2,3$ и значениями $P = 8, 16, 24$ [31]. Гистограммы для каждого масштаба строятся отдельно, после чего формируется результирующая гистограмма путем склейки гистограмм всех масштабов. Всего для каждого изображения рассчитывается 54 признака. Выбор многомасштабных локальных бинарных шаблонов, обусловлен их повышенной точностью, по сравнению с одномасштабными шаблонами.

Вычисленные признаки поступают на вход классификатора на основе рандомизированных деревьев. Параметры классификатора побирались во время обучения с использованием процедуры сеточного поиска.

2.6. Визуализация ЛБШ-признаков для различных типов искажений

В работах по оценке качества изображений [49] локальные бинарные шаблоны применялись к изображениям прошедшим предварительную процедуру фильтрации, в частности LOG-фильтр. В работах [31, 33, 35, 36] показано, что оператор ЛБШ может быть успешно применен к изображениям без предобработки. На рис. 2.6-2.9 представлено соответствие гистограммы равномерных инвариантных к повороту ЛБШ для изображений из базы LIVE искаженных различными типами помех.

Десятичные номера шаблонов соответствуют следующим бинарным кодам:

- 0 - 00000000
- 1 - 00000001
- 2 - 00000011
- 3 - 00000111
- 4 - 00001111
- 5 - 00011111
- 6 - 00111111
- 7 - 01111111
- 8 - 11111111
- 9 – все неравномерные шаблоны

Визуально можно заметить, что белый шум вызывает провал в центральной части гистограммы, при этом увеличивая число неравномерных шаблонов. Для сжатия типа JPEG резко увеличивается число шаблонов с номером 5 соответствующих границам, что можно объяснить границами, возникающими между блоками при блочности. Так при сжатии JPEG увеличивается шаблонов с номером 9, соответствующих областям с одинаковыми интенсивностями пикселей, визуально можно заметить, что

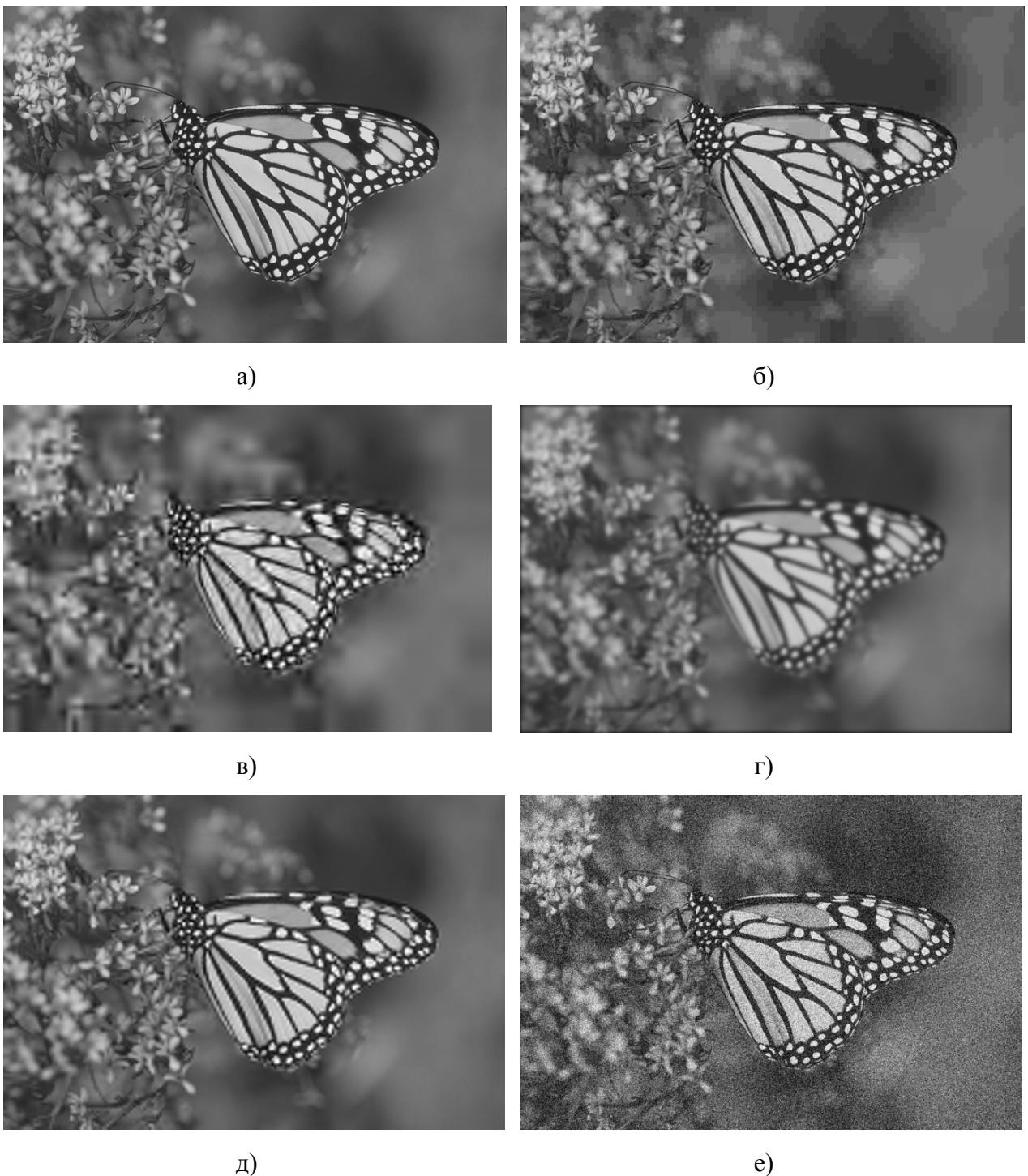
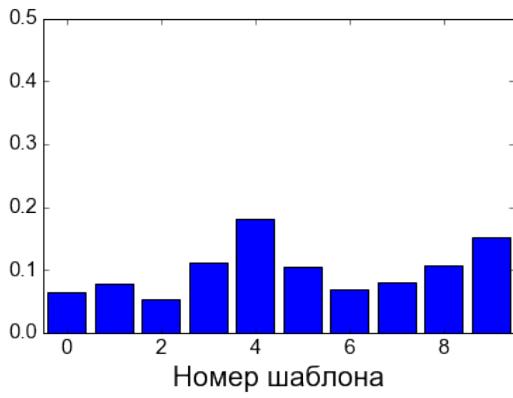
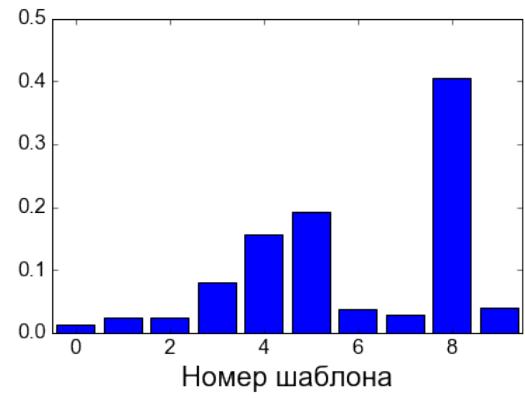


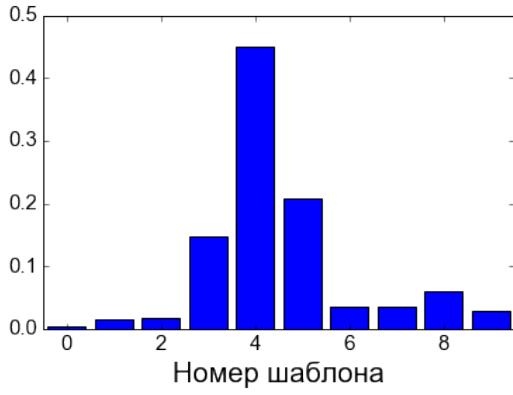
Рис. 2.6. Изображения с различными типами искажений: а) оригинал; б) jpeg; в) jpeg2000;
г) размытие; д) канал с замираниями; е) белый шум



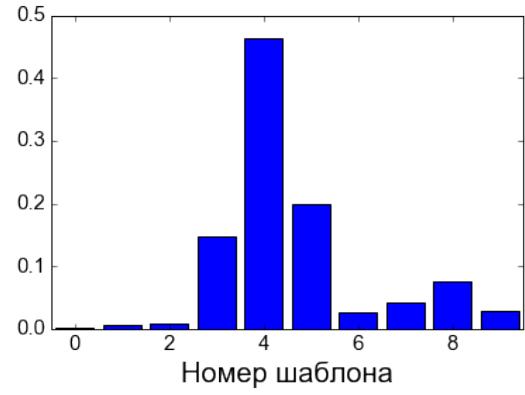
а)



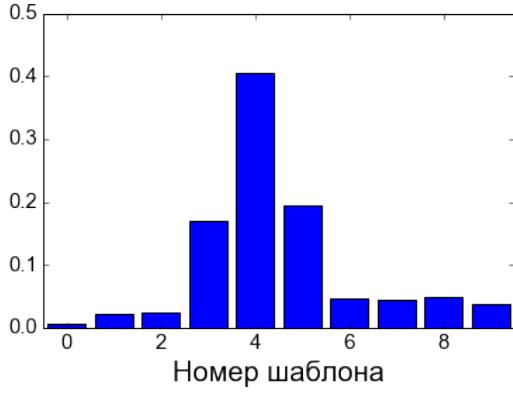
б)



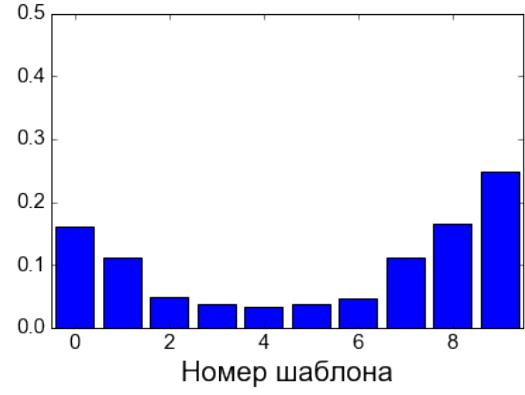
в)



г)



д)



е)

Рис. 2.7. Гистограмма равномерных инвариантных к сдвигу ЛБШ для изображений искаженных различными видами искажений: а) оригинал; б) jpeg; в) jpeg2000; г) размытие; д) канал с замираниями; е) белый шум

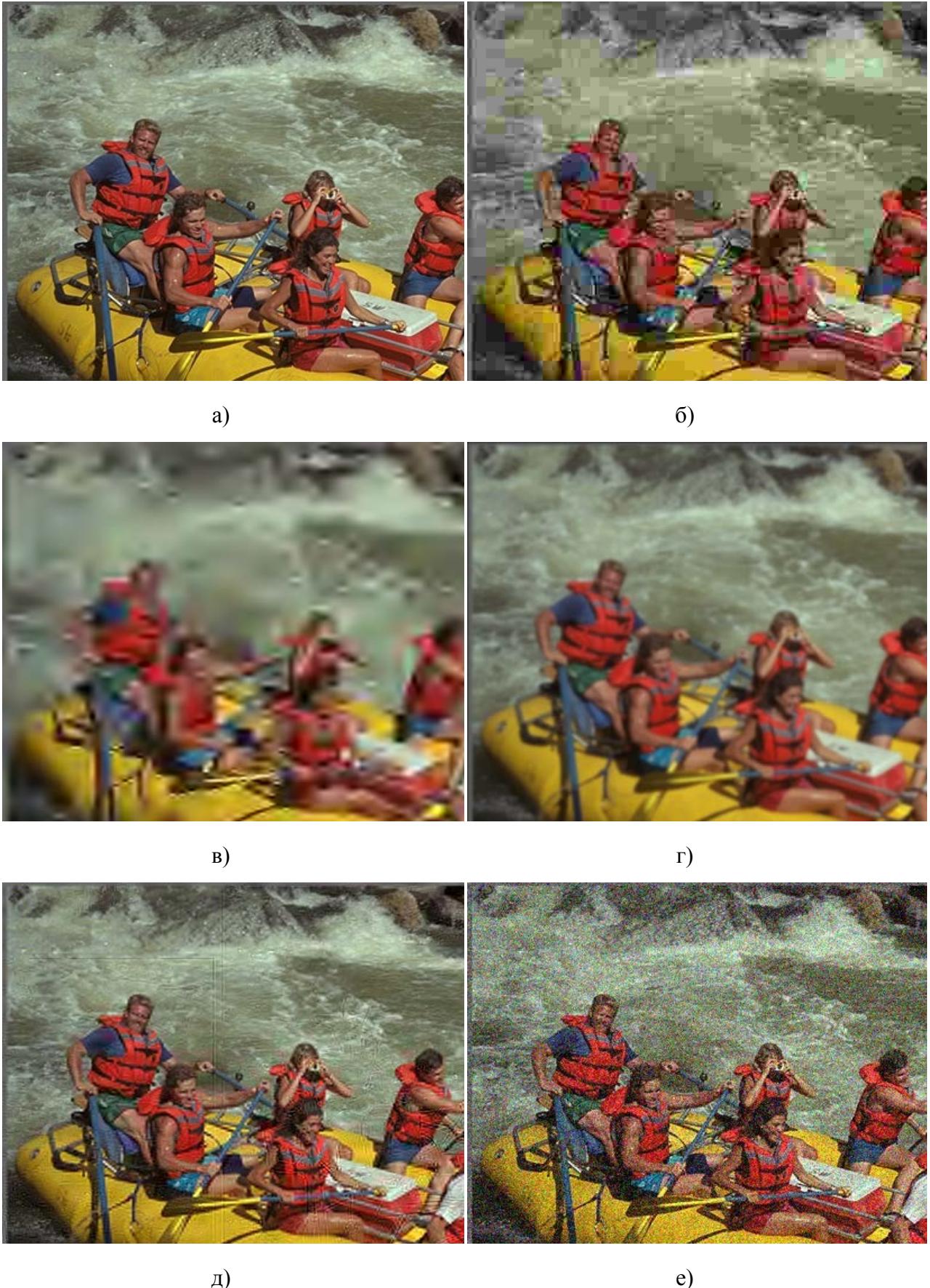
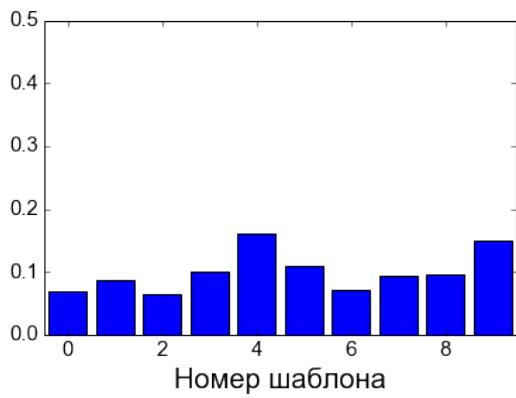
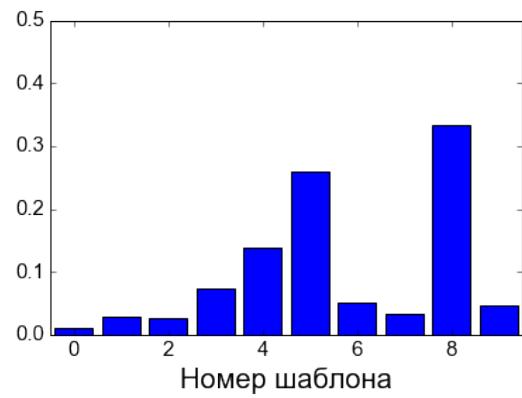


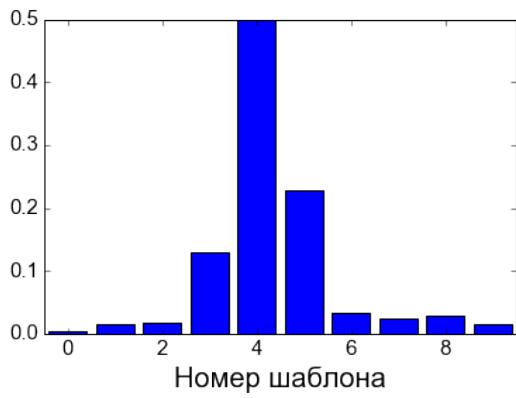
Рис. 2.8. Изображения с различными типами искажений: а) оригинал; б) jpeg; в) jpeg2000; г) размытие; д) канал с замираниями; е) белый шум



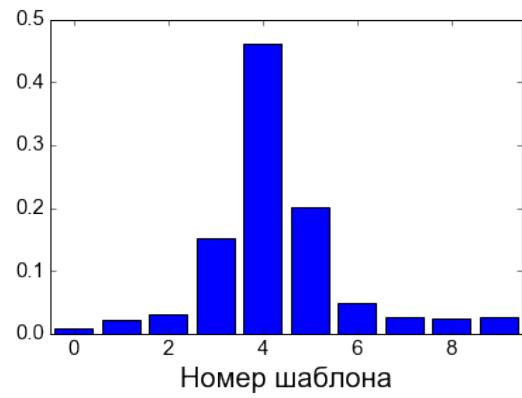
а)



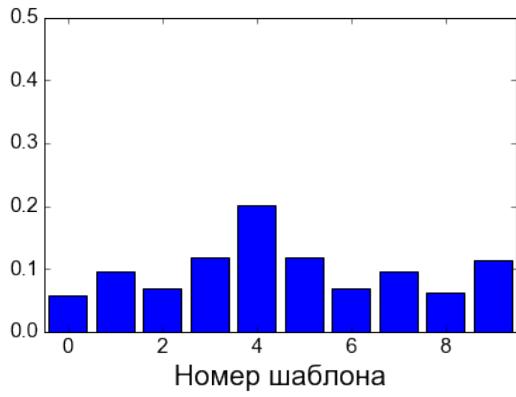
б)



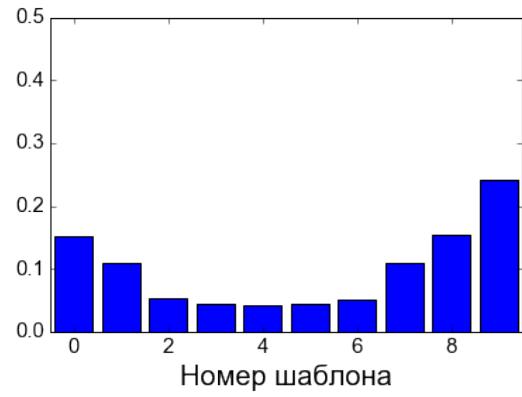
в)



г)



д)



е)

Рис. 2.9. Гистограмма равномерных инвариантных к сдвигу ЛБШ для изображений искаженных различными видами искажений: а) оригинал; б) jpeg; в) jpeg2000; г) размытие; д) канал с замираниями; е) белый шум

такие области возникают внутри блоков. Для искажения типа канал с замираниями гистограмма ЛБШ признаков не значительно отличается от гистограммы не искаженного изображения, тот же самый вывод можно сделать и о качестве изображения. Гистограммы для искажений JPEG2000 и размытие являются схожими, мы видим, что преимущественно все ЛБШ являются шаблонами с номерами 3, 4, 5, это можно объяснить тем, что при размытии происходит размытие границ объектов и один граничный переход превращается в несколько плавных переходов.

2.7. Неэталонная оценка качества в пространственной области

2.7.1. Описание алгоритма

Для получения точных оценок современные алгоритмы неэталонной оценки качества обучаются на тестовых изображениях с известными субъективными оценками. Алгоритм неэталонной оценки качества изображений в пространственной области (НОКПО) не исключение – из каждого анализируемого изображения извлекаются признаки, отражающие его качество, из которых позже рассчитывается значение метрики DMOS с помощью машины опорных векторов с ядром Гаусса.

В основе алгоритма извлечения признаков лежит тот факт, что значения усредненных нормализованных коэффициентов контраста (УНКК) для неискаженных изображений, удовлетворяют распределению Гаусса [62]:

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

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

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

где $i \in 1, 2 \dots M$, $j \in 1, 2 \dots N$ – пространственные индексы, M , N – высота и ширина изображения, а $C = 1$ – константа, необходима для исключения случаев деления на нуль, $\{\omega_{k,l} | k = -K, \dots, K, l = -L, \dots L\}$ – двумерная циклически-симметричная взвешенная функция Гаусса. Параметры K, L – небольшие целые числа.

