

Министерство образования и науки Российской Федерации
Ярославский государственный университет им. П. Г. Демидова
Кафедра дискретного анализа
Кафедра инфокоммуникаций и радиофизики

Д. В. Матвеев, А. Г. Седов, В. В. Хрящев, А. Л. Приоров

ОЦЕНКА КАЧЕСТВА ЦИФРОВЫХ ИЗОБРАЖЕНИЙ И ВИДЕОДАННЫХ

Учебно-методическое пособие

Ярославль
ЯрГУ
2018

УДК 621.391(075)
ББК 3973.235я73
М33

*Рекомендовано
Редакционно-издательским советом университета
в качестве учебного издания. План 2018 года*

Рецензент
кафедра инфокоммуникаций и радиофизики ЯрГУ

Матвеев, Дмитрий Вячеславович.
М33 Оценка качества цифровых изображений и видеоданных :
учебно-методическое пособие / Д. В. Матвеев, А. Г. Седов,
В. В. Хрящев, А. Л. Приоров ; Яросл. гос. ун-т им. П. Г. Демидо-
ва. — Ярославль : ЯрГУ, 2018. — 76 с.

Освещены основные вопросы оценки качества статических и динамических изображений. Рассмотрены эталонные и неэталонные критерии оценки качества цифровых изображений. Подробно описаны общедоступные базы тестовых изображений с экспертными оценками, а также метрики, используемые для тестирования точности алгоритмов неэталонной оценки качества изображений. Значительное внимание уделено оценке качества видеоизображений. Включено большое количество контрольных вопросов, охватывающих материал всех глав.

Пособие предназначено для студентов, обучающихся по направлению Прикладная математика и информатика. Кроме того, оно может быть использовано студентами и аспирантами смежных специальностей, изучающих современные технологии обработки, хранения, передачи и приема визуальной информации.

Материал пособия может быть также использован при подготовке студентами курсовых и выпускных квалификационных работ, а также для самообразования.

УДК 621.391(075)
ББК 3973.235я73

© ЯрГУ, 2018

Введение

*Мы, созерцая мир, даем оценку.
А как машине с этим быть?*

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 четкости, 3Dтелевидение) [1–10].

На современном этапе развития цифровая обработка изображений включает в себя следующие основные области [1]:

- коррекцию изображения, их «препарирование», т. е. сознательное разделение на части цифровыми средствами, видеоизменение этих частей и их обратную «сборку»;
- оценку параметров изображений с целью контроля качества их передачи и приема;
- преобразование и кодирование изображений для хранения и передачи по каналам связи;
- компьютерную графику, а также визуализацию информации, т. е. представление массивов данных в виде различных изображений, что очень эффективно, так как облегчает решение многих задач, сложных именно своей абстрактностью.

Алгоритмы обработки, сжатия и кодирования визуальной информации неизбежно вносят искажения различного типа, что делает актуальным вопрос об оценке параметров этих искажений и объективного оценивания качества искаженного изображения. Исследовательские лаборатории в области контроля качества изображений и видеоданных созданы во многих крупных университетах мира. Среди них стоит выделить University of Texas at Austin, Tampere University of Technology и другие ведущие мировые университеты. Практическими разработками в этой области занимаются как крупные глобальные компании: LG, Huawei,

Samsung, Tektronix, так и маленькие технологичные компании-стартапы, созданные при ведущих университетах.

Значительный вклад в данную область науки и техники внесли как отечественные ученые М. И. Кривошеев, А. С. Крылов, Д. С. Ватолин, А. С. Конушин, Б. А. Алпатов, М. К. Чобану, Ю. С. Бехтин, Е. П. Петров, И. С. Трубин, Е. В. Медведева, М. Н. Фаворская, Ю. С. Радченко, Н. А. Обухова, А. А. Мотыко, так и зарубежные – А. Bovik, Z. Wang, H. Sheikh, E. Simoncelli, J. Astola, K. Egiazarian, N. Ponomarenko, V. Lukin, M. Carli, F. Battisti, O. Ieremeiev, L. Jin, B. Vozel, K. Chehdi и др. [1–19].

На современном этапе развития все методы оценки качества классифицируются на две большие категории (рис. В.1):



Рис. В.1. Классификация современных методов оценки качества изображений

1. Субъективные критерии оценки качества изображений. Такая оценка в данном случае осуществляется человеком;

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.

Так как человек является приёмником 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 есть среднее арифметическое всех оценок членов группы (MOS).

Для того чтобы результаты оценки качества могли быть воспроизводимыми, Международный союз электросвязи принял нормативный документ ITU-R BT.500-11 [20], 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, 5 – отличное), так и непрерывные в диапазоне 0–100.