Экспериментально установлено, что внесение различных типов искажений в изображение влияет на распределение УНКК, причем между типом искажения и распределением УНКК существуют определенные закономерности [62]. Например, белый шум уменьшает значения «хвостов» распределения, а размытие искажает распределение так, что оно становится похожим по форме на распределение Лапласа. Множество всевозможных распределений искаженных изображений удобно записать в виде обобщенного распределения Гаусса:

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} \exp(-(\frac{|x|^\alpha}{\beta})),$$

$$\beta = \sigma \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}},$$

$$\Gamma(\alpha) = \int_0^\infty t^{\alpha-1} e^{-t} dt \quad \alpha > 0,$$

где α – параметр формы, σ^2 – дисперсия.

Для каждого изображения оценивается два параметра (α , σ^2) – их значения формируют первый набор признаков для алгоритма НОКПО. При вычислении следующего набора признаков принимаются во внимание статистические взаимоотношения между соседними пикселями. Для каждого пикселя считаются попарные произведения УНКК с коэффициентами соседних пикселей. Предполагается также, что совокупность распределений

попарных произведений можно описать с помощью асимметричного обобщенного распределения Гаусса:

$$f(x; \nu, \sigma_l^2, \sigma_r^2) = \begin{cases} \frac{\nu}{(\beta_l + \beta_r)\Gamma(1/\nu)} \exp\left(-\left(\frac{-x}{\beta_l}\right)^\nu\right) & x < 0 \\ \frac{\nu}{(\beta_l + \beta_r)\Gamma(1/\nu)} \exp\left(-\left(\frac{-x}{\beta_r}\right)^\nu\right) & x \geq 0 \end{cases}$$

$$\beta_l = \sigma_l \sqrt{\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}},$$

$$\beta_r = \sigma_r \sqrt{\frac{\Gamma(1/\nu)}{\Gamma(3/\nu)}}.$$

Второй набор признаков алгоритма НОКПО вычисляется из параметров $(\eta, \nu, \sigma_r^2, \sigma_l^2)$.

Первый параметр вычисляется по следующей формуле:

$$\eta = (\beta_r - \beta_l) \frac{\Gamma\left(\frac{2}{\nu}\right)}{\Gamma\left(\frac{1}{\nu}\right)}.$$

Параметр ν отвечает за форму, а параметры σ_r^2 и σ_l^2 за размер правой и левой частей распределения. Полный набор признаков вычисляется для двух изображений – исходного и изображения, полученного децимацией с коэффициентом 2. Суммарно для исследуемого изображения считается 36 признаков, на основе которых рассчитывается неэталонная оценка качества.

2.7.2. Результаты моделирования

Произведено исследование метрики НОКПО в задаче оценки качества восстановленных изображений, искаженных импульсным шумом. В качестве моделей импульсного шума использовались модели «соль и перец» и модель со случайными равномерно распределенными значениями импульсов. Для удаления импульсного шума типа «соль и перец» использовались следующие фильтры – переключающийся медианный фильтр (PSM), адаптивный

переключающийся фильтр (APSM), классический медианный фильтр (MF) и его адаптивная модификация (AMF) [102-104]. Для удаления импульсного шума со случайными значениями импульсов использовался классический медианный фильтр и его модификации SD-ROM, ROAD, DWM [3,105-106]. Результаты оценки качества восстановленных изображений усреднялись по пяти тестовым изображениям – Baboon, Barbara, Lenna, Peppers, Tank при различных вероятностях появления импульсного шума p и с использованием трех критериев качества: ПОСШ, КСП и НОКПО. При этом для метрик ПОСШ и КСП с ростом их значений увеличивается и субъективно воспринимаемое качество изображения, а для метрики НОКПО действует обратное правило. Соответствие метрик НОКПО, ПОСШ и КСП представлено в табл. 2.3 и табл. 2.4. Результаты получены на тестовых изображениях среды Piclab, сжатых по стандарту JPEG с различными значениями параметра компрессии.

Таблица 2.3

Соответствие показателей НОКПО и ПОСШ

НОКПО	ПОСШ
80, 100	24,66, 26,61
60, 80	26,7, 28,7
40, 60	28,74, 30,69
0, 40	30,78, 32,73

Таблица 2.4

Соответствие показателей НОКПО и КСП

НОКПО	КСП
80, 100	0,59, 0,67
60, 80	0,67, 0,75
40, 60	0,75, 0,83
0, 40	0,83, 0,91

Полученные зависимости для исследуемых восстанавливающих фильтров, а также для искаженных изображений без фильтрации (RAW) приведены на рис. 2.10. Для шума типа «соль и перец», при использовании метрик ПОСШ и КСП ранжир фильтров по качеству восстановления одинаков и выглядит следующим образом: APSM, AMF, PSM, MF, RAW. При использовании неэталонной метрики НОКПО порядок фильтров в большинстве диапазонов остается таким же, за исключением того, что результаты обработки фильтрами APSM и AMF становятся практически неразличимыми. Важно также заметить, что сложным для алгоритма неэталонной оценки качества является диапазон при вероятности зашумления $p>0,7$ в котором существенно нарушается статистика естественного изображения. При отсутствии восстанавливающего фильтра (кривая RAW) метрика НОКПО также неправильно различает слабые шумовые воздействия $p=0,1$ и $p=0,2$.

Эксперименты с импульсным шумом с равномерно распределенными случайными значениями импульсов показывают, что при использовании метрик ПОСШ и КСП выделяются две группы восстанавливающих фильтров по визуально воспринимаемому качеству: DWM/ROAD и MF/SD-ROM. Аналогичные выводы следуют из анализа кривых по метрике НОКПО при $p<0,7$. При больших значениях p неэталонное оценивание качества затруднено. Так же, как и в предыдущем случае остаются трудности при различении слабых шумовых воздействий $p<0,3$ на кривой RAW.

Перекрестный анализ зависимостей для разных моделей шумов при одинаковой метрике качества позволяет убедиться, что при одинаковых значениях вероятности p качество изображений восстановленных из шума типа «соль и перец» неизменно выше.

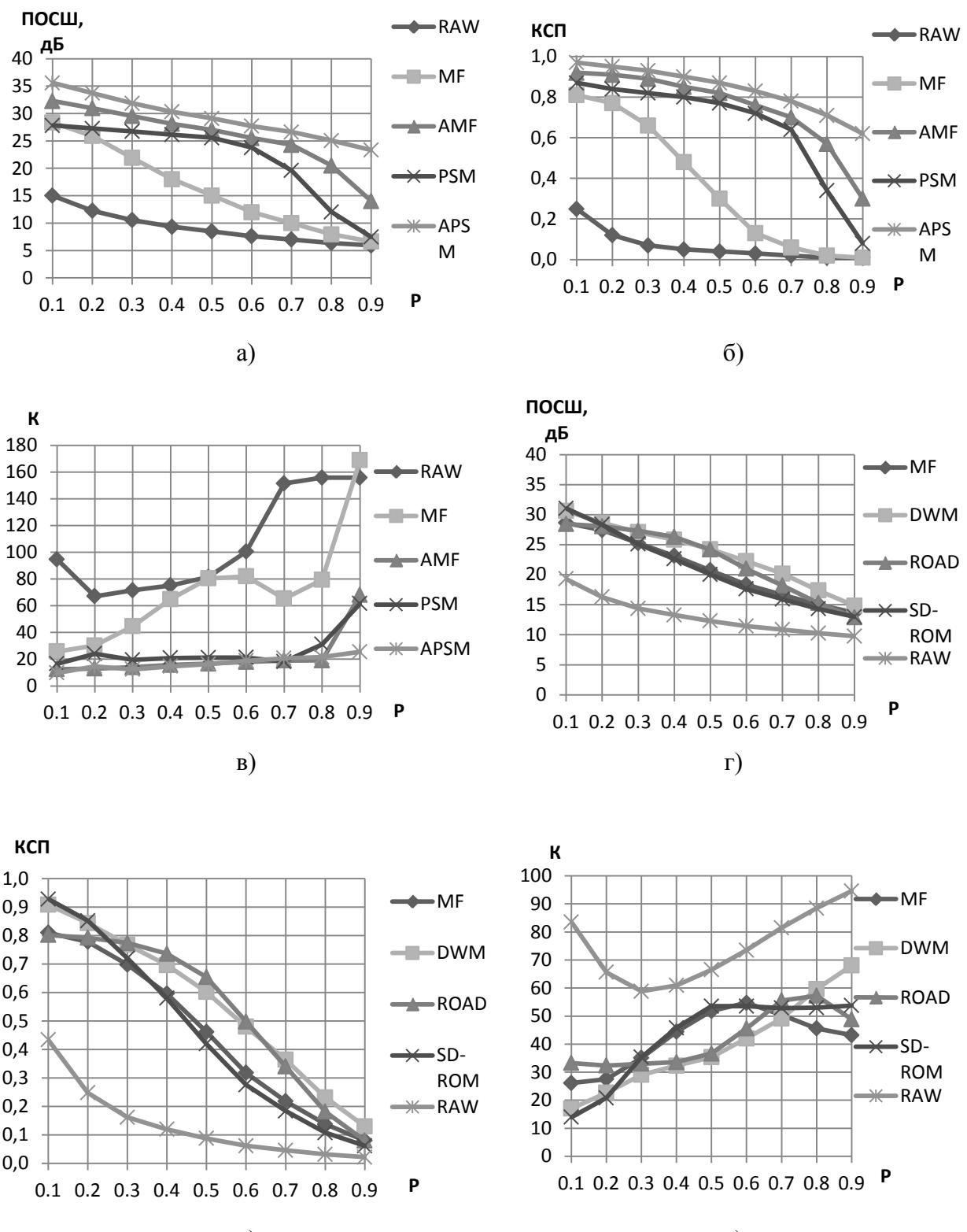


Рис. 2.10. Зависимость качества изображения от вероятности импульсного шума p :
 а) шум «соль и перец», метрика ПОСШ; б) шум «соль и перец», метрика КСП;
 в) шум «соль и перец», метрика НOKПО; г) шум со случ. значениями импульсов,
 метрика ПОСШ; д) шум со случ. значениями импульсов, метрика КСП;
 е) шум со случ. значениями импульсов, метрика НOKПО

2.8. Краткие выводы

1. На основе анализа современных научно-технических источников был выбран метод для разработки неэталонного алгоритма. В его качестве использован алгоритм на основе выделения признаков и алгоритмов машинного обучения.
2. Обоснован выбор локальных бинарных шаблонов в качестве признаков в задаче неэталонной оценки качества.
3. Произведен анализ алгоритма неэталонной оценки качества в пространственной области (НОКПО).
4. Произведено исследование алгоритма НОКПО в задаче фильтрации импульсного шума.

ГЛАВА 3. РЕЗУЛЬТАТЫ МОДЕЛИРОВАНИЯ НЕЭТАЛОННЫХ МЕТРИК ОЦЕНКИ КАЧЕСТВА ИЗОБРАЖЕНИЙ

3.1. Вводные замечания

В данной главе будет проведено многостороннее тестирование описанного подхода к неэталонному оцениванию качества изображений с использованием признаков на основе локальных бинарных шаблонов. Оно подразумевает выполнение следующих задач:

1. Разработка и тестирование алгоритма неэталонной оценки качества изображений (НОК ЛБШ-1);
2. Тестирование одномасштабных и многомасштабных локальных бинарных шаблонов в задаче неэталонной оценки качества изображений;
3. Тестирование классификаторов для изображений различного качества;
4. Разработка и тестирование модифицированного алгоритма неэталонной оценки качества (НОК ЛБШ-2) с использованием статистики естественных изображений;
5. Тестирование времени работы алгоритмов НОК ЛБШ, НОК ЛБШ-1, НОК ЛБШ-2 и НОКПО;
6. Суммарное тестирование алгоритмов на базе LIVE;
7. Суммарное тестирование алгоритмов на базе TID2013.

3.2. Тестирование различных видов классификаторов для неэталонного алгоритма оценки качества

3.2.1. Исследование для одномасштабных локальных бинарных шаблонов

В данном пункте приведены результаты исследования неэталонного алгоритма оценки качества изображений на тестовой базе изображений LIVE при использовании в качестве признаков ЛБШ с $r=1$ и числом точек в окрестности – 8. Исследование проводилось со следующими алгоритмами классификации: бустинг, машина опорных векторов, градиентный бустинг над решающими деревьями и рандомизированные деревья. Результаты приведены в табл. 3.1-3.3.

Таблица 3.1

Коэффициент корреляции Пирсона при использовании
одномасштабных ЛБШ и различных классификаторов на тестовой базе LIVE

Классификатор \ Тип искажения	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
Бустинг	0,918	0,934	0,948	0,940	0,870	0,916
Машина опорных векторов	0,851	0,928	-0,837	0,923	0,818	0,586
Градиентный бустинг	0,957	0,957	0,980	0,957	0,921	0,953
Рандомизированные деревья	0,964	0,964	0,985	0,964	0,924	0,959

Полученные результаты показывают, что при использовании одномасштабных шаблонов по всем трем метрикам наилучшие результаты показывает классификатор на основе рандомизированных деревьев. Очень близкие результаты показывает подход на основе градиентного бустинга над решающими деревьями. Применение машины опорных векторов показывает существенно большее расхождение с экспертными оценками для всех типов

искажений. Проведем аналогичные исследования для многомасштабных ЛБШ.

Таблица 3.2

Коэффициент корреляции Спирмена при использовании одномасштабных ЛБШ и различных классификаторов на тестовой базе LIVE

Классификатор	Тип искажения	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
Бустинг		0,888	0,917	0,957	0,942	0,843	0,916
Машина опорных векторов		0,882	0,922	-0,773	0,930	0,823	0,573
Градиентный бустинг		0,928	0,940	0,962	0,933	0,889	0,940
Рандомизированные деревья		0,944	0,898	0,941	0,953	0,898	0,948

Таблица 3.3

Корень из среднеквадратичной ошибки при использовании одномасштабных ЛБШ и различных классификаторов на тестовой базе LIVE

Классификатор	Тип искажения	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
Бустинг		9,92	9,80	7,72	9,40	10,98	9,70
Машина опорных векторов		17,05	18,33	18,33	15,34	16,19	19,18
Градиентный бустинг		7,09	7,05	4,28	6,67	8,52	6,97
Рандомизированные деревья		6,47	6,51	3,71	6,24	8,44	6,53

3.2.2. Исследование для многомасштабных локальных бинарных шаблонов

В данном пункте приведены результаты исследования неэталонного алгоритма оценки качества изображений на тестовой базе изображений LIVE при использовании в качестве признаков многомасштабных локальных бинарных шаблонов с радиусами окрестности 1, 2, 3 и числом точек в окрестности – 8, 16, 24, соответственно. Исследование проводилось со следующими алгоритмами классификации: бустинг, машина опорных векторов, градиентный бустинг над решающими деревьями, и рандомизированные деревья. Результаты приведены в табл. 3.4-3.6.

Таблица 3.4

Коэффициент корреляции Пирсона при использовании
многомасштабных ЛБШ и различных классификаторов на тестовой базе LIVE

Тип искажения Классификатор	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
Бустинг	0,923	0,943	0,950	0,945	0,886	0,924
Машина опорных векторов	0,809	0,884	-0,484	0,933	0,797	0,618
Градиентный бустинг	0,962	0,962	0,981	0,967	0,931	0,959
Рандомизированные деревья	0,970	0,968	0,987	0,971	0,936	0,965

Анализируя данные из табл. 3.4-3.6, можно сделать вывод о том, что, как и в случае с одномасштабными шаблонами, классификатор на основе рандомизированных деревьев показывает лучшие результаты по всем трем метрикам. Отрыв от ближайшего конкурента по метрике корня из среднеквадратичной ошибки составляет 0,48. По другим метрикам результаты немного выше, чем соответствующие результаты для

градиентного бустинга. Проведем исследования по вычислительной сложности рассмотренных алгоритмов классификации.

Таблица 3.5

Коэффициент корреляции Спирмена при использовании многомасштабных ЛБШ и различных классификаторов на тестовой базе LIVE

Тип искажения Классификатор	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
Бустинг	0,901	0,932	0,967	0,942	0,926	0,926
Машина опорных векторов	0,854	0,883	-0,364	0,931	0,796	0,567
Градиентный бустинг	0,937	0,945	0,964	0,945	0,904	0,947
Рандомизированные деревья	0,952	0,957	0,974	0,955	0,912	0,956

Таблица 3.6

Корень из среднеквадратичной ошибки при использовании многомасштабных ЛБШ и различных классификаторов на тестовой базе LIVE

Тип искажения Классификатор.	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
Бустинг	9,60	9,49	7,30	9,00	10,48	9,32
Машина опорных векторов	15,80	15,79	29,10	12,29	14,09	18,18
Градиентный бустинг	6,62	6,77	4,13	5,90	7,83	6,50
Рандомизированные деревья	5,95	6,22	3,40	5,63	7,56	6,02

3.2.3. Сравнительный анализ времени обучения

В данном пункте приведены результаты эксперимента по замеру времени обучения неэталонного алгоритма оценки качества изображений, а так же оперативной памяти необходимой для хранения обученных моделей.

В качестве признаков использовались многомасштабные локальные бинарные шаблоны с радиусами окрестности 1, 2, 3 и числом точек в окрестности 8, 16, 24 соответственно. Исследование проводилось со следующими алгоритмами классификации: бустинг, машина опорных векторов, градиентный бустинг над решающими деревьями, и рандомизированные деревья. Результаты приведены в табл. 3.7-3.8.

Таблица 3.7

Зависимость времени обучения от используемого алгоритма машинного обучения

Классификатор	Время обучения (мс)
Бустинг	431
Машина опорных векторов	86
Градиентный бустинг	2010
Рандомизированные деревья	441

Таблица 3.8

Размер модели в оперативной памяти при использовании различных алгоритмов машинного обучения

Классификатор	Размер модели (Мбайт)
Бустинг	0,62
Машина опорных векторов	1,45
Градиентный бустинг	1,68
Рандомизированные деревья	3,51

Из полученных результатов в табл. 3.1-3.8 видно, что самым точным алгоритмом неэталонной оценки качества из всех рассмотренных вариантов, является алгоритм, построенный по схеме: многомасштабные ЛБШ ($r = 1,2,3; P = 8, 16, 24$) и классификатор на основе рандомизированных деревьев. В дальнейших экспериментах такой алгоритм будем называть НОК ЛБШ-1. Данный алгоритм, требует больше оперативной памяти для хранения моделей 3,51 Мбайт, что, однако, не является серьезным недостатком в большинстве практических приложений. Вычисление

многомасштабных локальных бинарных шаблонов можно производить за один «проход» по изображению, поэтому вычислительная сложность построения гистограммы многомасштабных ЛБШ – $O(n)$. Стоит отметить, что такую же сложность, имеет алгоритм построения гистограммы одномасштабных шаблонов.

3.3. Тестирование модифицированного алгоритма неэталонной оценки качества

3.3.1. Расчет ЛБШ-признаков для различных изображений

В рассмотренном выше алгоритме НОКПО вычисление признаков происходит из преобразованного изображения. Изображение преобразуется к виду усредненных нормализованных коэффициентов контраста (УНКК-изображение). Для того что бы вычислить УНКК-изображение необходимо вычислить два изображения μ -изображение и σ -изображение (рис. 3.1).