Популярной на данный момент является метрика DMOS (Difference Mean Opinion Score), которая может быть записана в виде:

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

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

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

где $d_{i,j}$ – оценка i -го эксперта для j изображения, $r_{i,ref}(j)$ – обозначает оценку качества, которую выставил i -й эксперт эталонному изображению, \bar{d}_i – среднее по всем изображениям значение оценок, которое выставляет i -й эксперт, а σ_i – среднее квадратическое отклонение. Оценка производится по шкале качества, согласно которой наблюдатель должен оценить общее качество изображения, изменяющееся от «очень плохого» до «очень хорошего». Результатом экспертных оценок является линейное отображение значения, определенного по шкале качества в диапазоне оценок от 1 до 100 – чем большее значение, тем хуже качество тестируемого изображения.

Несмотря на свою точность, субъективные экспертизы имеют серьезные недостатки:

- дороговизну и низкую скорость работы;
- зависимость результатов от точности соблюдения экспериментальных условий и эмоционального состояния экспертов;
- отсутствие количественных оценок.

Последний пункт является главным недостатком субъективного критерия. Он не позволяет решать задачи оптимизации изображений в пространстве непрерывно изменяющихся параметров. Здесь возможен только лишь небольшой перебор параметров.

Под объективными критериями оценки качества понимаются критерии, расчет которых происходит автоматически, без участия группы экспертов. 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 Кендалла и величину среднеквадратичной ошибки. Чем выше значения коэффициентов корреляции и чем меньше величина среднеквадратичной ошибки, тем точнее работает рассматриваемый объективный алгоритм.

Как уже отмечалось, алгоритмы объективной оценки качества можно разделить на две большие группы. Первая – 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 – работают только с искаженным изображением. Такие алгоритмы более востребованы на практике, так как зачастую в «реальных» приложениях эталонное изображение неизвестно. Иногда в отдельную группу выносят так называемые частично-эталонные 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).

Учебно-методическое пособие состоит из 4 глав и охватывает широкий круг вопросов, связанных с алгоритмами эталонной и неэталонной оценки качества изображений. В его основу положены материалы лекций, которые в течение ряда лет читались бакалаврам, магистрантам и аспирантам ЯрГУ, а также результаты работы авторов и их коллег в данном научном направлении.

Выражаем глубокую благодарность выпускникам ЯрГУ, кандидатам технических наук И. Апалькову, Е. Саутову, В. Кобелеву, А. Ганину, С. Бухтоярову, Д. Зараменскому, В. Волохову, И. Мочалову, Е. Сергееву, Д. Куйкину, В. Соловьеву, М. Голубеву, Л. Шмаглицу, Е. Павлову, А. Никитину, И. Ненахову, работы которых оказали значительное влияние на формирование взглядов авторов в рассматриваемом научном направлении.

Глава 1. Эталонные алгоритмы оценки качества изображений

1.1. Пиковое отношение сигнал/шум

Пиковое отношение сигнал/шум (ПОСШ) – наиболее популярная объективная метрика качества изображений. Для двух цифровых изображений с 256 уровнями квантования яркости (что соответствует классическому 8-битовому представлению) и общим числом пикселей – N , эту метрику можно рассчитать как:

$$ПОСШ = 20 \log_{10} \frac{255}{\sqrt{СКО}},$$

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

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

Основными практическими преимуществами использования данной эталонной оценки являются:

- 1) отсутствие параметров и, как следствие, – простота расчета;
- 2) физическая интерпретация в виде энергии сигнала ошибки;

3) возможность использования приведенной оценки в оптимизационных задачах благодаря свойствам симметричности и дифференцируемости.

Однако при практическом использовании оценки ПОСШ возникает следующий ряд проблем:

1. Восприятие качества изображений не зависит от пространственных зависимостей между исходным и искаженным двумерным сигналом. Как результат, изменение порядка частей изображения не должно повлиять на результат оценки;

2. Визуальное восприятие не зависит от любых связей между сигналом изображения и сигналом ошибки. Это значит, что для сигнала с одинаковой ошибкой не имеет значения, каково исходное изображение. Результат будет одинаков;

3. Визуальное восприятие качества определяется только величиной сигнала ошибки. Как следствие, изменения в сигнале ошибки не влияют на результат оценки;

4. Все составляющие сигнала одинаково важны для визуальной оценки качества.

Иллюстрация этих положений приведена на рис. 1.1–1.3.

На рис. 1.1 изображение (б) создано добавлением белого гауссовского шума к исходному изображению (а). В изображении (в) амплитудные (яркостные) составляющие сигнала оставлены такими же, как и в изображении (а), но пространственный порядок изменен. Изображение (г) получено из изображения (б) путем использования процедуры перераспределения, как и в случае изображения (в). Другими словами, координатные преобразования, используемые при переходе из (б) в (г), такие же, как и при переходе из (а) в (в). Значения ПОСШ для изображений (б) и (г) относительно (а) и (в) соответственно абсолютно одинаковы. Однако изображение (г) воспринимается более зашумленным, нежели (б).

На рис. 1.2 одинаковый сигнал ошибки добавлен к исходному изображению (а) и (в). Сигнал ошибки (д) выбран таким образом, что полностью коррелирует с изображением (а). Оба искаженных сигнала (б) и (г) имеют одинаковое значение ПОСШ. Однако визуально искажение изображения (г) кажется сильнее, нежели (б).