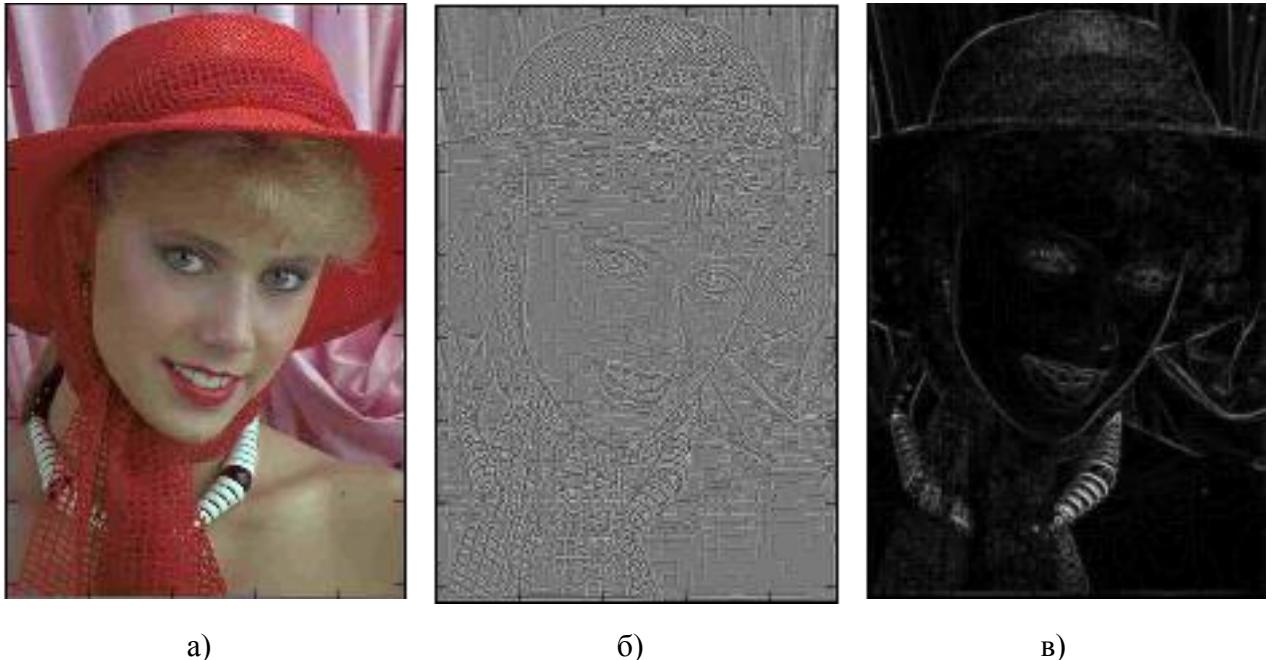


Рис. 3.1. Изображения, используемые алгоритмом НОКПО: а) оригинальное; б) нормализованные коэффициенты контраста (УНКК-изображение); в) σ -изображение

Проведем расчет ЛБШ-признаков из изображений, используемых в алгоритме НОКПО. Будем использовать многомасштабные ЛБШ с радиусами 1, 2, 3 и количеством пикселей в окрестности 8, 16, 24 соответственно. Сравним работу алгоритма НОК ЛБШ-1 с его модификациями для которых ЛБШ-признаки рассчитываются из преобразованных изображений (УНКК-изображение, σ -изображение). Результаты представлены в табл. 3.9-3.11.

Таблица 3.9

Коэффициент корреляции Пирсона при использовании различных изображений для расчета ЛБШ-признаков на тестовой базе LIVE

Тип искажения Алгоритм	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
НОК ЛБШ-1 (УНКК)	0,968	0,972	0,977	0,970	0,933	0,966
НОК ЛБШ-1 (УНКК+ σ)	0,970	0,973	0,973	0,974	0,946	0,969
НОК ЛБШ -1 (УНКК+ σ +ориг.)	0,971	0,973	0,988	0,979	0,949	0,972
НОК ЛБШ-1 (ориг.)	0,970	0,968	0,987	0,971	0,936	0,965

Таблица 3.10

Коэффициент корреляции Спирмена при использовании различных изображений для расчета ЛБШ-признаков на тестовой базе LIVE

Тип искажения Алгоритм	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
НОК ЛБШ-1 (УНКК)	0,948	0,961	0,974	0,954	0,906	0,958
НОК ЛБШ-1 (УНКК+ σ)	0,950	0,963	0,977	0,961	0,924	0,963
НОК ЛБШ -1 (УНКК+ σ +ориг.)	0,955	0,965	0,978	0,966	0,927	0,965
НОК ЛБШ-1 (ориг.)	0,952	0,957	0,974	0,955	0,912	0,956

Таблица 3.11

Корень из среднеквадратичной ошибки при использовании различных изображений для расчета ЛБШ-признаков на тестовой базе LIVE

Тип искажения Алгоритм	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
НОК ЛБШ-1 (УНКК)	6,05	5,78	3,80	5,23	7,7	5,95
НОК ЛБШ-1 (УНКК+ σ)	5,97	5,62	3,85	4,91	6,89	5,66
НОК ЛБШ-1 (УНКК+ σ +ориг.)	5,79	5,60	3,42	4,31	6,76	5,44
НОК ЛБШ-1 (ориг.)	5,95	6,22	3,40	5,63	7,56	6,02

Как видно из приведенных результатов, самым точным алгоритмом из рассматриваемых является алгоритм, использующий ЛБШ-признаки рассчитанные по трем изображениям УНКК, σ -изображению и оригинальному. Эти результаты подтверждаются для всех типов искажений, кроме белого шума. Однако следует заметить, что алгоритм НОК ЛБШ-1 (УНКК+ σ +ориг.), обладает повышенной сложностью, по сравнению с алгоритмом НОК ЛБШ-1 (ориг.). Из-за того что вычисление признаков необходимо производить сразу по трем изображениям, скорость работы алгоритма падает как минимум в три раза.

3.3.2. Вычисление смеси признаков

В работах Рудермана [62] показано, что распределение усредненных нормализованных коэффициентов контраста для естественных изображений удовлетворяет обобщенному распределению Гаусса, параметры которого представляют собой первый набор признаков алгоритма НОКПО:

$$f(x; \alpha, \sigma^2) = \frac{\alpha}{2\beta\Gamma(1/\alpha)} e^{-x^2/\beta} \left(-\left(\frac{|x|^\alpha}{\beta} \right) \right),$$

$$\beta = \sigma \sqrt{\frac{\Gamma(1/\alpha)}{\Gamma(3/\alpha)}}.$$

где α - параметр формы распределения, σ^2 – дисперсия.

Рассмотрим алгоритм, использующий смесь признаков из алгоритма НОК ЛБШ-1 и признаков алгоритма НОКПО – α и σ^2 . Назовем такой алгоритм НОК (ЛБШ + $\alpha + \sigma^2$). Результаты сравнения НОК (ЛБШ + $\alpha + \sigma^2$) с другими алгоритмами приведены в табл. 3.12-3.14.

Таблица 3.12

Коэффициент корреляции Пирсона при использовании смеси признаков на тестовой базе LIVE

Алгоритм	JPEG 2000	JPEG	Белый шум	Размытие	Битовые ошибки	Все
НОК (ЛБШ + $\alpha + \sigma^2$)	0,969	0,970	0,993	0,973	0,935	0,966
НОК ЛБШ-1 (оригинал)	0,970	0,968	0,987	0,971	0,936	0,965

Таблица 3.13

Коэффициент корреляции Спирмена при использовании смеси признаков на тестовой базе LIVE

Алгоритм	JPEG 2000	JPEG	Белый шум	Размытие	Битовые ошибки	Все
НОК (ЛБШ + $\alpha + \sigma^2$)	0,959	0,956	0,984	0,956	0,912	0,958
НОК ЛБШ-1 (оригинал)	0,952	0,957	0,974	0,955	0,912	0,956

Из приведенных результатов видно, что добавление признаков алгоритма НОКПО к алгоритму НОК ЛБШ-1 увеличивает точность последнего для изображений искаженных белым шумом, а так же для всех типов искажений по критерию корня из среднеквадратичной ошибки.

Таблица 3.14

Корень из среднеквадратичной ошибки при использовании смеси признаков на тестовой базе LIVE

Алгоритм	JPEG 2000	JPEG	Белый шум	Размытие	Битовые ошибки	Все
НОК (ЛБШ + $\alpha + \sigma^2$)	5,90	6,15	2,54	5,57	7,62	5,86
НОК ЛБШ-1 (оригинал)	5,95	6,22	3,40	5,63	7,56	6,02

3.3.3. Итоговый модифицированный алгоритм

Проведем совмещение идей расчета признаков на различных изображениях с вычислением смеси признаков. Результаты подобного моделирования представлены в 3.15-3.17.

Таблица 3.15

Коэффициент корреляции Пирсона при использовании комбинированных алгоритмов на тестовой базе LIVE

Изображения, по которым рассчитываются признаки	Тип признаков	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
УНКК-изобр. + σ -изобр. +ориг.	ЛБШ+ $\alpha + \sigma^2$	0,972	0,973	0,990	0,98	0,952	0,973
УНКК-изобр. + σ -изобр. +ориг.	ЛБШ	0,971	0,973	0,988	0,979	0,949	0,972
ориг.	ЛБШ	0,969	0,968	0,987	0,972	0,946	0,966

Таблица 3.16

Коэффициент корреляции Спирмена при использовании комбинированных алгоритмов на тестовой базе LIVE

Изображения, по которым рассчитываются признаки	Тип признаков	JPEG	JPEG	Белый шум	Размытие	Битовые ошибки	Все
УНКК-изобр. + σ -изобр. +ориг.	ЛБШ+ $\alpha+$ σ^2	0,955	0,966	0,983	0,967	0,929	0,966
УНКК-изобр. + σ -изобр. +ориг.	ЛБШ	0,955	0,965	0,978	0,966	0,927	0,965
ориг.	ЛБШ	0,952	0,957	0,974	0,955	0,912	0,956

Таблица 3.17

Корень из среднеквадратичной ошибки при использовании комбинированных алгоритмов на тестовой базе LIVE

Изображения, по которым рассчитываются признаки	Тип признаков	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все
УНКК-изобр. + σ -изобр. +ориг.	ЛБШ+ $\alpha+$ σ^2	5,74	5,59	2,91	4,26	6,54	5,31
УНКК-изобр. + σ -изобр. +ориг.	ЛБШ	5,79	5,6	3,42	4,31	6,76	5,44
ориг.	ЛБШ	5,95	6,22	3,40	5,63	7,56	6,02

Самый точный алгоритм по результатам всех исследований это алгоритм, использующий многомасштабные ЛБШ с радиусами окрестности 1, 2, 3; числом точек в окрестности 8, 16, 24 соответственно и первый набор

признаков из алгоритма НОКПО. Все признаки рассчитываются для трех изображений УНКК, σ -изображения и самого тестируемого изображения соответственно. Далее такой алгоритм будем называть НОК ЛБШ-2. Важно заметить, что признаки, заимствованные из алгоритма НОКПО увеличивают точность оценки для изображений искаженных белым шумом.

3.4. Тестирование времени работы алгоритмов оценки качества

Время работы алгоритмов НОКПО, НОК ЛБШ-1, НОК ЛБШ-2, НОК ЛБШ тестировалось на процессоре Intel Core i5 2.5 ГГц. Для алгоритмов НОКПО, НОК ЛБШ-1, НОК ЛБШ-2 указаны лучшие результаты запуска из 50 экспериментов. Для алгоритма НОК ЛБШ указано время из оригинальной статьи (рис. 3.2).

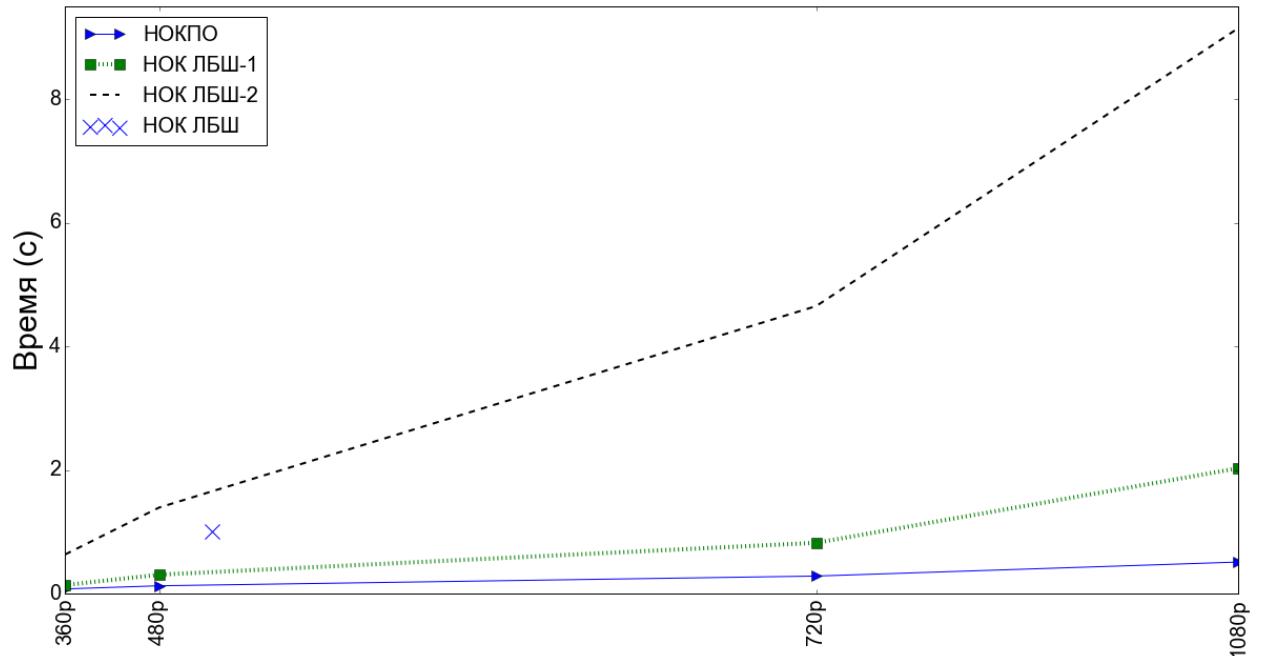


Рис. 3.2. Время работы алгоритмов НОКПО, НОК ЛБШ, НОК ЛБШ-1, НОК ЛБШ-2 для изображений различного разрешения

Приведенные результаты показывают, что время работы алгоритма НОК ЛБШ-2 при росте разрешения изображения увеличивается существенно,

поэтому использование его целесообразно только в тех приложениях, когда требуется оценка качества изображения в режиме онлайн.

3.5. Суммарное тестирование алгоритмов на базе LIVE

Проведем сравнительный анализ разработанных алгоритмов неэталонной оценки качества изображений на тестовой базе LIVE с известными из литературы 8 эталонными и неэталонными критериями. Результаты приведены в табл. 3.18-3.19. В таблицах представлены значения медианы коэффициентов корреляции, вычисленные по результатам 1000 экспериментов.

Таблица 3.18
Значения коэффициентов корреляции Пирсона для различных алгоритмов оценки качества изображений на базе LIVE

Название алгоритма	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все типы
ПОСШ (PSNR) [3]	0,903	0,876	0,917	0,780	0,880	0,859
КСП (SSIM) [45]	0,946	0,941	0,982	0,900	0,951	0,907
МКСП (MSSIM) [46]	0,979	0,974	0,988	0,918	0,948	0,954
LBIQ [47]	0,935	0,910	0,976	0,910	0,838	0,909
LD-TS [48]	0,851	0,827	0,954	0,932	0,878	0,877
GRNN [60]	0,880	0,828	0,989	0,825	0,819	0,837
НОКПО	0,973	0,923	0,985	0,951	0,903	0,942
НОК ЛБШ	0,950	0,933	0,965	0,944	0,913	0,937
НОК ЛБШ-1	0,971	0,969	0,988	0,972	0,946	0,966
НОК ЛБШ-2	0,973	0,974	0,992	0,981	0,960	0,974

Результаты показывают, что предложенный алгоритм НОК ЛБШ-2 показывает наилучшие результаты для следующих типов искажений: белый шум, размытие, а также по всей базе LIVE.

Таблица 3.19

Значения коэффициентов корреляции Спирмена для различных алгоритмов оценки качества изображений на базе LIVE

Название алгоритма	JPEG	JPEG 2000	Белый шум	Размытие	Битовые ошибки	Все типы
ПОСШ (PSNR) [3]	0,883	0,865	0,941	0,752	0,874	0,864
КСП (SSIM) [45]	0,947	0,939	0,964	0,905	0,939	0,913
МКСП (MSSIM) [46]	0,978	0,962	0,977	0,914	0,939	0,951
LBIQ [47]	0,929	0,904	0,970	0,898	0,822	0,906
LD-TS [48]	0,833	0,820	0,957	0,925	0,886	0,883
GRNN [60]	0,872	0,816	0,979	0,833	0,735	0,827
НОКПО	0,965	0,914	0,979	0,951	0,877	0,940
НОК ЛБШ	0,934	0,928	0,948	0,943	0,889	0,932
НОК ЛБШ-1	0,956	0,960	0,977	0,960	0,924	0,957
НОК ЛБШ-2	0,957	0,966	0,985	0,970	0,937	0,967

3.6. Суммарное тестирование алгоритмов на базе TID2013

Проведем сравнительный анализ разработанных алгоритмов неэталонной оценки качества на тестовой базе TID2013 с известными из литературы 10 эталонными критериями. Результаты приведены в табл. 3.20-3.21.

Таблица 3.20

Результаты работы алгоритмов НОК ЛБШ-1 и НОК ЛБШ-2

на тестовой базе TID2013

Тип искажения	Коэффициент корреляции Спирмена для НОК ЛБШ-1	Коэффициент корреляции Спирмена для НОК ЛБШ-2	Коэффициент корреляции Кендалла для НОК ЛБШ-1	Коэффициент корреляции Кендалла для НОК ЛБШ-2
Аддитивный гауссовский шум	0,68	0,89	0,48	0,74
Аддитивный шум в цветовых компонентах	0,76	0,73	0,55	0,6
Шум с пространственной корреляцией	0,80	0,86	0,63	0,66

Маскирующий шум	0,80	0,81	0,62	0,63
Высокочастотный шум	0,88	0,87	0,71	0,71
Импульсный шум	0,65	0,83	0,46	0,65
Шум квантования	0,80	0,72	0,60	0,53
Размытие по Гауссу	0,84	0,89	0,64	0,71
Искажения, возникающие в результате применения алгоритмов шумоподавления	0,82	0,88	0,60	0,72
Сжатие JPEG	0,73	0,84	0,56	0,64
Сжатие JPEG2000	0,83	0,90	0,64	0,74
Битовые ошибки в JPEG потоке	0,77	0,60	0,58	0,48
Битовые ошибки в JPEG2000 потоке	0,69	0,62	0,51	0,44
Специальный тип шума - 1 (non eccentricity pattern noise)	0,74	0,59	0,58	0,44
Локальные внутри блоковые искажения разной интенсивности	0,81	0,80	0,65	0,60
Сдвиг среднего	0,12	0,74	0,08	0,56
Изменение контрастности	0,24	0,30	0,18	0,20
Изменение цветовой насыщенности	0,84	0,85	0,67	0,67
Мультипликативный шум Гаусса	0,89	0,86	0,75	0,68
Специальный тип шума - 2 (comfort noise)	0,75	0,80	0,56	0,62
Сжатие с потерей информации	0,34	0,84	0,26	0,64
Квантование цветных изображений (image color quantization with dither)	0,81	0,87	0,61	0,69
Хроматические абберации	0,69	0,80	0,54	0,61
Специальный тип шума - 3 (sparse sampling and reconstruction)	0,84	0,92	0,65	0,77
Все типы	0,79	0,86	0,6	0,67

Таблица 3.21

Сравнительные результаты работы алгоритмов НОК ЛБШ-1 и НОК ЛБШ-2
на тестовой базе TID2013

Алгоритм	Коэффициент корреляции Спирмена	Коэффициент корреляции Кендалла
НОК ЛБШ-2	0,859	0,670
FSIMx [67]	0,851	0,667
PSNR-HA[67]	0,819	0,643
PSNR-HMA [67]	0,813	0,632
FSIM [67]	0,801	0,630
НОК ЛБШ-1	0,790	0,600
МКСП (MSSIM) [46]	0,787	0,608
PSNRc [59]	0,687	0,508
VSNR [32]	0,681	0,508
PSNR-HVS [59]	0,654	0,496
ПОСШ (PSNR) [3]	0,640	0,482
KCP (SSIM) [45]	0,637	0,470

Таким образом, проведенный анализ подтвердил преимущество разработанной метрики НОК ЛБШ-2.

3.7. Краткие выводы

Результаты моделирования неэталонных метрик и их тестирования на унифицированных тестовых базах изображений LIVE и TID2013 позволяют сформулировать следующие результаты и рекомендации.

1. Разработан и протестирован алгоритм неэталонной оценки качества на основе локальных бинарных шаблонов и классификации на основе методов машинного обучения (НОК ЛБШ-1).
2. Установлено, что использование многомасштабных локальных бинарных шаблонов позволяет увеличить коэффициент корреляции Пирсона на 0,006, коэффициент корреляции Спирмена на 0,009 и

метрику корня из среднеквадратичной ошибки на 0,51 при тестировании на базе LIVE.