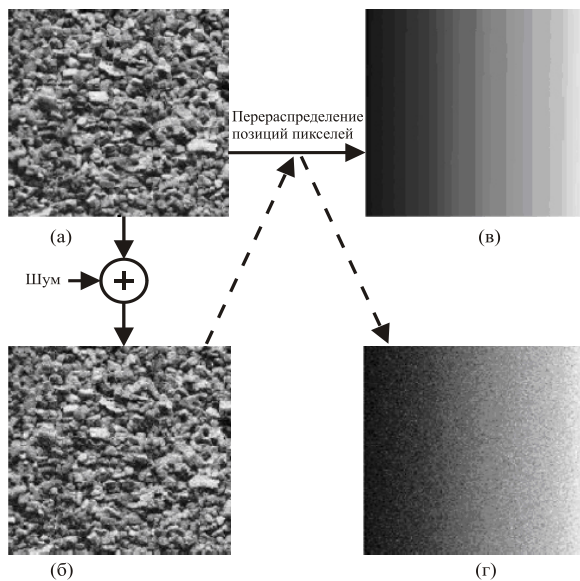


Рис. 1.1. Визуальное различие при добавлении гауссовского шума

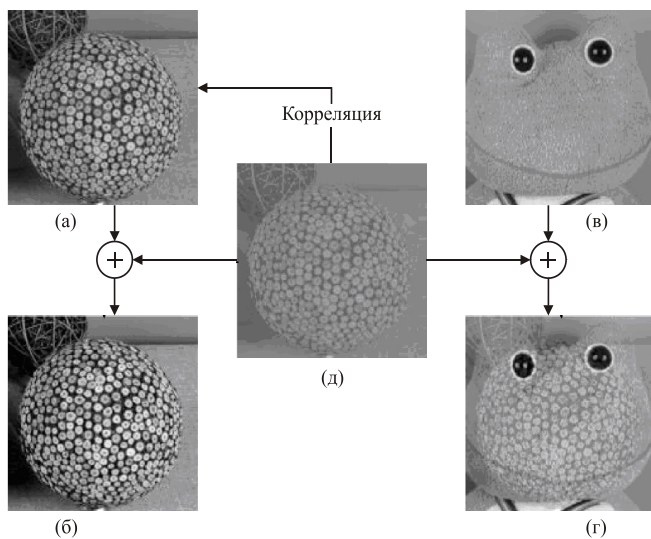


Рис. 1.2. Визуальное различие при добавлении коррелированного шума

На рис. 1.3 два искаженных изображения созданы из одного исходного. Первое изображение подвержено добавлению постоянной компоненты ко всем значениям пикселей. Второе изображение получено путем добавления ко всем значениям пикселей исходного изображения константы со случайным знаком. Визуальное качество изображений значительно различается, в то время как ПОСШ полностью игнорирует знаки и значения таких оценок качества для обоих изображений равны.

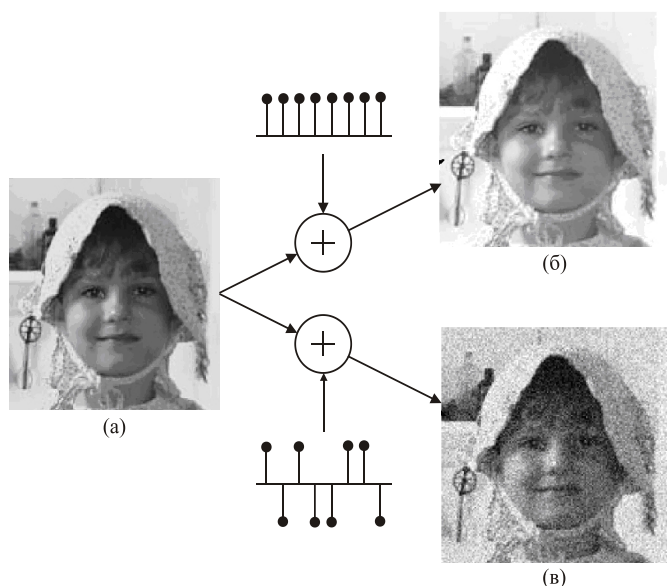
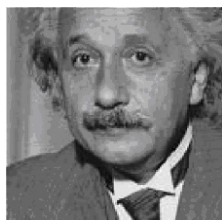


Рис. 1.3. Визуальное различие при добавлении к исходному изображению константы и знакопеременной последовательности

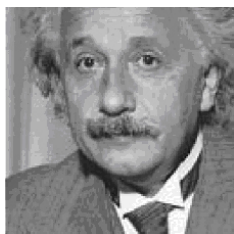
Другим примером может служить рис. 1.4., где исходное тестовое изображение «Эйнштейн» (а) изменено несколькими типами искажений: средний сдвиг яркости (б), увеличение контраста (в), искажение импульсным шумом (г), искажение гауссовским шумом (д), размытие (е), JPEGсжатие (ж). Необходимо отметить, что изображения (б) – (ж) имеют одинаковое значение СКО, но, очевидно, различное визуальное восприятие качества.

Также СКО чувствительна к пространственным изменениям, изменению масштаба и вращению. Это легко заметить на рис. 1.5, где то же самое эталонное изображение «Эйнштейн» подвержено изменению путем пространственного сдвига влево (а), зум-эффекта (б) и вращения (в).

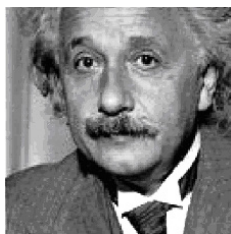
Объясняются столь некоррелированные результаты тем, что пиксельная структура, используемая в ПОСШ, неточно отображает систему человеческого визуального восприятия. Психофизические и психофизиологические исследования показали, что функции зрительной системы человека не учитываются в ПОСШ.



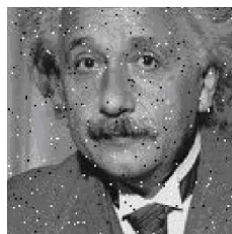
(а)



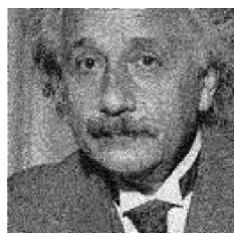
(б) СКО=306



(в) СКО=309



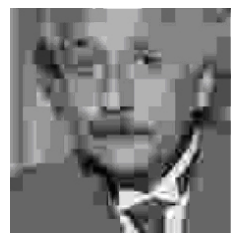
(г) СКО=313



(д) СКО=309

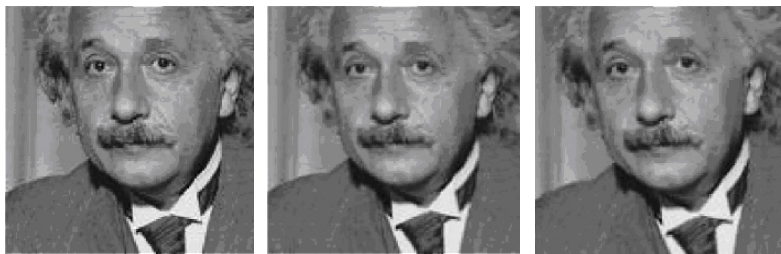


(е) СКО=308



(ж) СКО=309

Рис. 1.4. Изображение «Эйнштейн» с близкими значениями СКО



(а) СКО=871

(б) СКО=694

(в) СКО=590

Рис. 1.5. Реакция СКО на пространственные изменения

1.2. Универсальный индекс качества

Начиная с 2000 года проведено множество исследований по разработке эталонных алгоритмов визуального качества обрабатываемых изображений, которые бы стояли ближе к результатам субъективных оценок. Одним из таких алгоритмов является универсальный индекс качества изображений (УИК, UQI – Universal Quality Index), используемый для определения подобности восстановленного изображения оригиналу в зависимости от вида и степени искажения эталонного изображения. Универсальность данного критерия заключается в том, что он не только отражает некоторую схожесть обработанного изображения по отношению к оригиналу, но и должным образом учитывает различные виды искажений. В отличие от многих других алгоритмов данный метод не привязан к специфике изображения и искажениям, присутствующим в нем. Он основывается на статистическом анализе отдельных блоков входного сигнала и дальнейшем сравнении полученных результатов со значениями эталонного изображения. Рассмотрим расчет метрики УИК более подробно.

1.2.1. Определение УИК для одномерных сигналов

Пусть даны две последовательности $\{x_i\}$ и $\{y_i\}$, где $i = 1, 2, \dots, N$, соответствующие стационарному входному и выходному сигналам. Тогда универсальный индекс качества будет определяться комбинацией статистических характеристик (математического ожидания, дисперсии и корреляционной функции) указанных последовательностей по следующей формуле:

$$\text{УИК} = \frac{4 \sigma_{xy} \bar{x} \bar{y}}{[\sigma_x^2 + \sigma_y^2][(\bar{x})^2 + (\bar{y})^2]}, \quad (1.1)$$

где

$$\begin{aligned} \bar{x} &= \frac{1}{N} \sum_{i=1}^N x_i, \\ \bar{y} &= \frac{1}{N} \sum_{i=1}^N y_i, \\ \sigma_x^2 &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})^2, \\ \sigma_y^2 &= \frac{1}{N} \sum_{i=1}^N (y_i - \bar{y})^2, \\ \sigma_{xy} &= \frac{1}{N} \sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y}). \end{aligned}$$

Из выражения (1.1) следует, что значение метрики УИК изменяется в интервале $[-1, 1]$, при этом значение $\text{УИК}=1$ соответствует максимально возможному качеству сигнала. Оно получается только тогда, когда $y_i = x_i$ на протяжении всего сигнала ($i = 1, 2, \dots, N$). В свою очередь минимальное значение $\text{УИК}=-1$ достигается в случае, если $y_i = 2\bar{x} - x_i$, $i = 1, 2, \dots, N$.