3. Установлено, что использование рандомизированных деревьев оказывается наилучшим выбором из всех тестируемых алгоритмов машинного обучения. Их использование совместно с многомасштабными локальными бинарными шаблонами позволяет достичь следующих показателей для алгоритма НОК ЛБШ-1 на базе LIVE: коэффициент корреляции Пирсона – 0,966, коэффициент корреляции Спирмена – 0,957 корень из среднеквадратичной ошибки – 6,02.
4. Разработан и протестирован комбинированный алгоритм неэталонной оценки качества (НОК ЛБШ-2). Он использует смешанные признаки и несколько изображений для их расчета. Его использование позволяет достичь следующих показателей для алгоритма НОК ЛБШ-2 на базе LIVE: коэффициент корреляции Пирсона – 0,974, коэффициент корреляции Спирмена – 0,967 корень из среднеквадратичной ошибки – 5,31.
5. Тестирование времени работы алгоритмов неэталонной оценки качества показывает, что время работы алгоритма НОК ЛБШ-2 при росте разрешения изображения увеличивается существенно, поэтому использование его целесообразно только в тех приложениях, когда требуется оценка качества изображения в режиме онлайн. Для приложений реального времени рекомендуется использовать алгоритм НОК ЛБШ-1.
6. Итоговое тестирование известных эталонных и неэталонных алгоритмов на базе LIVE показывает, что предложенный алгоритм НОК ЛБШ-2 показывает абсолютно лучшие результаты для следующих типов искажений: белый шум, размытие, а также в среднем по всей базе LIVE.

7. Итоговое тестирование эталонных и неэталонных алгоритмов на базе TID2013 показывает, что предложенный алгоритм НОК ЛБШ-2 показывает абсолютно лучший результат, достигая следующих значений: коэффициент корреляции Спирмена – 0,859, коэффициент корреляции Кендалла – 0,670.

ГЛАВА 4. АНАЛИЗ ПРАКТИЧЕСКИХ ПРИМЕНЕНИЙ НЕЭТАЛОННЫХ АЛГОРИТМОВ ОЦЕНКИ КАЧЕСТВА ДЛЯ ТЕЛЕВИЗИОННЫХ ИЗОБРАЖЕНИЙ

4.1. Вводные замечания

Проведем анализ практического применения разработанных алгоритмов неэталонной оценки качества изображений в следующих задачах цифровой обработки изображений и технического зрения:

1. Оценивание качества видеопоследовательностей.
2. Оценка качества изображений лиц в задаче идентификации при использовании алгоритма распознавания на основе сверточной нейронной сети.
3. Оценка качества в задаче выбора наилучшего лица для задач биометрической идентификации.

Во всех перечисленных задачах будем использовать метрику НОК ЛБШ-1, заточенную под оценку качества изображения в режиме реального времени.

4.2. Оценка качества видеопоследовательностей

В данном разделе представлены результаты исследования на стандартной видеопоследовательности «Foreman» (рис. 4.1).

Характеристики данной видеопоследовательности:

1. Разрешение – 720x576 пикселей.
2. Частота кадров – 60 кадр/с.
3. Цветовая модель – YUV.
4. Формат субдискретизации – 4:2:2.

4.2.1. Покадровая оценка качества видео

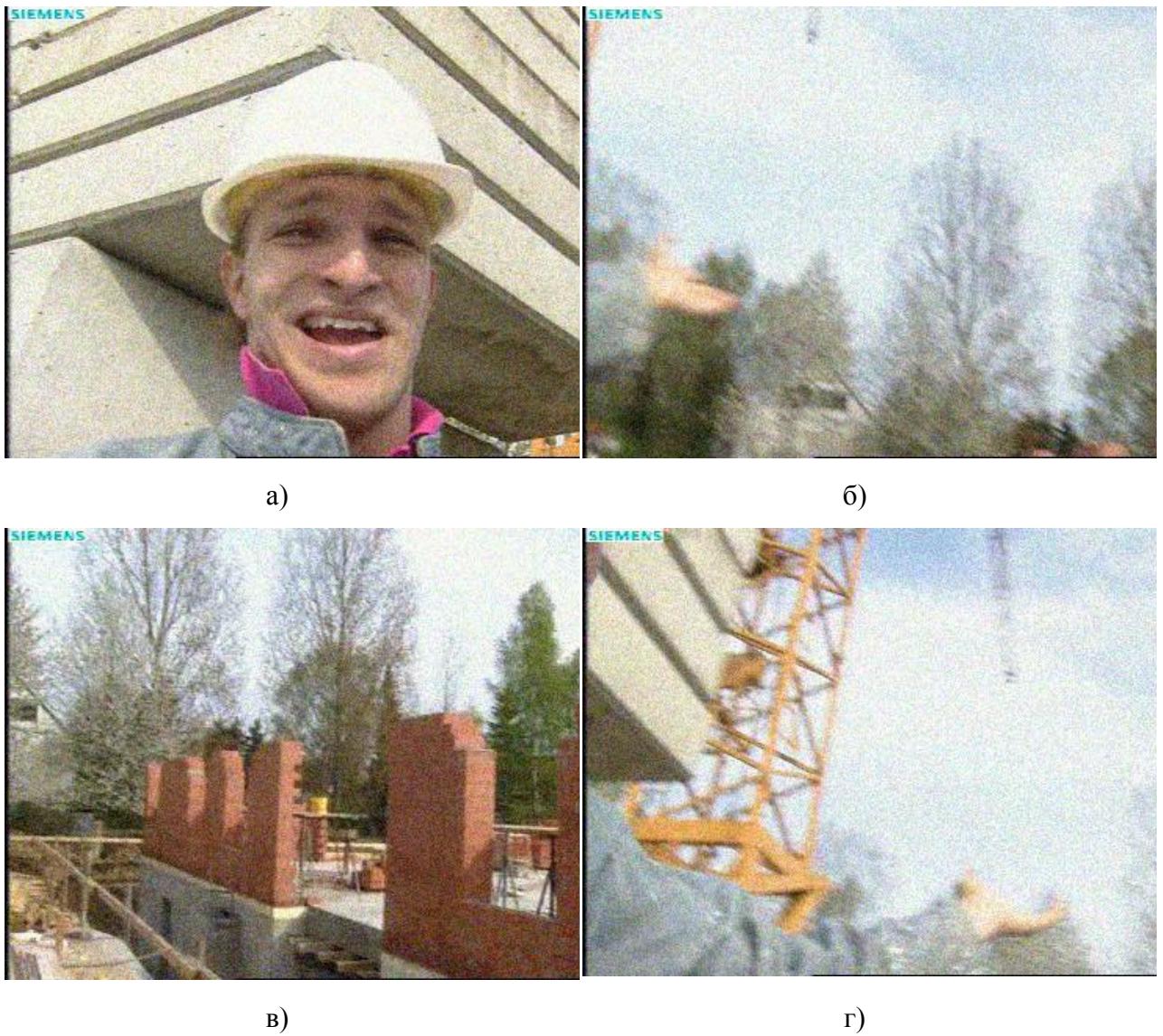


Рис. 4.1. Кадры из тестовой видеопоследовательности «Foreman»:
а) 15-й; б) 160-й; в) 170-й; г) 230-й

Покадровые значения метрик КСП (SSIM), ПОСШ и НОК ЛБШ-1 приведены на рис. 4.2. Анализ данных зависимостей показывает, что ухудшение качества входных видеоданных, наблюдаемое в районе 170-го кадра успешно отслеживается как по эталонным метрикам КСП и ПОСШ, так и по предложенной неэталонной метрике НОК ЛБШ-1.

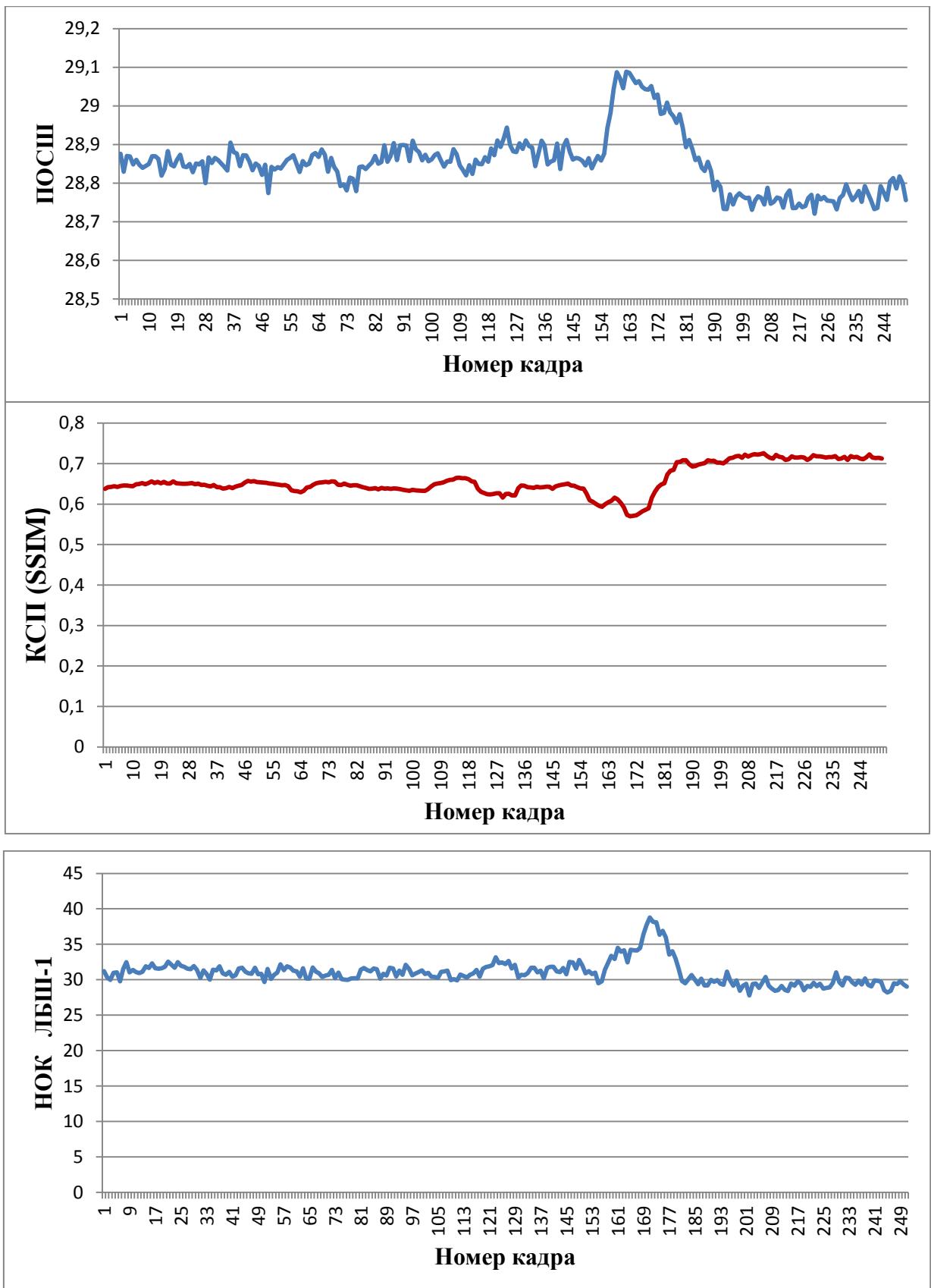


Рис. 4.2. Изменение метрик ПОСП, КСП (SSIM) и НОК ЛБШ-1 в зависимости от номера кадра для видеопоследовательности «Foreman»

На рис. 4.3 приведены кадры из другой тестовой видеопоследовательности «Football».

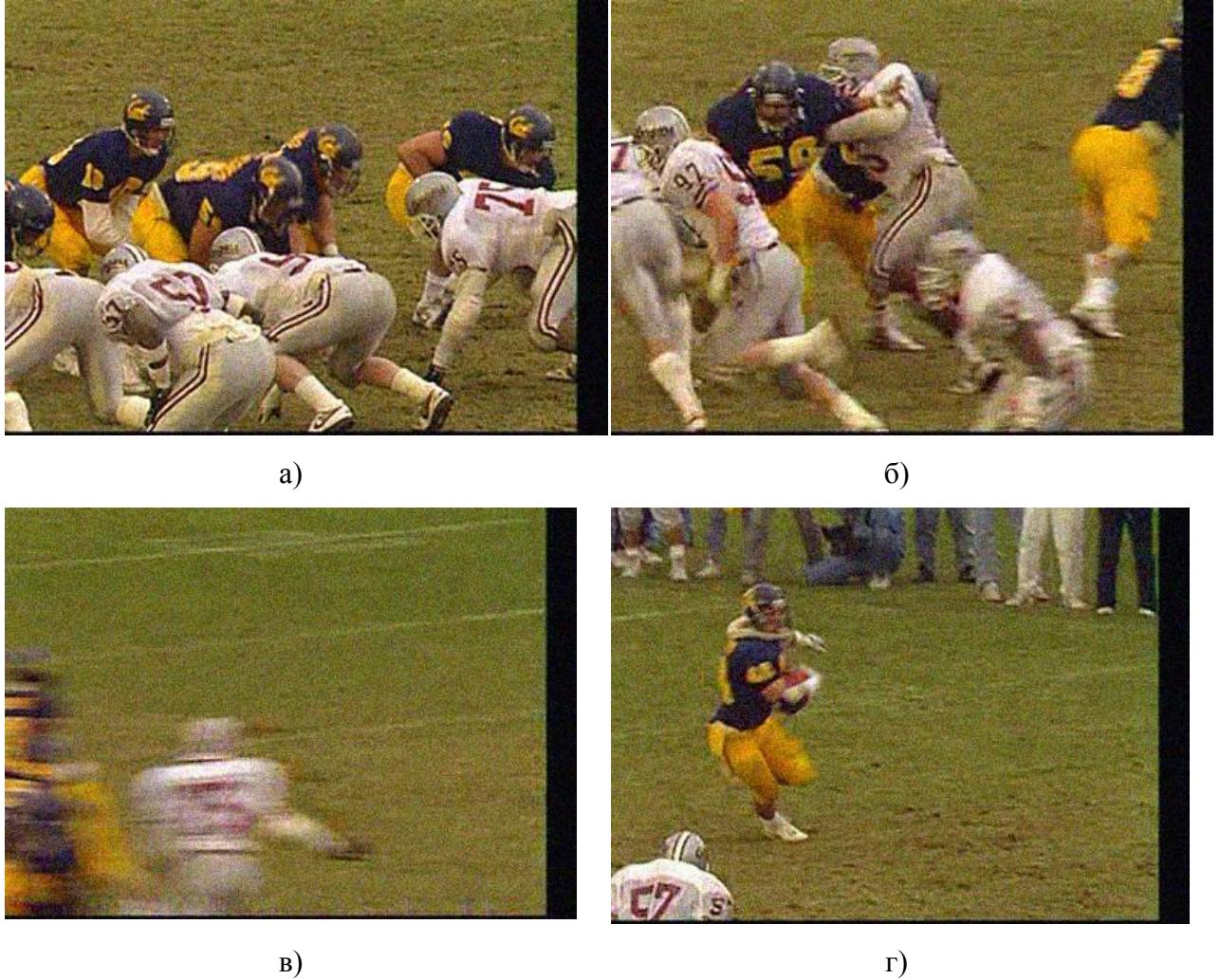


Рис. 4.3. Кадры из тестовой видеопоследовательности «Football»:

а) 1-й кадр; б) 26-й; в) 70-й; г) 103-й

На рис. 4.4 показаны покадровые значения метрик КСП (SSIM), ПОСШ и НОК ЛБШ-1. Анализ данных зависимостей показывает, что ухудшение качества в районе 70-го кадра фиксируется рассматриваемыми метриками. Изменение качества в районе 169-го кадра фиксируется метриками КСП и НОК ЛБШ-1. Значения метрики ПОСШ колеблются в узком диапазоне значений от 28,12 дБ до 28,2 дБ.

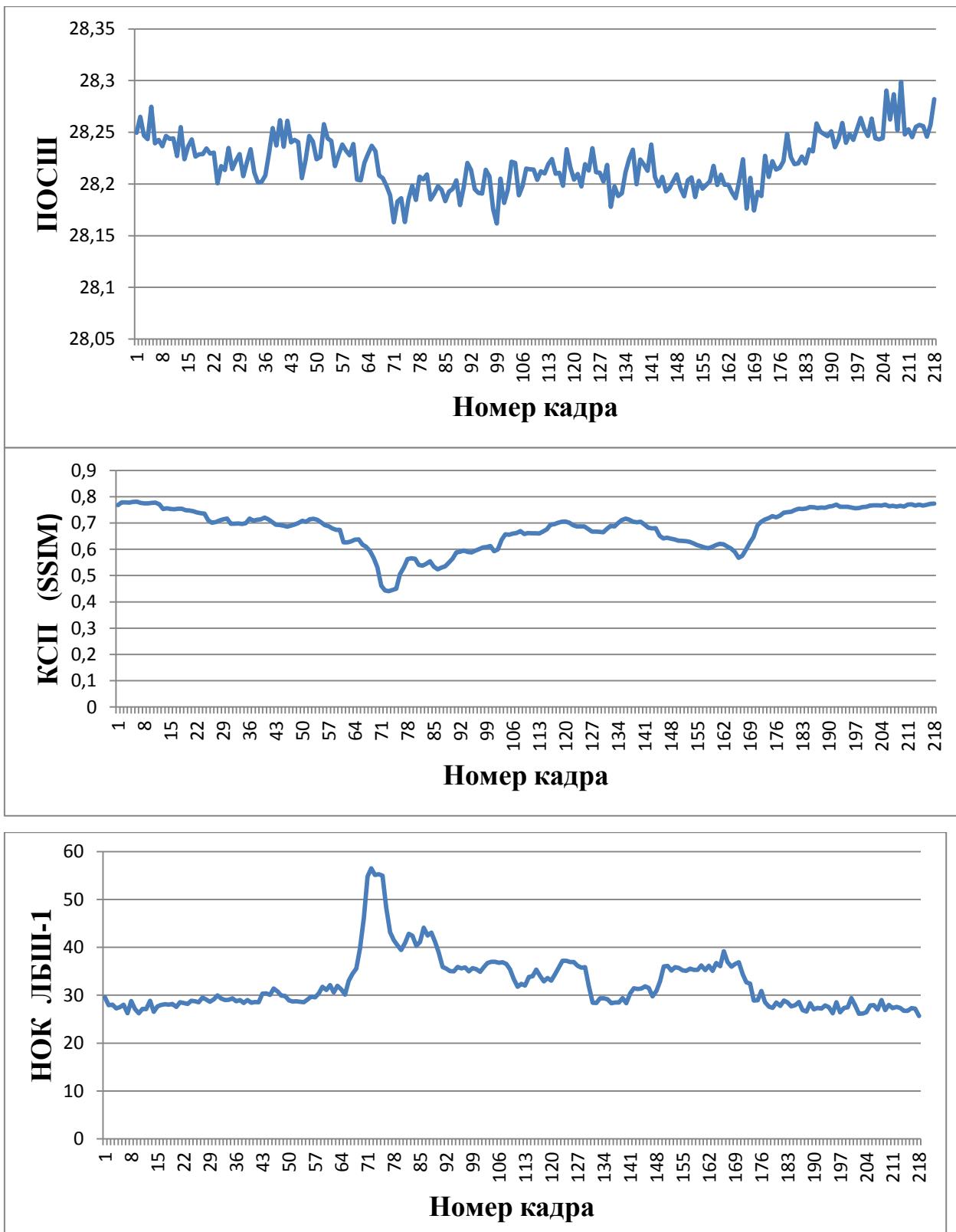


Рис. 4.4. Изменение метрик ПОСП, КСП (SSIM) и НОК ЛБШ-1 в зависимости от номера кадра для видеопоследовательности «Football»

На рис. 4.5 приведены кадры из другой тестовой видеопоследовательности «ParkJoy».



а)



б)



в)



г)

Рис. 4.5. Кадры из тестовой видеопоследовательности «ParkJoy»:

а) 3-й; б) 17-й; в) 108-й; г) 184-й

На рис. 4.6 показаны покадровые значения метрик КСП (SSIM), ПОСШ и НОК ЛБШ-1 для рассматриваемой видеопоследовательности. Все три метрики достаточно четко фиксируют скачки качества на 9-м, 96-м, 193-м кадрах.

Из анализа рис. 4.2, рис. 4.4, рис. 4.6 можно сделать вывод о том, что значения предложенной неэталонной метрики НОК ЛБШ-1 соответствуют результатам оценки качества эталонной метрикой КСП, что делает

пригодным использование первой в задачах неэталонной оценки качества видеопоследовательностей.

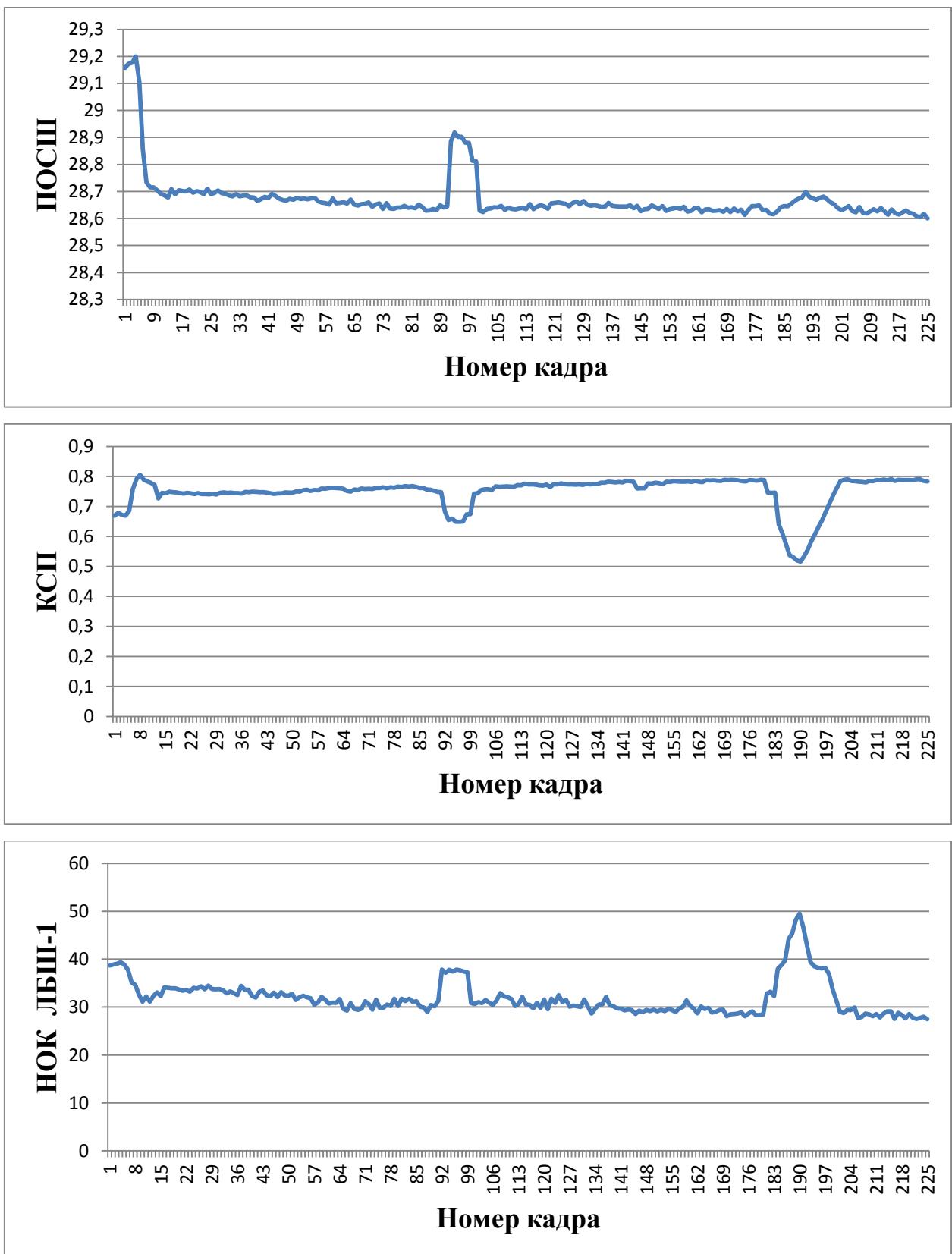


Рис. 4.6. Изменение метрик ПОСЧ, КСП (SSIM) и НОК ЛБШ-1 в зависимости от номера кадра для видеопоследовательности «ParkJoy»

4.2.2. Неэталонное оценивание качества видеоданных относительно субъективных экспертных оценок

В данном пункте представлены результаты экспериментов по неэталонной оценке качества видео метриками НОК ЛБШ-1 и НОКПО на тестовой базе ReTRiEVED [107]. Усредненные на 8-ми тестовых видеопоследовательностях результаты приведены на рис. 4.7. Их анализ показывает, что неэталонные метрики НОКПО и НОК ЛБШ-1 коррелируют с визуально воспринимаемым качеством изображения (характеризуемым средними экспертными оценками MOS). При падении оценок MOS с 3 до 1 – уровень неэталонных метрик возрастает на величину порядка 20%.

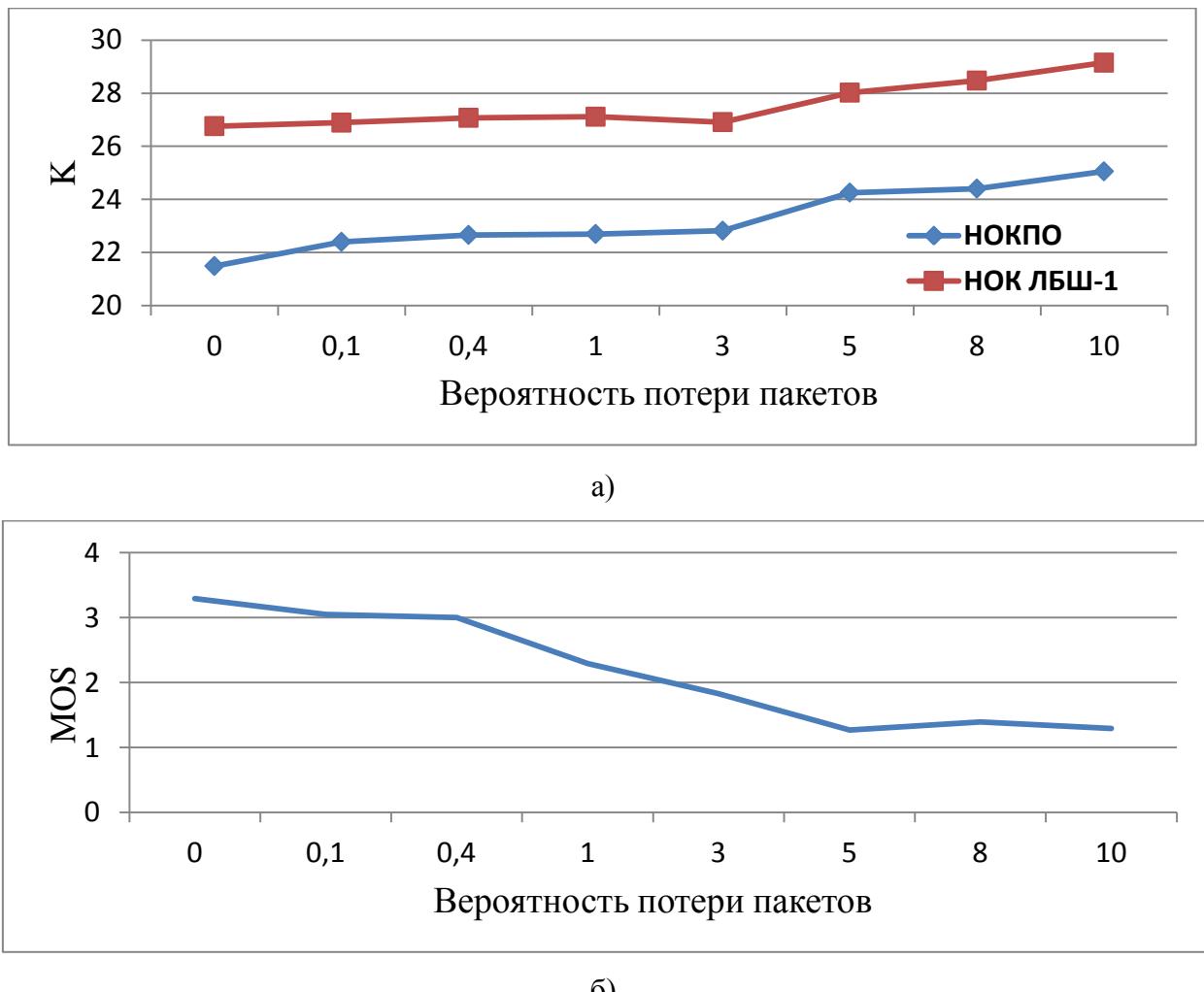


Рис. 4.7. Зависимость метрик качества от вероятности потери видео в канале:

а) НОКПО, НОК ЛБШ-1; б) экспертные оценки MOS

4.3. Оценка качества изображений лиц в задаче идентификации при использовании алгоритма распознавания на основе сверточной нейронной сети

4.3.1. Описание системы распознавания

Работу системы распознавания можно разделить на несколько этапов:

1. Детектирование лица.
2. Нормализация лица.
3. Получение представления лица с помощью глубокой сверточной сети.
4. Классификация.

Этапы работы системы проиллюстрированы на рис. 4.8.

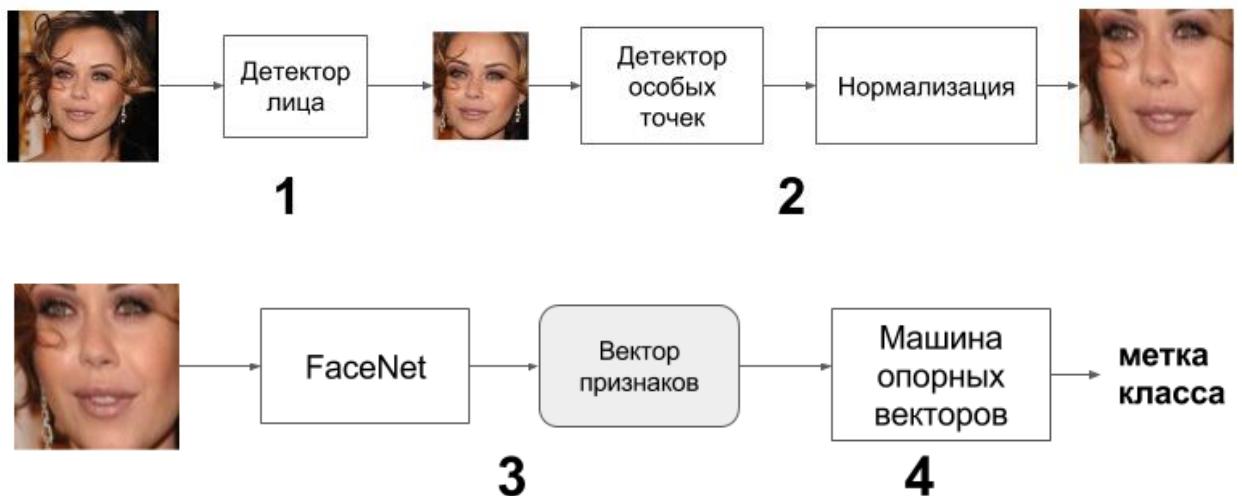


Рис. 4.8. Система распознавания лиц на основе сверточной нейронной сети

4.3.2. Детектор лиц

Используемый подход к детектированию лиц был предложен П. Виолой и М. Джонсом в работе [101]. Он использует процедуру обучения, основанную на бустинге [95]. Такой алгоритм состоит из трех этапов: переход к интегральному изображению, построение классификатора на основе бустинга, комбинирование классификаторов в каскадную структуру.

Предлагаемый алгоритм использует три вида простых признаков. Значение двухпрямоугольного признака вычисляется как разность между суммами значений пикселей, принадлежащих двум прямоугольным областям. Области имеют одинаковую длину и ширину и ориентированы вертикально или горизонтально, как показано на рис. 4.9.

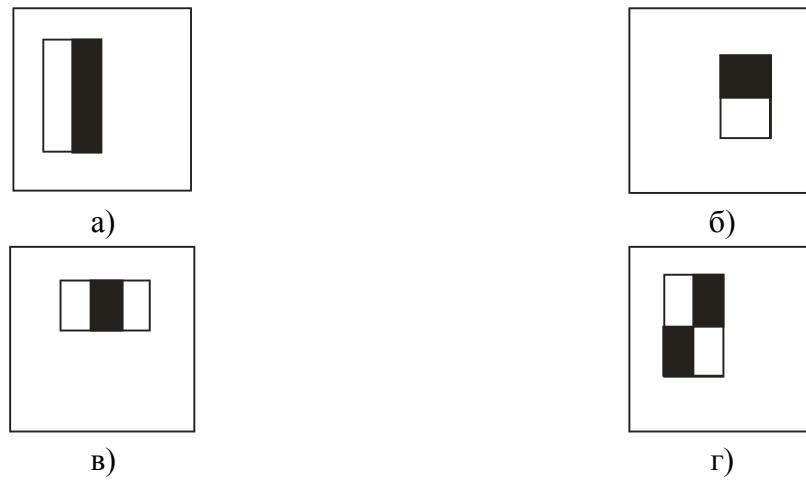


Рис. 4.9. Примеры прямоугольных признаков и их расположения внутри окна детектора: а), б) двухпрямоугольные признаки; в) трехпрямоугольный признак; г) четырехпрямоугольный признак

4.3.3. Алгоритм определения особых точек

Детектор лиц дает на выходе набор рамок, содержащих лицо. Распознавание изображений из этих рамок без предварительной обработки может привести к уменьшению точности всей системы, так как человек на изображении может находиться в разных позах и условиях освещенности. Один из способов борьбы с этой проблемой – сбор огромного числа данных, включающий в себя изображения с разными углами поворота и условиями освещенности. Другой способ снижения вариативности входных данных – нормализация лиц. Под ней понимается процесс, результатом которого является изображение лица, особые точки (глаза, нос, рот) которого находятся в заранее определенных местах. Нормализация лиц – популярная тема в области технического зрения. Многие современные алгоритмы, такие

как DeepFace [108], приводят изображение лица к нормальному виду путем построения 3D модели. В данной системе используется более простой метод, использующий двумерные аффинные преобразования входного изображения (рис. 4.10).

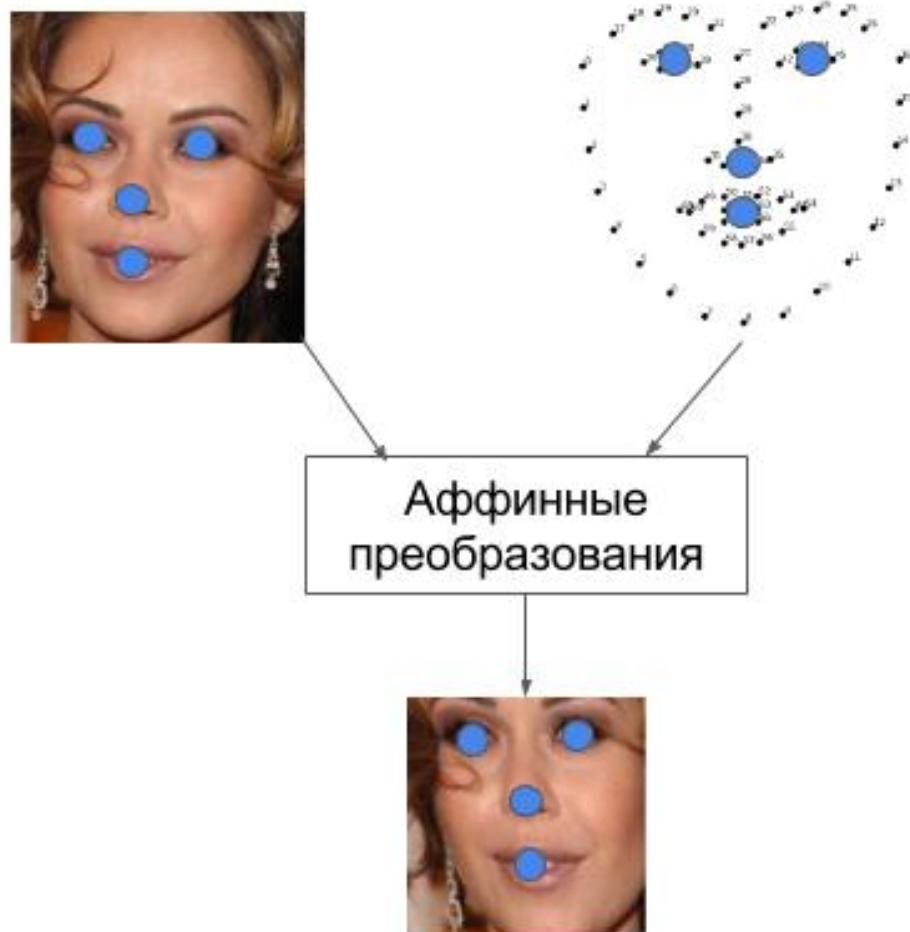


Рис. 4.10. Аффинное преобразование входного изображения

Алгоритм [109] используется для детектирования 68 особых точек лица. Нумерация особых точек представлена на рис. 4.11.

Окончательные значения краев глаз и носа на выходных изображениях вычисляются как средние значения координат особых точек глаз и носа на всем тренировочном наборе данных. Так же в ходе нормализации происходит изменение размера входного изображения: оно обрезается по крайним особым точкам. Итоговое нормализованное изображение имеет разрешение – 96x96 пикселей.

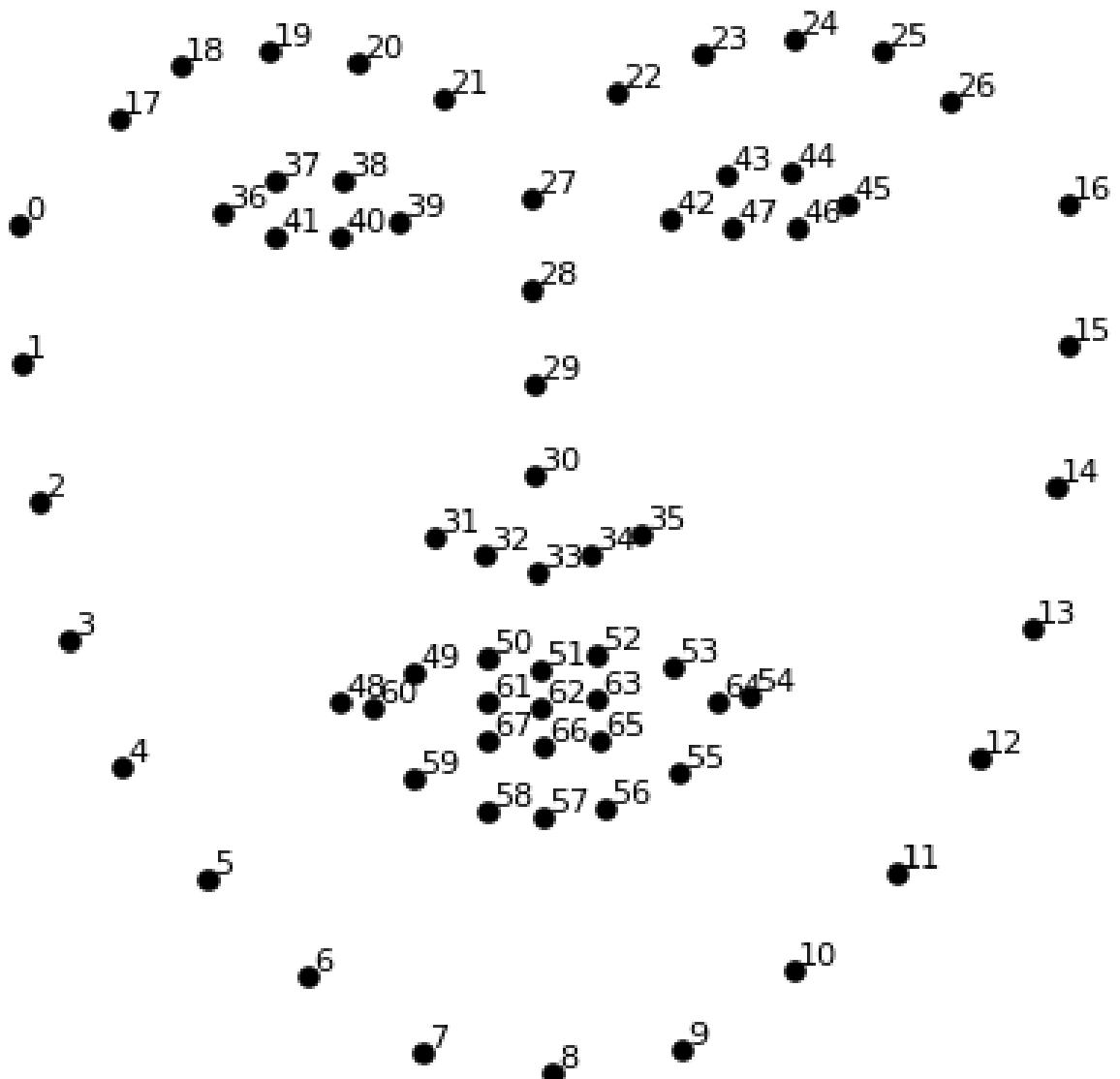


Рис. 4.11. Нумерация особых точек на лице человека

4.3.4 Получение представления лица с помощью глубокой сверточной сети

На данном этапе происходит трансформация изображения в компактный вектор признаков размерностью 128. Извлечение вектора из изображения происходит с помощью глубокой сверточной искусственной нейросети, построенной в соответствии с архитектурой из табл. 4.1.