Определенный таким образом индекс качества учитывает в себе три искажающих фактора: степень коррелированности отчетов двух сигналов, изменение значений математического

ожидания и среднеквадратичных отклонений сигнала относительно оригинала. Такое представление позволяет наиболее точно оценивать качество восстановленного сигнала и величину искажений, присутствующих в нем. В результате этого исходное выражение (1.1) может быть представлено в виде произведения трех множителей:

$$\text{УИК} = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x} \bar{y}}{(\bar{x})^2 + (\bar{y})^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2}. \quad (1.2)$$

Первый множитель есть коэффициент взаимной корреляции сигналов x и y , который определяет степень их линейной подобности и принимает значения в интервале $[-1, 1]$. При этом максимальное значение достигается в случае, если $y_i = ax_i + b$, где $a > 0$, $b > 0$ и $i = 1, 2, \dots, N$. Если восстановленный сигнал линейно соотносится с входным, но в нем присутствуют какие-либо искажения, они будут учитываться во втором и третьем множителях. Второй множитель (1.2) определяет соотношение между математическими ожиданиями входного и восстановленного сигнала и принимает значения в интервале $[0, 1]$, при этом максимум достигается в случае их равенства. В случае если сигналом является изображение, этот коэффициент отвечает за изменения его яркости.

Значения среднеквадратичных отклонений σ_x и σ_y определяют динамический диапазон изменения сигналов x и y соответственно. Вследствие этого они могут рассматриваться как характеристики контраста в том случае, если сигналом является изображение. Таким образом, третий множитель в выражении (1.2) для УИК соответствует искажению контрастности восстановленного сигнала относительно оригинала и принимает значения в интервале $[0, 1]$.

Предложенный индекс качества обладает хорошей согласованностью со значениями субъективных экспертных оценок. К тому же, в отличие от оценки ПОСШ, он имеет абсолютное значение, равное единице, то есть при показаниях УИК, близких к единице, можно говорить о высоком качестве восстановленного сигнала. Рассмотрим теперь, как выражение (1.1) можно использовать в роли критерия качества для цифровых изображений.

1.2.2. Применение УИК для оценки качества изображений

Изображение является двумерным и в общем случае нестационарным сигналом, поэтому прямое использование формулы (1.1) для вычисления УИК в данном случае недопустимо. Для решения этой проблемы предлагается выделять в изображении отдельные локальные области размером $A \times B$, в пределах которых изображение можно считать стационарным и вычислять статистические характеристики (математическое ожидание, дисперсию и корреляционную функцию) внутри этой области, а уже на основании их вычислять значение $УИК_i$, соответствующее выделенному блоку. Такой подход обеспечивает наиболее правильное и точное суждение о качестве восстановленного изображения.

В результате этого процесс расчета метрики УИК для конкретного изображения будет сводиться к следующему: «скользящее окно» (маска) размером $A \times B$, начиная из верхнего левого угла, проходит пиксель за пикселем по всем строкам и столбцам этого изображения, на каждом шаге вычисляется значение статистических характеристик и собственно сама оценка $УИК_i$ для текущего положения окна. Итоговый критерий, характеризующий качество изображения в целом, будет определяться как среднее арифметическое значений УИК, на протяжении всего изображения:

$$УИК = \frac{1}{M} \sum_{i=1}^M УИК_i ,$$

где M – число блоков, внутри которых вычислялись $УИК_i$.

Описанный алгоритм особенно удобен в расчете для изображений, имеющих блочную структуру. Как известно, одним из примеров таких изображений является формат JPEG, основанный на дискретном косинусном преобразовании (ДКП) отдельных блоков размером 8×8 пикселей. В результате этого для оценки качества JPEG-изображений возможно использование оценки УИК с размером «скользящего окна», равным размеру блоков ДКП.

Благодаря комбинированной структуре УИК (1.2), оценка качества восстановленных изображений осуществляется наиболее

точно. В ряде практических случаев использование УИК вместо оценки ПОСШ дает существенный выигрыш. Это наблюдается, например, при изменении яркости изображений, трансформации гистограмм, гамма-коррекции и т. д. Особенно важно то, что в таких приложениях УИК позволяет дифференцировать изображения, имеющие одинаковые или близкие значения ПОСШ, что способствует построению действительно оптимальных фильтров для обработки изображений.

Исходя из выражения (1.1) для вычисления УИК следует, что в некоторых случаях могут возникать ситуации, когда значение нового индекса качества имеет неопределенный характер; математически это означает неопределенность типа «нуль на нуль». Причиной этого могут служить два фактора. Знаменатель и числитель могут обращаться в нуль, во-первых, при совместном равенстве нулю дисперсий для каких-либо блоков разбиения изображений, во-вторых, при совместном равенстве нулю математических ожиданий внутри этих блоков. На практике подобные ситуации встречаются довольно часто, поэтому предлагается исключать их из рассмотрения и не включать такие значения в конечный результат (рис. 1.6).