Таблица 4.1

Архитектура сети FaceNet

Тип уровня	Размерность выходных данных
Сверточный (7x7x3)	$112 \times 112 \times 64$
Субдискретизация (макс.) + нормализация	$56 \times 56 \times 64$
Inception (2)	$56 \times 56 \times 192$
Субдискретизация (макс.) + нормализация	$28 \times 28 \times 192$
Inception (3a)	$28 \times 28 \times 256$
Inception (3b)	$28 \times 28 \times 320$
Inception (3c)	$14 \times 14 \times 640$
Inception (4a)	$14 \times 14 \times 640$
Inception (4b)	$14 \times 14 \times 640$
Inception (4c)	$14 \times 14 \times 640$
Inception (4d)	$14 \times 14 \times 640$
Inception (4e)	$7 \times 7 \times 1024$
Inception (5a)	$7 \times 7 \times 1024$
Inception (5b)	$7 \times 7 \times 1024$
Субдискретизация (усреднение)	$1 \times 1 \times 1024$
Полносвязанный	$1 \times 1 \times 128$
L2 нормализация	$1 \times 1 \times 128$

Обучение сети происходило на данных из двух публично доступных наборов лиц FaceScrub [110] и CASIA-WebFace [111], с использованием стохастического градиентного спуска в качестве оптимизатора. В качестве функции потерь выбиралась следующая функция:

$$L = \sum_i^N \|f(x_i^a) - f(x_i^p)\|_2^2 - \|f(x_i^a) - f(x_i^n)\|_2^2 + \alpha,$$

где x_i^a – тестируемое изображение; x_i^p – изображение того же человека, как и x_i^a ; x_i^n – изображение другого человека; а $f(x) \in R^{128}$ функция

рассматриваемой нейросети. Значение параметра α подбиралось в ходе процедуры обучения и равнялось 0,2.

4.3.5. Классификация

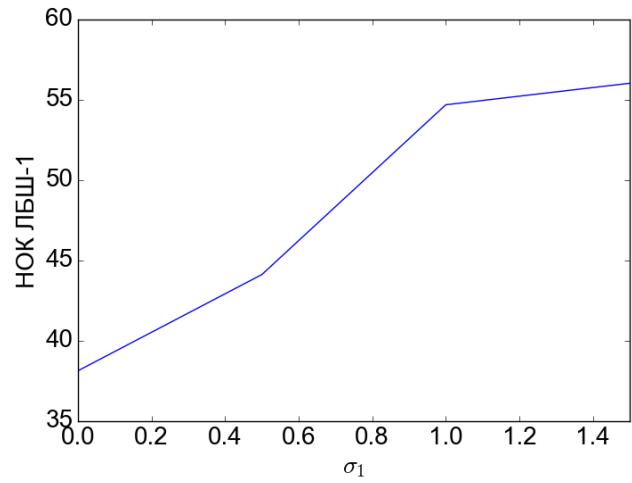
Последний этап в рассматриваемой системе распознавания – классификации. Как уже отмечалось ранее, на выходе сверточной нейросети формируется 128-ми размерное представление изображения лица. Для получения соответствующей метки класса, решается задача мультиклассовой классификации. В качестве классификатора выбрана машина опорных векторов с ядром Гаусса. Классификатор обучался на тестовой базе, состоящей из случайным образом отобранных 100 классов из базы лиц Casia Web Face. Размер тестовой базы составил 4630 изображений. На этапе обучения использовалось 80% базы, остальная ее часть отбиралась для тестирования производительности системы распознавания. В тренировочном и тестовом наборах данных не было одинаковых изображений. Разделение базы на тренировочный и тестовый наборы данных проводилось в соответствии с распределением изображений между классами. После процедуры обучения относительное число правильно классифицированных изображений составило 0,852.

4.3.6. Анализ точности распознавания лиц при наличии искажений на изображении

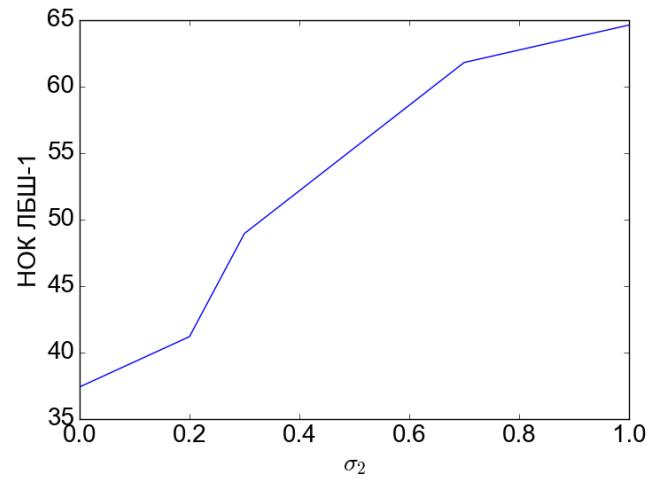
Цель проведенного ниже эксперимента: провести анализ точности системы распознавания при внесении во входные изображения различных типов искажений: размытие по Гауссу, белый шум, JPEG. Оценить корреляцию между точностью распознавания и значениями предложенной метрики НОК ЛБШ-1 для искаженных изображений.

На рис. 4.12 представлены усредненные результаты работы метрики НОК ЛБШ-1 на тестовом наборе изображений, искаженных различными типами помех.

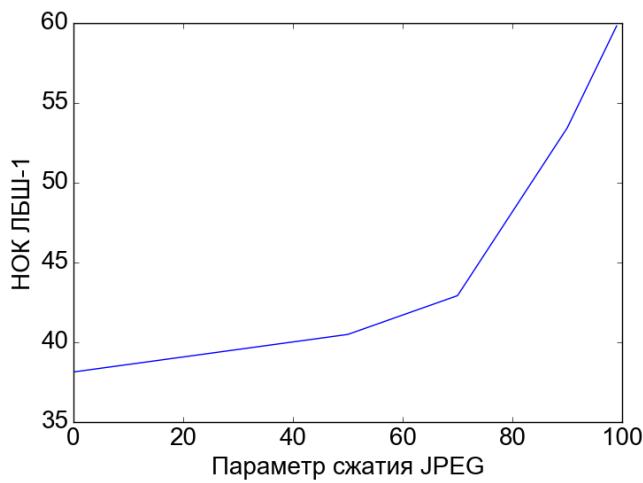
Из приведенных зависимостей видно, что среднее значение метрики НОК ЛБШ-1 для неискаженных изображений равно 37, что говорит о том, что тестовый набор уже изначально содержит изображения не идеального качества. При изменении уровня искажающих факторов значение указанной метрики меняется в диапазоне от 37 до 65.



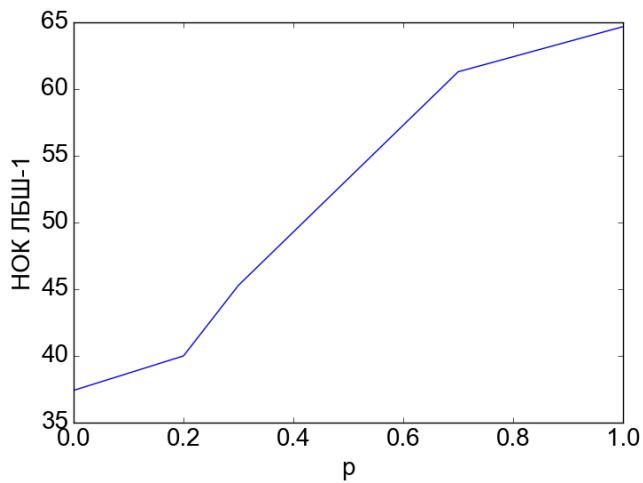
a)



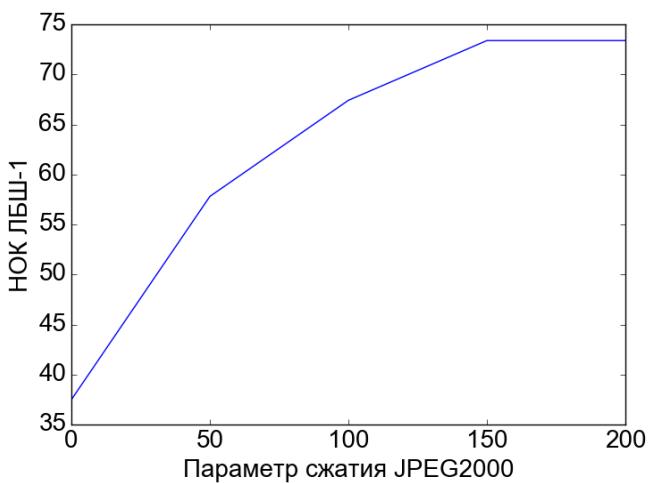
б)



в)



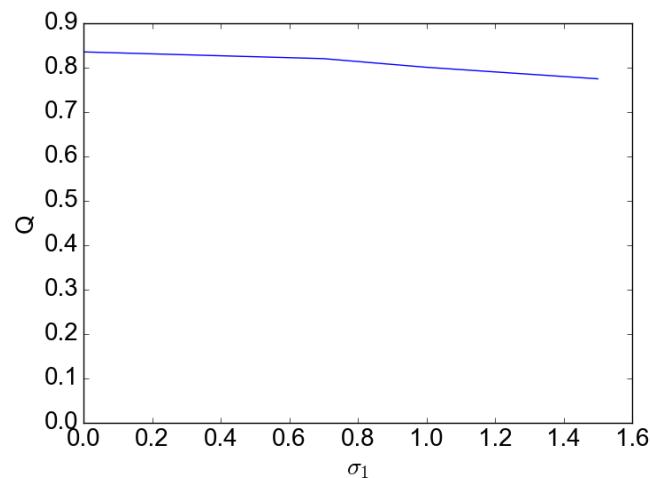
г)



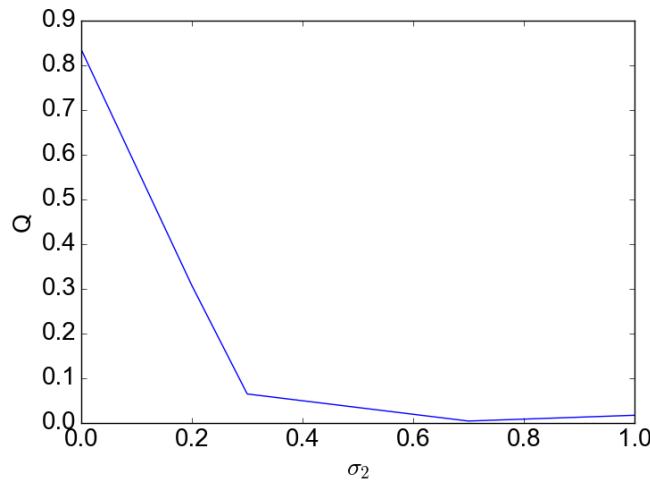
д)

Рис. 4.12. Зависимость метрики НОК ЛБШ-1 от параметров искажения:
а) размытие по Гауссу; б) белый шум; в) JPEG; г) импульсный шум; д) JPEG2000

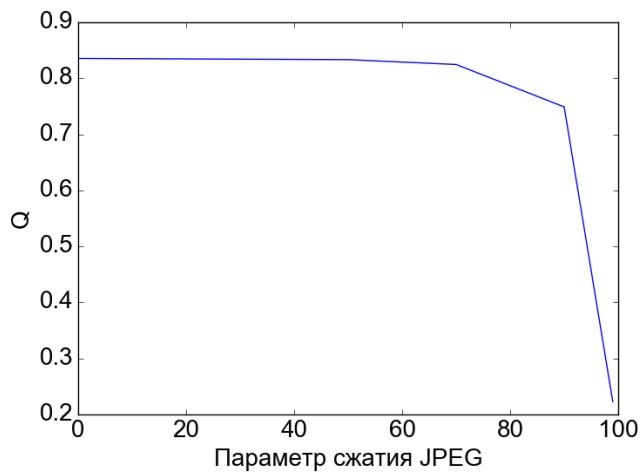
На рис. 4.13 представлены результаты точности распознавания от величины параметров искажения. Все изображения из тестового набора подвергались искажению с определенным значением параметра, определяющего силу воздействия последнего. Под точностью распознавания Q имеется ввиду отношение числа правильно классифицированных лиц к общему числу изображений лиц в тестовом наборе данных.



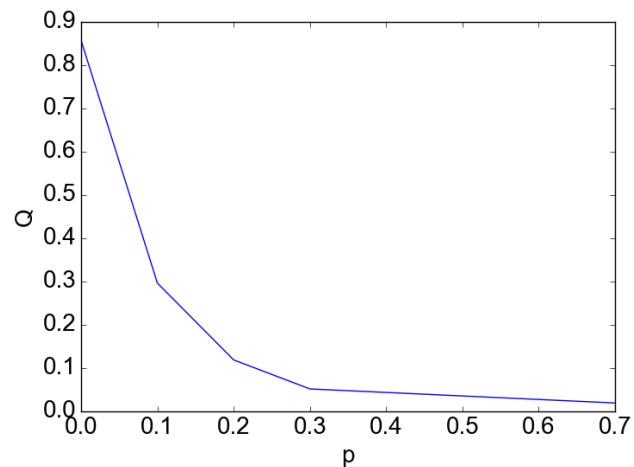
a)



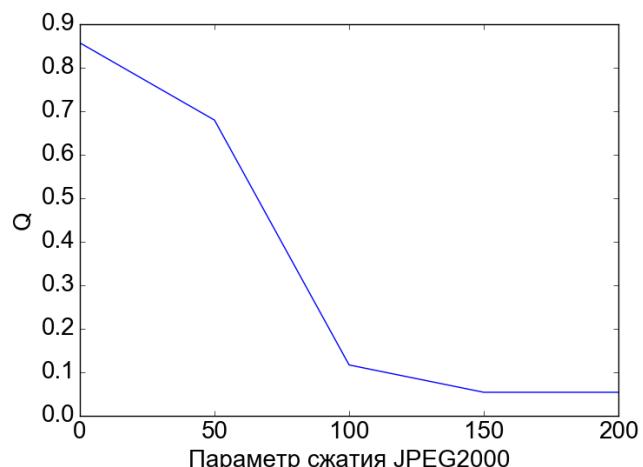
б)



б)



г)



д)

Рис. 4.13. Относительное число правильно распознанных лиц в зависимости от величины искажения: а) размытие по Гауссу; б) белый шум; в) JPEG;
г) импульсный шум; д) JPEG2000

Зависимости, приведенные на рис. 4.13 показывают, что рассматриваемая система распознавания относительно устойчива к размытию по Гауссу, о чем свидетельствует падение точности с 0,83 до 0,77.

Анализируя совместно данные рис. 4.12 и рис. 4.13 можно сделать следующие выводы:

- искажения типа размытие по Гауссу и JPEG оказывают незначительное влияние на уровень распознавания. Искажения типа белый шум, импульсный шум, JPEG2000 существенно влияют на уровень распознавания;
- значение предложенной метрики НОК ЛБШ-1 ниже 50 при всех типах искажений соответствует рабочему диапазону работы системы распознавания на основе сверточной нейронной сети.

4.4. Выбор наилучшего лица для задач биометрической идентификации на основе алгоритмов неэталонной оценки качества

Когда человек попадает в поле зрения камеры видеонаблюдения, установленной на объекте интереса, на жесткий диск, как правило, сохраняется несколько изображений его лица. Большинство из них бесполезны для системы биометрической идентификации ввиду ряда причин: движение человека в кадре вызывает размытие, человек может находиться в области низкой освещенности, в кадр может попасть только часть лица или сильно повернутое лицо. Алгоритмы, с помощью которых производится процедура идентификации человека, вычислительно достаточно сложны, поэтому распознавание всей последовательности изображений может привести к замедлению работы системы видеонаблюдения в целом [72, 73]. Таким образом, актуальной является задача выбора изображений наилучшего качества из всего полученного на практике набора лиц, по которым будет происходить идентификация человека.

Указанная проблема широко рассматривается в современной научно-технической литературе [74–82]. Одним из первых подходов к решению данной задачи являлся подход, основанный на применении алгоритма кластеризации по методу К-средних [82]. Как показали практические эксперименты – он обладает низкой точностью в том случае, когда большая часть полученного набора лиц является низкокачественной. Принципиально другой подход состоит в поиске самых качественных для распознавания лиц путем последовательной оценки качества всех изображений [26]. Типичная схема системы распознавания с модулем оценки качества изображений лиц приведена на рис. 4.14. Качество лиц здесь оценивается на этапе предобработки. Изображения низкого качества отбрасываются или архивируются, распознавание происходит только по высококачественным изображениям. В [73] показано, что использование системы распознавания с модулем оценки качества лиц приводит к значительному приросту производительности всей системы видеонаблюдения.

Существует несколько стандартов, определяющих качество изображений лиц – ISO/IEC 19794-5, ICAO 9303 [84,85], которые содержат описание характеристик, влияющих на принятие решения о пригодности изображения для систем автоматического распознавания. Все стандартизованные характеристики можно разделить на два класса: текстурные (резкость, контрастность, освещенность, степень компрессии, наличие прочих искажений) и относящиеся непосредственно к признакам лица (симметричность, поза, поворот, видимость глаз, наличие бликов или теней на лице). Для их автоматического определения в научной литературе описаны следующие методы – определение позы [87], не фронтальности освещения и поворота [86, 88].

Алгоритмы оценки качества лиц пользуются рассмотренными методами, суммируя результаты их работы с определенными весовыми коэффициентами. Например, в [80] суммируются метрики контрастности,

симметричности, разрешения и неравномерности освещения. Удельный вес каждой метрики находится методами машинного обучения. Стоит отметить, что метрика качества, полученная таким образом, будет привязана к конкретной базе обучающих изображений, а также к конкретной системе распознавания. Для решения этой проблемы разработаны принципиально другие подходы для оценки качества изображений лиц, не использующие стандартизованные характеристики, например, метод на основе статистической модели лица [77], метод на основе обучения ранжированию [78].



Рис. 4.14. Система распознавания лиц с модулем оценки качества

Ни в одной из доступных работ не используются экспертные оценки качества изображений лиц в качестве соответствующих эталонных оценок. В свою очередь экспертные оценки широко применяются при анализе качества изображений и видеопоследовательностей [37, 38]. Стоит отметить, что для человека не составляет труда определить большинство характеристик изображения лица. Поэтому задача анализа применимости субъективных экспертных оценок в качестве эталонных для задач распознавания лиц является актуальной.

Таким образом, является важным исследование ряда метрик оценки качества изображений лиц, использующихся в задаче распознавания, и сравнение результатов их работы с аналогичной экспертной оценкой.

4.4.1. Описание используемых метрик

Симметричность и неоднородность освещения

Метрика симметричности – оценка того, насколько сильно текущее положение лица отлично от фронтального и насколько неоднородным является освещение. В [80] метрика симметричности рассчитывается как пересечение соответствующих гистограмм направленных градиентов (HoG-дескрипторов) в симметричных особых точках лица:

$$d(i) = \sum_i \min(H_i^L(i), H_i^R(i)),$$

$$Sym = \frac{1}{N} \sum_{i=1}^N d(i),$$

где $H_i^L(i)$ – функция для вычисления гистограммы ориентированных градиентов в особой точке с номером i . Симметрично расположенные особые точки имеют одинаковое значение i , то есть $H_i^L(i)$ рассчитывается для i -й

левой точки. Величина Sym – значение рассматриваемой метрики симметричности, а N – число пар особых точек лица.

Четкость

Так как в реальных условиях видеонаблюдения объекты перед камерой могут быть не статичны, поэтому их изображения могут подвергаться размытию. Определение того, насколько сильно размыто изображение – одна из важных проблем в оценке качества изображений лиц. Четким изображениям должны присваиваться высокие значения данной метрики. Для вычисления метрики четкости используется модифицированный дискретный оператор Лапласа:

$$L(I) = \left| \frac{\partial^2 I}{\partial x^2} \right| + \left| \frac{\partial^2 I}{\partial y^2} \right|.$$

Дискретные вторые производные могут быть вычислены с помощью свертки с соответствующими ядрами $(1, -2, 1)$ и $(1, -2, 1)^T$. Оценка четкости производится непосредственно в области лица.