1.3. Коэффициент структурного подобия

Сигналы естественных изображений хорошо структурированы. Выборки, взятые из сигналов изображений, имеют сильную зависимость, и эти зависимости несут важную информацию о структуре: объектах, их расположении, текстурах. Главная идея, лежащая в основе расчета коэффициента структурного подобия, заключается в том, что зрительная система человека хорошо приспособлена к выделению структурной информации и поэтому измерение степени структурного подобия должно обеспечивать хорошее приближение к визуально воспринимаемому качеству изображения.

Для того чтобы преобразовать общий принцип структурного подобия в конкретный алгоритм оценки качества изображения, необходимо разделить структурные и неструктурные искажения.

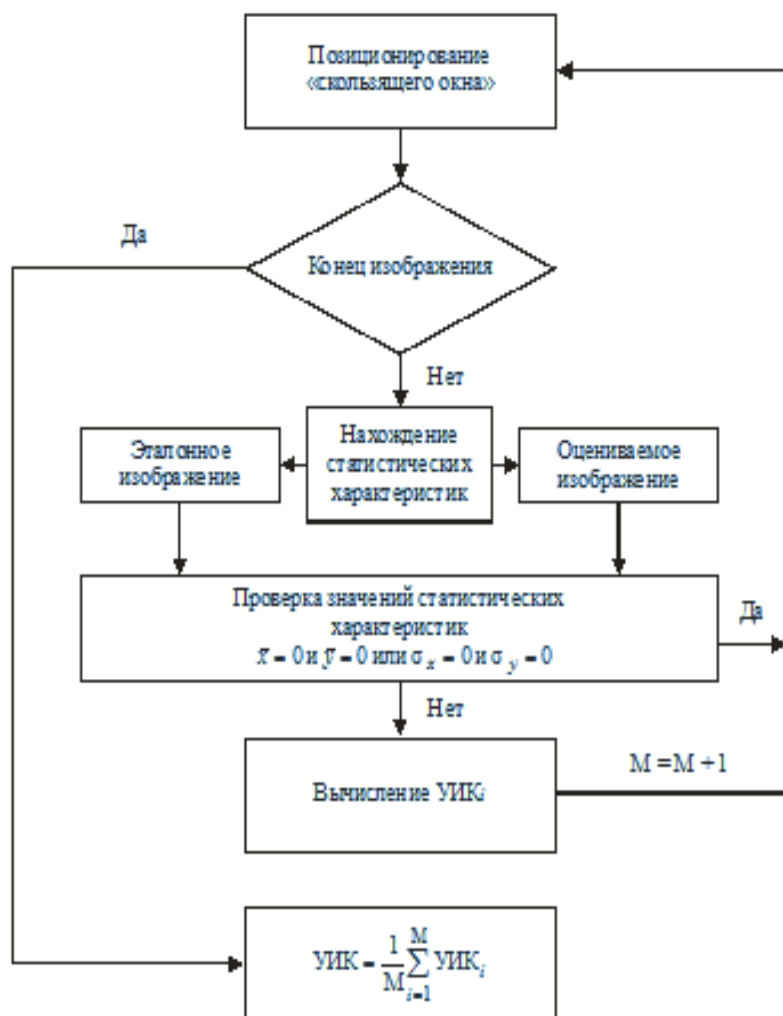


Рис. 1.6. Алгоритм нахождения универсального индекса качества для изображений

Неструктурными искажениями считаются те из них, которые не вносят изменения в структуру объектов на изображении, остальные называются структурными.

Изображение с измененным контрастом (рис. 1.7б) сохраняет фактически всю существенную информацию об объектах исходного (эталонного) изображения (рис. 1.7а). Таким образом, согласно принципу структурного подобия, такое искажение – неструктурное.

Изображение, искаженное артефактами сжатия с потерями по стандарту JPEG (рис. 1.7д), имеет то же самое СКО, что и изображение с измененным контрастом (рис. 1.7б), однако можно видеть, что артефакты сжатия значительно изменяют структуру объектов на изображении. Искажения такого рода являются структурными.

На рис. 1.7 исходное изображение представлено вектором в пространстве значений пикселей этого изображения. Любые искажения исходного изображения можно понимать как добавление вектора искажения к «эталонному» вектору. В частности, векторы искажения одинаковой длины определяют гиперсферу равного СКО в пространстве изображений. Длина вектора искажения не является эффективной оценкой его степени, тогда как направление имеет значение для визуально воспринимаемого качества.

На рис. 1.8 показано, что искажения яркости и контраста отделяются от структурных искажений в рассмотренном пространстве. Векторы яркостных искажений параллельны вектору изображения равной яркости $x_1 = x_2 = \dots = x_N$, который перпендикулярен гиперплоскости $\sum_{i=1}^N x_i = 0$. Изменения контраста определяются направлением $\mathbf{x} - \bar{x}$.