Освещенность

Метрика освещенности рассчитывается как длина R диапазона значений яркости изображения, за вычетом 5% самых темных и самых ярких пикселей, нормированная на максимально возможную длину диапазона:

$$Lum = \frac{R}{256}.$$

Симметричность особых точек

Для определения симметричности лица на изображении используются координаты особых точек, полученные детектором, описанным в [89]. Каждая особая точка имеет номер от 1 до 68. Для увеличения

производительности в предложенном алгоритме рассматриваются только 22 особые точки с номерами: 9, 31, 32, 36–49, 52, 55, 58, 63, 67.

Для определения степени отклонения лица от фронтального положения проводится прямая линия через точки 9 и 31. Далее вычисляется расстояние w между точками 37 и 46. Пусть d_i – расстояние от точки с индексом i до построенной прямой. Тогда результирующая метрика S рассчитывается по формуле:

$$S = \frac{|d_{32} - d_{36}| + |d_{40} - d_{43}| + |d_{37} - d_{46}| + |d_{49} + d_{55}| + d_{52} + d_{67}}{w}.$$

4.4.2. Результаты моделирования

Для проведения эксперимента записывались 10 тестовых видеопоследовательностей одного человека при различных условиях освещенности в 20, 50, 75, 130, 180, 500 лк. Из тестовых данных извлекались изображения двумя разными способами: экспертная выборка (10 разнокачественных изображений) и последовательная (извлекался каждый 25-й кадр). Примеры таких изображений представлены на рис. 4.15 и рис. 4.16.



Рис. 4.15. Тестовые изображения с различными искажениями:
а) поворот лица; б) размытие; в) низкое разрешение лица



Рис. 4.16. Тестовые изображения, полученные путем последовательной выборки через 25 кадров

Из каждого тестового изображения выделялись лица с использованием типового детектора лиц [90]. Далее, для каждого выделенного лица вычислялись значения следующих метрик: контрастность, разрешение, четкость, симметричность, размытие, метрика симметричности особых точек S и метрики качества K (на основе обучения ранжированию [78]). Совместно с вычислением описанных выше метрик, для каждого изображения из тестового набора определялись средние экспертные оценки со значениями от 1 (наилучшее качество) до 10 (наихудшее) включительно. Между рангом, полученным экспертом, и рангом для каждой метрики для изображений с одной освещенностью вычислялся коэффициент корреляции Спирмена. В табл. 4.2 представлены результаты измерения коэффициента ранговой корреляции Спирмена для тестовых изображений из экспертной выборки.

Таблица 4.2

Ранговая корреляция для различных метрик качества

Освещенность (лк)	K	Разрешение	Четкость	S	Симметричность	НОК ЛБШ-1	НОКПО
20	0,02	0,40	-0,17	-0,05	0,02	0,25	0,33
50	0,36	-0,03	0,03	0,05	0,15	-0,13	-0,23
75	-0,37	-0,09	-0,06	-0,15	0,23	-0,004	-0,28
130	0,1	0,36	-0,1	0,45	0,28	0,33	0,73
180	-0,03	-0,09	-0,21	0,06	0,05	0,1	-0,25
500	0,10	-0,15	-0,30	0,28	0,22	0,79	-0,06

Результаты моделирования показывают, что при низком уровне освещенности (<50 лк) лучшую корреляцию с субъективными оценками показывает метрика размытия, что вызвано низким средним качеством соответствующих видеопоследовательностей. При нормальном и высоком уровне освещенности (>130 лк) лучшую корреляцию показывает предложенная метрика на основе симметричности особых точек S . Низкую корреляцию на некоторых наборах тестовых изображениях (50 лк и 180 лк) можно объяснить тем, что работа алгоритмов определения симметричности зависит от точности детектора особых точек, которая уменьшается при сильных поворотах лица или размытии, вызванном движением человека.

Ранжир на основе метрики K в большинстве случаев имеет низкую ранговую корреляцию с экспертным ранжиром. Значения метрики K часто попадают в узкий диапазон значений, к примеру, диапазон значений для тестового набора с освещенностью 75 лк составляет 15 единиц между изображениями самого низкого и самого высокого качества, хотя весь диапазон значений метрики составляет 100 единиц.

Проведен также эксперимент по выбору трех наиболее качественных изображений (использовались тестовые наборы из последовательной выборки в 10 изображений) на основе субъективной оценки, а также на основе использования рассмотренных объективных метрик качества. Исследования проводились при различном уровне освещенности и наличии/отсутствии очков на лице человека. Результаты, представленные в табл. 4.3, показывают, что наиболее точный выбор высококачественных изображений позволяет сделать метрику на основе обучения ранжированию K . Второй результат показывает предложенная метрика на основе анализа симметричности особых точек S . Данный результат можно объяснить тем, что и субъективная оценка качества и метрика на основе анализа симметричности особых точек рассчитываются в пространственной области.

Таблица 4.3

Количество совпадений по результатам выбора трех лучших изображений

Освещенность (лк)	Наличие очков	K	Размер	Четкость	S	НОК ЛБШ-1	НОКПО	Симметричность
20	-	1	1	1	1	0	0	1
20	+	2	0	2	1	0	2	2
50	-	1	1	1	0	0	0	1
50	+	3	0	0	2	0	1	2
75	-	1	1	0	2	0	1	0
75	+	2	1	1	1	2	1	0
130	-	2	1	1	1	1	0	0
180	-	3	0	1	2	1	0	1
180	+	0	0	0	1	1	1	1
500	+	2	0	0	1	1	0	1
Всего		17	5	7	12	6	6	9

Проведено исследование ряда метрик оценки качества изображения лица применительно к задаче выбора наилучшего изображения для биометрической идентификации. Результаты моделирования показывают, что при низком уровне освещенности лучшую корреляцию с субъективными экспертными оценками показывает метрика размытия, что вызвано низким средним качеством соответствующих видеопоследовательностей. При нормальном и высоком уровне освещенности лучшую корреляцию показывает разработанная метрика на основе симметричности особых точек. В эксперименте с выбором трех изображений наилучший результат показывает метрика на основе обучения ранжированию. Полученные результаты будут полезны инженерам в области построения систем видеонаблюдения и биометрической идентификации/верификации по изображению лица.

4.5. Краткие выводы

Результаты проведенных исследований по практическому применению разработанной метрики НОК ЛБШ-1 позволяют сделать следующие выводы и дать соответствующие рекомендации:

- анализ результатов оценивания качества видеопоследовательностей показывает, что неэталонный алгоритм НОК ЛБШ-1 хорошо коррелирует с визуально воспринимаемым качеством изображения (характеризуемым средними экспертными оценками MOS). При падении оценок MOS с 3 до 1 – уровень неэталонных метрик возрастает на величину порядка 20%;
- анализ применения алгоритма НОК ЛБШ-1 в задаче распознавания показал, что искажения типа размытие по Гауссу и JPEG оказывают незначительное влияние на уровень распознавания. Искажения типа белый шум, импульсный шум, JPEG2000 существенно влияют на уровень распознавания. Значение метрики НОК ЛБШ-1 ниже 50 соответствуют при всех 5 типах искажений рабочему диапазону для работы алгоритмов распознавания на сверточной нейронной сети;
- в задаче выбора наилучшего лица для биометрической идентификации показано, что использование алгоритма НОК ЛБШ-1 целесообразно при уровне освещения лица выше 100 лк. При низком уровне освещения целесообразно использование специализированных метрик оценки качества.

ЗАКЛЮЧЕНИЕ

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

1. На основе анализа современных научно-технических источников выбран метод для разработки неэталонного алгоритма оценки качества изображений. В его качестве использован алгоритм на основе выделения признаков и алгоритмов машинного обучения. Обоснован выбор локальных бинарных шаблонов в качестве признаков в задаче неэталонной оценки качества изображений.
2. Разработан и протестирован алгоритм неэталонной оценки качества изображений на основе локальных бинарных шаблонов и классификации на основе методов машинного обучения (НОК ЛБШ-1).
3. Установлено, что использование многомасштабных локальных бинарных шаблонов позволяет увеличить коэффициент корреляции Пирсона на 0,006, коэффициент корреляции Спирмена на 0,009 и метрику корня из среднеквадратической ошибки на 0,51 при тестировании на базе LIVE.
4. Установлено, что использование рандомизированных деревьев оказывается наилучшим выбором из всех тестируемых алгоритмов машинного обучения. Их применение совместно с многомасштабными локальными бинарными шаблонами позволяет достичь следующих показателей для алгоритма НОК ЛБШ-1 на базе LIVE: коэффициент корреляции Пирсона – 0,966; коэффициент корреляции Спирмена – 0,957; корень из среднеквадратической ошибки – 6,02.
5. Разработан и исследован модифицированный алгоритм неэталонной оценки качества (НОК ЛБШ-2). Он использует смешанные признаки и несколько изображений для их расчета. Его использование позволяет достичь следующих показателей на базе LIVE: коэффициент

корреляции Пирсона – 0,974; коэффициент корреляции Спирмена – 0,967; корень из среднеквадратической ошибки – 5,31.

6. Тестирование времени работы алгоритмов неэталонной оценки качества показывает, что вычислительная сложность разработанного алгоритма НОК ЛБШ-2 при росте разрешения изображения увеличивается существенно, поэтому использование его целесообразно только в тех приложениях, где требуется оценка качества изображения в режиме онлайн. Напротив, для приложений реального времени рекомендуется использовать алгоритм НОК ЛБШ-1.
7. Итоговое тестирование известных эталонных и неэталонных алгоритмов на базе LIVE показывает, что предложенный алгоритм НОК ЛБШ-2 показывает абсолютно лучшие результаты для следующих типов искажений: белый шум, размытие, а также в среднем по всей базе LIVE.
8. Итоговое тестирование эталонных и неэталонных алгоритмов на базе TID2013 показывает, что предложенный алгоритм НОК ЛБШ-2 показывает абсолютно лучший результат, достигая следующих значений: коэффициент корреляции Спирмена – 0,859; коэффициент корреляции Кендалла – 0,670.
9. Анализ результатов оценивания качества видеопоследовательностей показывает, что разработанный неэталонный алгоритм НОК ЛБШ-1 хорошо коррелирует с визуально воспринимаемым качеством изображения (характеризуемым средними экспертными оценками MOS). При падении оценок MOS с 3 до 1 – уровень неэталонных метрик возрастает на величину порядка 20%.
10. Анализ применения алгоритма НОК ЛБШ-1 в задаче распознавания показал, что искажения типа размытие по Гауссу и JPEG оказывают незначительное влияние на уровень распознавания. Напротив, искажения типа белый шум, импульсный шум, JPEG2000 существенно влияют на уровень распознавания. Значение метрики НОК ЛБШ-1

ниже 50 соответствуют при всех 5 типах искажений рабочему диапазону для работы алгоритмов распознавания на сверточной нейронной сети.

11. В задаче выбора наилучшего лица для биометрической идентификации показано, что использование алгоритма НОК ЛБШ-1 целесообразно при уровне освещения лица выше 100 лк. При низком уровне освещения целесообразно использование специализированных метрик оценки качества.

12. Цель и задачи диссертационной работы успешно выполнены. Разработанные алгоритмы неэталонной оценки качества изображений интегрированы в программу SkifVideo.QMonitoring – приложение для неэталонной оценки качества видео, на которую получено свидетельство о государственной регистрации программы для ЭВМ.

СПИСОК ЛИТЕРАТУРЫ

1. Гонсалес Р., Вудс Р. Цифровая обработка изображений – М.: Техносфера, 2005.
2. Кравченко В.Ф. (под ред.) Цифровая обработка сигналов и изображений в радиофизических приложениях – М.: Физматлит, 2007.
3. Приоров А.Л., Апальков А.В., Хрящев В.В. Цифровая обработка изображений: Учеб. пособие / Яросл. гос. университет – Ярославль, 2007.
4. Борисов Ю.И., Дворкович В.П., Зубарев Ю.Б. Технические проблемы и перспективы внедрения цифрового телевидения и радиовещания в России. Часть 1 // Broadcasting. Телевидение и радиовещание, 2007. № 1. С. 22-27.
5. Зубарев Ю.Б., Дворкович В.П., Дворкович А.В. Проблемы и перспективы внедрения информационных мультимедийных систем в России // Электросвязь, 2004. № 10. С. 11-16.
6. Дворкович В.П., Дворкович А.В. Цифровые видеоинформационные системы (теория и практика) // Техносфера, 2012. 1008 с.
7. Минкомсвязь России: Телевидение нового поколения, <http://minsvyaz.ru/ru/activity/directions/370/>.
8. Bovik A. The Handbook of Image and Video Processing (Communication, Networking and Multimedia) // Academic Press, 2005.
9. Wang Z., Bovik A. Modern Image Quality Assessment. Morgan and Claypool Publishing Company, 2006.
10. Marr D. Vision. W.H. Freeman, 1980.
11. Илюшкина Н., Чобану М. Применение новых критериев оценки качества изображений после их сжатия с потерями // Современная электроника, 2007. № 3. С. 66-69.

12. Илюшкина Н.С., Чобану М.К. Три эталонных критерия оценки качества сжатия изображений // Доклады 10-й Междунар. конф. «Цифровая обработка сигналов и ее применение» (DSPA-2008), М., 2008. Т. 2, С. 521-524.
13. Арляпов С.А., Приоров А.Л., Хрящев В.В. Модифицированный критерий оценки качества восстановленных изображений // Цифровая обработка сигналов, 2006. №2. С. 27-33.
14. Павлов Е.А., Гущина О.Н., Приоров А.Л., Хрящев В.В. Неэталонный алгоритм оценки качества изображений с использованием дискретного косинусного преобразования // Успехи современной радиоэлектроники. 2013. № 11.
15. Приоров А.Л., Хрящёв В.В., Павлов Е.А., Герасимов Н.Б., Шемяков А.М. Оценка качества цветных изображений при подавлении импульсного шума // Радиотехника. 2013. № 5. С. 41-49.
16. Приоров А.Л., Апальков И.В., Герасимов Н.Б., Павлов Е.А. Алгоритмы оценки качества видеопоследовательностей, сжатых по стандарту Motion JPEG2000 // Успехи современной радиоэлектроники. 2012. № 10. С. 59-65.
17. Радченко Ю.С., Булыгин А.В., Радченко Т.А. Модификация индекса структурного подобия (MSSIM) на основе методов непараметрической статистики // Доклады 10-й Междунар. конф. «Цифровая обработка сигналов и ее применение» (DSPA-2008), М., 2008. Т. 2, С. 505-508.
18. Gushina O., Nenakhov I., Tarakanov A. Influence of Packets Losses on Video Quality in Case of Using Multiple Description Coding with // Proceedings of the 17th conference of FRUCT association, 2015. P. 230-236.
19. Nenakhov I., Khryashchev V., Priorov A. No-Reference Image Quality Assessment based on Local Binary Patterns // Proceedings of the IEEE East-West Design & Test Symposium (EWDTs), 2016.
20. Гущина О.Н., Ненахов И.С., Тараканов А.Н. Анализ качества изображения в системе многодескрипторного кодирования видео с

- пространственным мультиплексированием // Радиоэлектронные устройства и системы для инфокоммуникационных технологий – РЭУС-2015: сб. докл. междунар. конф. - Москва, 2015. С. 317-321.
21. Приоров А.Л., Хрящев В.В., Ненахов И.С. Исследование метода неэталонной оценки качества изображений в пространственной области // Радиоэлектронные средства получения, обработки и визуализации информации (РСПОВИ-2015): сб. докл. 3-й Всерос. конф. - Москва 2015. С. 147-151.
 22. Тараканов А.Н., Балашова О.Н., Ненахов И.С. Сравнительный анализ алгоритмов многодескрипторного видеокодирования с пространственным и временным разделением на субпотоки // Радиотехнические и телекоммуникационные системы, 2016. № 1. С. 76-86.
 23. Приоров А.Л., Ненахов И.С., Хрящев В.В. Разработка и анализ алгоритма неэталонной оценки качества изображений на основе локальных бинарных шаблонов // Успехи современной радиоэлектроники, 2016. № 7. С. 46–52.
 24. Ненахов И.С., Шемяков А.М., Хрящев В.В. Использование неэталонного алгоритма оценки качества изображений в задачах нелинейной фильтрации // Цифровая обработка сигналов и ее применение (DSPА-2016): докл. 18-й междунар. конф. – Москва, 2016. Т. 2. С. 743-748.
 25. Ненахов И.С., Хрящев В.В., Павлов В.А. Оценка качества изображения лица в задачах биометрической идентификации // Цифровая обработка сигналов и ее применение (DSPА-2016): докл. 18-й междунар. конф. – Москва, 2016. Т. 2. С. 954-959.
 26. Ненахов И.С. Разработка и анализ алгоритмов оценки качества изображений для задач распознавания // Радиоэлектронные устройства и системы для инфокоммуникационных технологий – РЭУС-2016: сб. докл. междунар. конф. - Москва, 2016. Т. 2. С. 445-450.

27. Ненахов И.С., Брюханов Ю.А., Хрящев В.В. Неэталонная оценка качества изображений на основе многомасштабных локальных бинарных шаблонов // Радиоэлектронные устройства и системы для инфокоммуникационных технологий – РЭУС-2016: сб. докл. междунар. конф. - Москва, 2016. Т. 2. С. 378-382.
28. Шемяков А.М., Хрящев В.В., Ненахов И.С., Лебедев А.А. Распознавание лиц на изображениях при потере фронтальности // Радиоэлектронные устройства и системы для инфокоммуникационных технологий – РЭУС-2016: сб. докл. междунар. конф. - Москва, 2016. Т. 2. С. 525-529.
29. Хрящев В.В., Ненахов И.С., Приоров А.Л. Использование многомасштабных локальных бинарных шаблонов в задаче оценки качества изображений // Телевидение: передача и обработка изображений: Тезисы 13-й международной конференции. - Санкт-Петербург, 2016. С. 119-123.
30. Pietikainen M., Hadid A., Zhao G., Ahonen T. Computer vision using local binary patterns. Springer, 2011.
31. Ojala T., Pietikainen M., Maenpaa T. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns // IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002. V. 24, № 7. P. 971–987.
32. Lin W. Pattern Masking Estimation in Image with Structural Uncertainty // IEEE Transactions on Image Processing, 2013. V. 12, № 4. P. 892-904.
33. Ahonen T., Hadid A., Pietikäinen M. Face description with local binary patterns: Application to face recognition // IEEE Trans. Pattern Anal. Mach. Intell, 2006. V 28, № 12. P. 2037–2041.
34. Ojala T., Pietikäinen M., Harwood D. A comparative study of texture measures with classification based on feature distributions // Pattern Recognition, 1996. V 29, № 1. P. 51–59.
35. Zhang C., Zhang Z. Boosting-Based Face Detection and Adaptation. Morgan & Claypool Publishers, 2010.

36. Ojala T., Pietikäinen M., Mäenpää T. Gray Scale and Rotation Invariant Texture Classification with Local Binary Patterns // Lecture Notes in Computer Science, 2003. V. 1842. P. 404–420.
37. Mohammadi P., Ebrahimi A., Shirani S. Subjective and Objective Quality Assessment of Image: A Survey // Elsevier Journal of Visual Communication and Image Representation, 2014. P. 1–50.
38. Recommendation ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures, 2002.
39. Egiazarian K., Jaakko A., Ponomarenko N., Lukin V., Battisti F., Carli M. Two new full-reference quality metrics based on HVS // Proc. Of the 2nd int. workshop on video process. and quality metrics for consumer electronics, VPQM, 2006.
40. Chandler D., Hemami S. VSNR: a wavelet-based visual signal-to-noise ratio for natural images // IEEE Trans. Image Process. 2007. V. 16, № 9. P. 2284-2298.
41. Saad M., Bovik A., Charrier C. Blind image quality assessment: A natural scene statistics approach in the DCT domain // IEEE Trans. Image Process, 2012. V. 21, № 8. P. 3339–3352.
42. Tong H., Li M., Zhang H., Zhang C. Learning no-reference quality metric by examples // Proc. 11th Int. Multimedia Modelling Conf., 2005. P. 247–254.
43. Ye P., Doermann D. No-reference image quality assessment using visual codebook // IEEE International Conference on Image Processing., 2012. V. 21, № 7. P. 3129–3138.
44. Sheikh H., Sabir M., Bovik A. A statistical evaluation of recent full reference image quality assessment algorithms // IEEE Trans. Image Process., 2006. V. 15, № 11. P. 3440-3451.
45. Wang Z., Bovik A., Sheikh H. Image quality assessment: From error visibility to structural similarity // IEEE Transactions on Image Processing., 2004. V. 13, № 4, P. 600–612.

46. Wang Z., Simoncelli E., Bovik A. Multi-scale structural similarity for image quality assessment // IEEE Asilomar Conf. Signals, Systems, and Computers, 2003. P. 1398–1402.
47. Tang H., Joshi N., Kapoor A. Learning a blind measure of perceptual image quality // International Conference on Computer Vision and Pattern Recognition, 2011. P. 305–312.
48. Gao X., Li X., Tao D. Universal no reference image quality assessment metrics based on local dependency // IEEE The First Asian Conference on Pattern Recognition., 2011. P. 298–302.
49. Zhang M., Xie J., Zhou X., Fujita H. No reference image quality assessment based on local binary pattern statistics // Vis. Commun. Image Process., 2013. P. 1–6.
50. Zhang M., Muramatsu C., Zhou X., Hara T., Fujita H. Blind Image Quality Assessment Using the Joint Statistics of Generalized Local Binary Pattern // IEEE signal processing letters., 2015. V. 22, № 2 P. 207-210.
51. Wang Z., Bovik A., Mean squared error: Love it or leave it? A new look at signal fidelity measures // IEEE Signal Processing Magazine., 2009. V. 26, № 1. P. 98–117.
52. Girod B. What's wrong with mean squared error? // Visual Factors of Electronic Image Communications, 1993. P. 207–220.
53. Wang Z., Bovik A., Lu L. Why is image quality assessment so difficult? // Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing, 2002. V. 4. P. 3313–3316.
54. Chen M., Bovik A. No reference Image Blur Assessment Using Multiscale Gradient // 1st International Workshop on Quality of Multimedia Experience, 2009.
55. Marziliano P., Dufaux F., Winkler S., Ebrahimi T. Perceptual blur and ringing metrics: Application to JPEG2000 // Sig. Proc.: Image Comm., 2004. V. 19, № 2. P. 163–172.

56. Wang Z., Sheikh H., Bovik A. No-reference perceptual quality assessment of JPEG compressed images // IEEE Int'l Conf. Image Proc., 2002. P. 477-480.
57. Ong E., Lin W., Lu Z., Yao S., Yang X., Jiang L. No-reference JPEG-2000 image quality metric // International Conference on Multimedia and Expo, 2003. V. 1. P. 6-9.
58. Olsen S. Estimation of noise in images: An evaluation // Comput. Vision Graphics Image Process., 1993. V.55, № 4. P. 319-323.
59. Immerkaer J. Fast Noise Variance Estimation // Computer Vision and Image Understanding, 1996. V.64, № 2. P. 300-302.
60. Li C., Bovik A., Wu X. Blind image quality assessment using a general regression neural network // IEEE Trans. Neural Networks. 2011. V. 22. P. 793–799.
61. Mittal A., Moorthy A., Bovik A. No-reference image quality assessment in the spatial domain // IEEE Trans. Image Process. 2012. V. 7, № 12. P. 4695–4708
62. Ruderman D. The statistics of natural images // Network: Computation in Neural System, 1994. V. 5, № 4. P 517–548.
63. Sharifi K., Leon-Garcia A. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video // IEEE Transactions on Circuits and Systems for Video Technology, 1995. V. 5, № 1. P. 52–56.
64. Lasmar N., Stitou Y., Berthoumieu Y. Multiscale skewed heavy tailed model for texture analysis // Image Processing, IEEE International Conference, 2009. P. 2281–2284.
65. Sheikh H., Wang Z., Cormack L., Bovik A. LIVE image quality assessment database release 2, 2006. <http://live.ece.utexas.edu/research/quality/subjective.htm>.
66. Ponomarenko N. TID2008 – a database for evaluation of full-reference visual quality assessment metrics. 2008. <http://www.ponomarenko.info/tid2008.htm>.

67. Ponomarenko N., Jin L., Ieremeiev O., Lukin V., Egiazarian K., Astola J., Vozel B., Chehdi K., Carli M., Battisti F., Jay Kuo C., Image database TID2013: Peculiarities, results and perspectives // Signal Processing: Image Communication, 2015. V. 30, P. 57-77.
68. Kodak database, <http://r0k.us/graphics/kodak/>
69. Кобзарь А. И. Прикладная математическая статистика. – М.: Физматлит, 2006.
70. LeCun Y., Kavukcuoglu K., Farabet C. Convolutional networks and applications in vision. In Circuits and Systems (ISCAS) // Proceedings of 2010 IEEE International Symposium, 2010. P. 253–256.
71. LeCun Y., Bottou L., Bengio Y., Haffner P., Gradient Based Learning Applied to Document Recognition // Proceedings of IEEE, 1998. V. 86, № 11. P. 2278-2324.
72. Zhao W., Chellappa R., Phillips P., Rosenfeld A. Face recognition: A literature survey // ACM Computing Surveys, 2003. V. 35, № 4. P. 399–458.
73. Ozay N., Tong Y., Frederick W., Liu X. Improving face recognition with a quality-based probabilistic framework // In Computer Vision and Pattern Recognition Biometrics Workshop, 2009. P. 134–141.
74. Zhu X., Ramanan D. Face detection, pose estimation and landmark localization in the wild // Proc. of the IEEE Conference on Computer Vision and Pattern Recognition, 2012. P. 2879–2886.
75. Gao X., Li S. Z., Liu R., Zhang P. Standardization of face image sample quality // Proc. Int. Conf. Biometrics, 2007. P. 242–251.
76. Nasrollahi K., Moeslund T.B. Face quality assessment system in video sequences // BIOID, Lecture Notes in Computer Science, 2008. V. 5372. P. 10–18.
77. Wong Y., Chen S., Mau S., Sanderson C., Lovell B.C. Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition // Computer Vision and Pattern Recognition (CVPR) Workshops, 2011. P. 74–81.

78. Chen J., Deng Y., Bai G., Su G. Face Image quality assessment based on learning to rank // IEEE Signal Processing Letters, 2015. V. 22, № 1. P. 90–94.
79. Chen J., Yang C., Deng Y., Zhang G., Su G. Exploring facial asymmetry using optical flow // IEEE Signal Processing Letters, 2014. V. 21, № 7. P. 792–795.
80. Nikitin M., Konushin A., Konushin V. Face quality assessment for face verification in video // Proceedings of GraphiCon'2014, 2014. P. 111–114.
81. Hadid A., Pietikainen M. From still image to video-based face recognition: An experimental analysis // In Proc. Automatic Face and Gesture Recognition, 2004. P. 813–818.
82. Berrani S., Garcia C. Enhancing face recognition from video sequences using robust statistics // In IEEE International Conference on Video and Signal Based Surveillance, 2005. P. 324–329.
83. Никитин А.Е., Хрящев В.В., Приоров А.Л., Матвеев Д.В. Разработка и анализ алгоритма распознавания лиц на основе локальных квантованных шаблонов // Нелинейный мир, 2014. № 8. С. 35–42.
84. ISO/IEC 19794-5 (published version). Information technology – Biometric Data Interchange Formats, 2005.
85. Machine readable travel documents. International Civil Aviation Organization, 2006.
86. Sang J., Lei Z., Li S. Z. Face image quality evaluation for ISO/IEC standards 19794-5 and 29794-5 // In ICB, Lecture Notes in Computer Science (LNCS). 2009. V. 5558. P. 229–238.
87. Yang Z., Ai H., Wu B., Lao S., Cai L. Face pose estimation and its application in video shot selection // In International Conference on Pattern Recognition (ICPR), 2004. P. 322–325.
88. Zhang G., Wang Y. Asymmetry-based quality assessment of face images // In ISVC, Lecture Notes in Computer Science, 2009. V. 5876. P. 499–508.

89. Feng Z., Huber P., Kittler J., Christmas W., Wu X.J. Random cascaded-regression copse for robust facial landmark detection // IEEE Signal Processing Letters, 2015. V. 22, № 1. P. 76–80.
90. Howse J. OpenCV Computer vision with Python. Packt Publishing Ltd., 2013.
91. Geurts P., Ernst D., Wehenkel L. Extremely randomized trees // Machine Learning, 2006. V. 36, № 1. P. 3–42.
92. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M., and Duchesnay E. Scikit-learn: machine learning in python // Journal of Machine Learning Research. 2011. V 12. P. 2825–2830.
93. Cortes C, Vapnik V. Support-vector networks. Machine Learning, 1995. V. 20. P. 273–297.
94. Schölkopf B., Smola A. New support vector algorithms // Neural Computations, 2000. V. 12, P. 1207–1245.
95. Freund Y., Schapire R., A Decision-Theoretic Generalization of on-Line Learning and an Application to Boosting // Journal of Computer and System Sciences, 1995. P. 119–139.
96. Ducker H. Improving Regressors using Boosting Techniques // Machine Learning: Proceedings of the Fourteenth International Conference, 1997. P. 107–115.
97. Friedman J. Greedy Function Approximation: A Gradient Boosting Machine, // The Annals of Statistics, 2001. V. 29, № 5. P. 1189–1232.
98. Phillips P.J., Wechsler H., Huang J., Rauss P. The FERET database and evaluation procedure for face recognition algorithms // Image and Vision Computing, 1998.
99. Turk M., Pentland A. Face Recognition Using Eigenfaces // Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1991. P. 586-591.

100. Pentland A., Moghaddam B., Starner T. View-Based and Modular Eigenspaces for Face Recognition // Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1994. P. 84-91.
101. Viola P., Jones M. Rapid object detection using a boosted cascade of simple features // Proc. Int. Conf. on Computer Vision and Pattern Recognition, 2001. No 1. P. 511-518.
102. Бухтояров С.С., Приоров А.Л., Апальков И.В., Хрящев В.В. Переключающийся медианный фильтр с блоком предварительного детектирования // Цифровая обработка сигналов, 2006. № 4. С. 2-8.
103. Wang Z., Zhang D. Progressive switching median filter for the removal of impulse noise from highly corrupted images // IEEE Trans. Circuits Systems II, 1999. V. 46, №1, P. 78-80.
104. Dong Y., Hu S. A new directional weighted median filter for removal of random-valued impulse noise // IEEE Signal Processing Letters, 2003. V. 14, №3. P. 193-196.
105. Abreu E., Mitra S. A signal-dependent rank ordered mean filter-a new approach for removal of impulses from highly corrupted images // IEEE Acoustics, Speech, and Signal Processing, 1995. V.4. P. 2371–2374.
106. Garnett R., Huegerich T., Chui C., He W. A universal noise removal algorithm with an impulse detector // IEEE Transactions on Image Processing, 2005. V. 14, P. 1747–1754.
107. Paudyal P., Battisti F., Carli M. A study on the effects of quality of service parameters on perceived video quality // 5-th European Workshop on Visual Information Processing, 2014. P. 1-6.
108. Taigman Y., Yang M., Ranzato M., L. Wolf. Deepface: Closing the gap to human-level performance in face verification // Computer Vision and Pattern Recognition, 2014. P. 1701-1708.
109. King D. Dlib-ml: A machine learning toolkit // The Journal of Machine Learning Research, 2009. V. 10. P. 1755–1758.

110. Ng H.-W., Winkler S. A data-driven approach to cleaning large face datasets // Proc. IEEE International Conference on Image Processing, 2014. P. 27-30.
111. Yi D., Lei Z., Liao S., Li S. Learning Face Representation from Scratch // arXiv preprint, 2014. <http://arxiv.org/abs/1411.7923>.

ПРИЛОЖЕНИЕ 1. ТЕСТОВАЯ БАЗА RETRIEVED

Данная база была получена из набора исходных видеопоследовательностей различного содержания и длительности. Восемь исходных цветных видеопоследовательностей включают в себя изображения людей, животных и искусственных объектов. Характеристики исходных видео представлены в табл. П.1.

Таблица П.1

Характеристики исходных видеопоследовательностей

Название	Разрешение	Частота кадров	Длительность (с)
Crowdrun	704x576	25	9
Duckstakeoff	704x576	25	9
Harbour	704x576	30	9
Ice	704x576	30	7
Parkjoy	704x576	25	8
Soccer	704x576	30	7
Running	720x576	25	9
Restaurant	720x576	25	8

Всего в базе 184 искаженных тестовых видео с соответствующими субъективными оценками MOS. Для внесения искажений использовался модуль ядра Linux Network Emulator (NETEM) и Video LAN. Типы вносимых искажений: потеря пакетов с различной вероятностью, джиттер, задержка, передача видео с различной пропускной способностью. Значения параметров искажений: величина задержки - 100 мс, 300 мс, 500 мс, 800 мс и 1000 мс, вероятность потери пакетов - 0.1%, 0.4%, 1%, 3%, 5%, 8% и 10%, значение пропускной способности – 0.5 Мбит/с, 1 Мбит/с, 2 Мбит/с, 3 Мбит/с, 5 Мбит/с и величина джиттера - 1 мс, 2 мс, 3 мс, 4 мс и 5 мс.

На рис. П.1. представлены примеры кадров тестовых видеопоследовательностей «Parkjoy», «Soccer», «Running», «Restaurant», искаженных потерей пакетов.



а)



б)



в)



г)

Рис П.1. Кадры видео а) «Parkjoy», б) «Soccer», в) «Running», г) «Restaurant»
искаженные потерей пакетов

**ПРИЛОЖЕНИЕ 2. АКТЫ ВНЕДРЕНИЯ
РЕЗУЛЬТАТОВ РАБОТЫ**

«УТВЕРЖДАЮ»

Первый проректор Ярославского
государственного университета

им. П.Г. Демидова



Ю.А. Кащенко

2016 г.

АКТ

внедрения результатов диссертационной работы

Ненахова Ильи Сергеевича на тему

**«Неэталонная оценка качества телевизионных изображений на основе локальных бинарных шаблонов и алгоритмов машинного обучения»
в учебный процесс**

Мы, нижеподписавшиеся, заведующий кафедрой динамики электронных систем, профессор, д.т.н. Брюханов Ю.А. и доцент кафедры динамики электронных систем, к.т.н. Волохов В.А. составили настоящий акт о том, что результаты диссертационной работы Ненахова И.С. внедрены в учебный процесс на кафедре динамики электронных систем физического факультета ЯрГУ (направление «Радиотехника»):

- в курсе «Цифровая обработка изображений» – неэталонный алгоритм оценки качества изображений на основе локальных бинарных шаблонов и методов машинного обучения;
- в курсе «Системы технического зрения» – результаты практического использования неэталонного алгоритма оценки качества изображений в условиях искажений, характерных для задач биометрической идентификации/верификации личности по изображению лица.

Заведующий кафедрой динамики
электронных систем, д.т.н., профессор

Ю.А. Брюханов

Доцент кафедры динамики
электронных систем, к.т.н.

В.А. Волохов

«УТВЕРЖДАЮ»

Первый проректор Ярославского
государственного университета
им. П.Г. Демидова



С.А. Кащенко

2016 г.

АКТ

внедрения результатов диссертационной работы Ненахова Ильи Сергеевича на тему «Неэталонная оценка качества телевизионных изображений на основе локальных бинарных шаблонов и алгоритмов машинного обучения» в научно-исследовательские работы

Результаты диссертационной работы Ненахова И.С., представленной на соискание ученой степени кандидата технических наук по специальности 05.12.04 Радиотехника, в том числе системы и устройства телевидения, использованы в научно-исследовательских работах:

- при выполнении НИР «Разработка и анализ алгоритмов технического зрения для систем анализа аудитории и управления мобильными роботами» (грант РФФИ № 15-07-08674-А) внедрены результаты анализа работы неэталонного алгоритма оценки качества изображений в условиях искажений, характерных для задач биометрической идентификации/верификации личности по изображению лица;
- при выполнении НИР «Разработка нелинейных алгоритмов цифровой обработки речевых сигналов, изображений и видеопоследовательностей на основе модифицированных критериев оценки качества» (грант РФФИ № 15-08-99639-А) внедрен алгоритм неэталонной оценки качества изображений на основе локальных бинарных шаблонов и рандомизированных деревьев.

Заведующий кафедрой динамики
электронных систем, профессор, д.т.н.

Ю.А. Брюханов

(a)вижн

Общество с ограниченной ответственностью «А-Вижн»
Юр. adr.: ул. Угличская, д. 31, кв. 43, г. Ярославль,
150054, Россия
Почт. adr.: ул. Б. Октябрьская, д. 45, г. Ярославль,
150000, Россия

Тел.: +7 (4852) 26-50-10
connect@a-vsn.ru, www.a-vsn.ru

ИНН 7604082087 КПП 760401001
ОГРН 1057600741549

УТВЕРЖДАЮ

Директор ООО «А-Вижн»

Апальков И.В.

«20» сентября 2016 г.



АКТ

внедрения результатов диссертационной работы Ненахова Ильи Сергеевича, выполненной в Ярославском государственном университете имени П.Г. Демидова (ЯрГУ), на тему «Неэталонная оценка качества телевизионных изображений на основе локальных бинарных шаблонов и алгоритмов машинного обучения».

Комиссия в составе: председатель комиссии – заместитель директора Саутов Е.Ю., члены комиссии – технический директор Конюхов А.С., коммерческий директор Дрязгин В.Н., рассмотрев диссертационную работу Ненахова И.С., составила настоящий акт о том, что ее результаты нашли применение в работе ООО «А-Вижн». Особый практический интерес представляет следующий результат диссертации:

- неэталонный алгоритм оценки качества видеопоследовательностей на основе локальных бинарных шаблонов и методов машинного обучения (НОК ЛБШ-1).

Данный алгоритм использован при проектировании систем охранного телевидения. Использование предложенного Ненаховым И.С. алгоритма позволило оптимизировать параметры системы охранного телевидения для различных типов используемых видеокодеков и подобрать эффективные параметры используемого оборудования.

Председатель комиссии,
заместитель директора

Е.Ю. Саутов

Члены комиссии

А.С. Конюхов

В.Н. Дрязгин

**ОБЩЕСТВО С ОГРАНИЧЕННОЙ
ОТВЕТСТВЕННОСТЬЮ
«ТЕХНОЛОГИИ ПОИСКА»**

ИНН 7604288730, КПП 760401001

ОГРН 1157627028129,

150002, г. Ярославль,

пос. Текстилей, д. 19, пом. 57

УТВЕРЖДАЮ

Генеральный Директор

ООО «Технологии поиска»

Карелин П.В.

«15 сентября 2016 г.



АКТ

внедрения результатов докторской работы Ненахова Ильи Сергеевича, выполненной в Ярославском государственном университете имени П.Г. Демидова (ЯрГУ), на тему «Неэталонная оценка качества телевизионных изображений на основе локальных бинарных шаблонов и алгоритмов машинного обучения».

Результаты докторской работы Ненахова И.С. нашли применение в работе ООО «Технологии поиска». Особый практический интерес представляют следующие результаты докторской:

1. Анализ применения алгоритма НОК ЛБШ-1 в задаче распознавания показал, что искажения типа размытие по Гауссу и JPEG оказывают незначительное влияние на уровень распознавания. Искажения типа белый шум, импульсный шум, JPEG2000 существенно влияют на уровень распознавания. Значение метрики НОК ЛБШ-1 ниже 50 соответствуют при всех 5 типах искажений рабочему диапазону для работы алгоритмов распознавания на сверточной нейронной сети.
2. В задаче выбора наилучшего лица для биометрической идентификации показано, что использование алгоритма НОК ЛБШ-1 целесообразно при уровне освещения лица выше 100 лк. При низком уровне освещения целесообразно использование специализированных метрик оценки качества.

Разработанные Ненаховым И.С. алгоритмы неэталонной оценки качества изображений на основе алгоритмов машинного обучения применяются в составе программно-аппаратного комплекса для биометрической идентификации по изображению лица.

Ведущий программист
ООО «Технологии поиска»

Федорин Д.В.

РОССИЙСКАЯ ФЕДЕРАЦИЯ



СВИДЕТЕЛЬСТВО

о государственной регистрации программы для ЭВМ

№ 2016618772

**SkifVideo.QMonitoring - приложение для неэталонной оценки
качества видео**

Правообладатель: *Общество с ограниченной ответственностью
«СКИФВИДЕО» (RU)*

Авторы: *Мочалов Иван Сергеевич (RU), Ненахов Илья Сергеевич
(RU), Приоров Андрей Леонидович (RU)*

Заявка № 2016616387

Дата поступления **17 июня 2016 г.**

Дата государственной регистрации

в Реестре программ для ЭВМ **05 августа 2016 г.**

Руководитель Федеральной службы
по интеллектуальной собственности

Г.П. Ильин