Таким образом, векторы искажений яркости и контраста определяют в исходном пространстве плоскость неструктурных искажений. Структурные же искажения соответствуют повороту этой плоскости на некоторый угол (рис. 1.8).

Коэффициент структурного подобия (КСП, SSIM – Structural Similarity Index Measure) – это эталонный алгоритм оценки качества изображений, воплощающий идею структурного подобия в пространственной области. На вход алгоритма подаются изображения I и J , одно из которых эталонное. На выходе рас-

считывается индекс структурного подобия, который служит основной оценкой качества. В ходе расчета метрики КСП происходят сравнения яркости, контраста и структуры изображений (рис. 1.9).

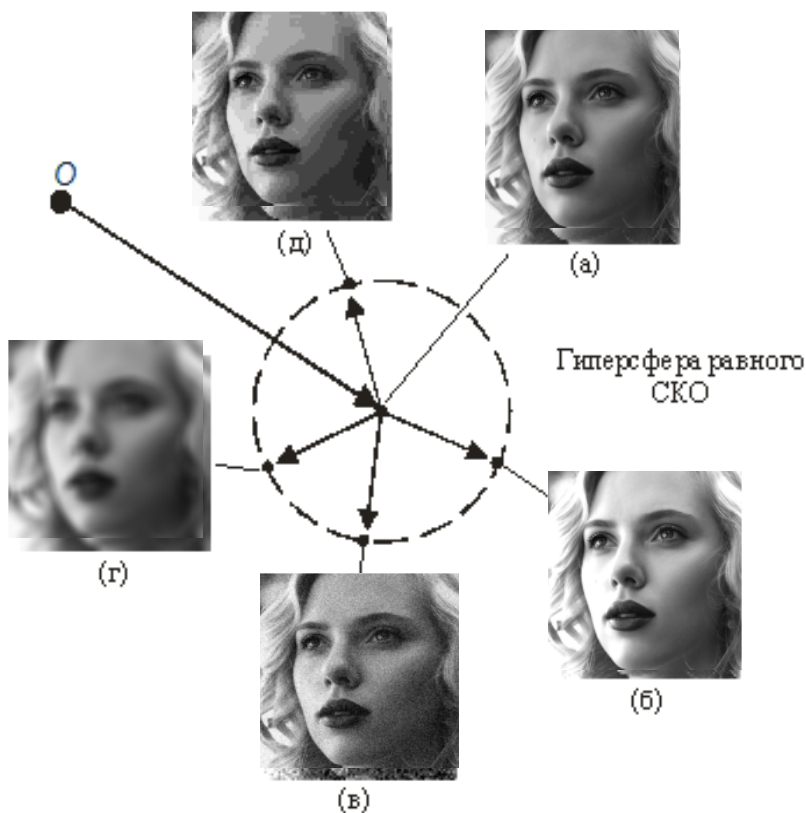


Рис. 1.7. Структурные и неструктурные искажения:
а) оригинальное изображение; б) увеличение контраста;
в) внесение шума; г) размытие;
д) артефакты сжатия по стандарту JPEG

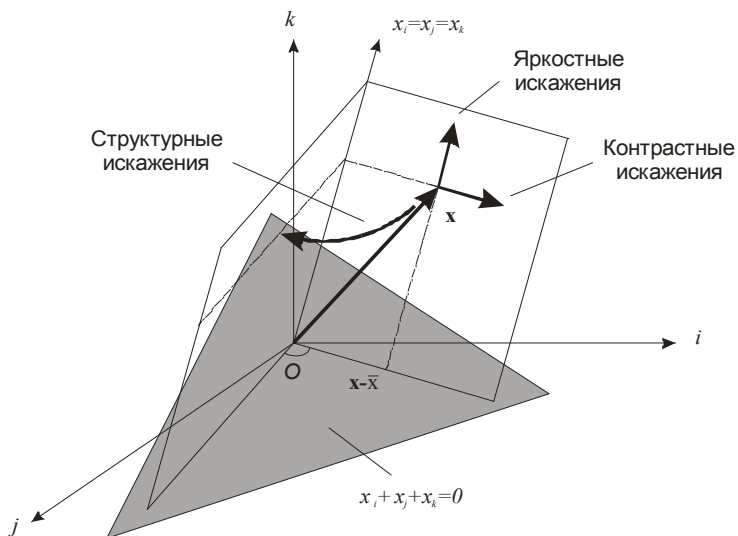


Рис. 1.8. Разделение структурных и неструктурных искажений в пространстве изображения

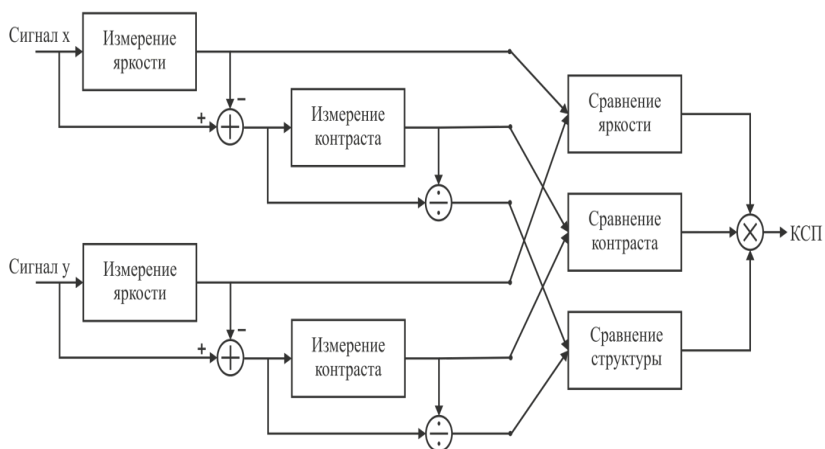


Рис. 1.9. Алгоритм расчета коэффициента структурного подоби

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 КСП внутри одного окна, состоящего из пикселей. Локальная яркость каждого изображения в пикселе с координатами (x,y) находится как усредненное значение яркости в локальной окрестности этого пикселя:

$$\mu_x = \bar{x} = \frac{1}{N} \sum_{i=1}^N x_i, \quad (1.3a)$$

$$\mu_y = \bar{y} = \frac{1}{N} \sum_{i=1}^N y_i, \quad (1.3b)$$

где N – количество пикселей в рассматриваемой окрестности.

Функция $l(x,y)$, характеризующая искажения яркости, будет определяться через значения μ_x и μ_y :

$$l(x,y) = l(\mu_x, \mu_y),$$

Если вычесть среднее значение μ_x из окрестности сигнала \mathbf{x} , то результат $\mathbf{x} - \mu_x$ будет являться проекцией вектора \mathbf{x} на гиперплоскость $\sum_{i=1}^N x_i = 0$, изображенную на рис. 1.8. В качестве оценки

локальной контрастности сигнала используется среднеквадратичное отклонение:

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

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

Функция $c(x,y)$, характеризующая искажения контраста, будет определяться через значения σ_x и σ_y :

$$c(x,y) = c(\sigma_x, \sigma_y)$$

Структурное подобие между эталонным и оцениваемым изображениями определяется через нормированные на свои средние значения соответствующие среднеквадратичные отклонения:

$$s(x,y) = s\left(\frac{x - \mu_x}{\sigma_x}, \frac{y - \mu_y}{\sigma_y} \right).$$

Все три составляющие объединяются в общую локальную оценку следующим образом:

$$S(x,y) = f[l(x,y), c(x,y), s(x,y)].$$

На функцию $S(x,y)$ накладываются следующие ограничения:

1. **Симметрия:** $S(x,y) = S(y,x)$. Когда определяется подобие между двумя сигналами, изменение порядка входящих сигналов не должно влиять на результирующее измерение;

2. **Ограниченность:** $S(x,y) \leq 1$. Верхняя граница служит показателем того, насколько близко два сигнала к тому, чтобы быть идентичными;

3. **Единственный максимум:** $S(x,y) = 1$. Функция $S(x,y)$ принимает максимальное значение тогда и только тогда, когда $\mathbf{x} = \mathbf{y}$.

Для удовлетворения этих условий входящие в состав $S(x,y)$ функции $l(x,y)$, $c(x,y)$, $s(x,y)$ определяются следующим образом:

$$l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1}, \quad (1.4)$$

где константа C_1 включена для того, чтобы избежать обращения в бесконечность, когда значение μ_x^2 и μ_y^2 близко к нулю. В частности, значение C_1 выбирают равным $(K_1 \cdot L)^2$, где L – динамический диапазон значений пикселей изображения, а $K_1 \ll 1$ – малая константа. Заметим, что уравнение (1.4) подчиняется трем свойствам, описанным выше.

Функция искажения контраста имеет подобный вид:

$$c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (1.5)$$

где $C_2 = (K_2 \cdot L)^2$ и $K_2 \ll 1$. Эти определения также удовлетворяют трем свойствам, описанным выше.

Функция структурных искажений вводится следующим образом:

$$s(x,y) = \frac{2\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3}, \quad (1.6)$$

где

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

В выражение (1.6), так же, как и для двух предыдущих функций, вводится малая константа C_3 . Геометрически $s(x,y)$ – это косинус угла между векторами $(\mathbf{x} - \mu_x)$ и $(\mathbf{x} - \mu_y)$.

Три рассмотренные составляющие (1.4), (1.5) и (1.6) объединяются в локальный коэффициент структурного подобия изображений \mathbf{x} и \mathbf{y} :

$$S(x,y) = l(x,y)^\alpha, c(x,y)^\beta, s(x,y)^\gamma,$$

где $\alpha > 0$, $\beta > 0$, $\gamma > 0$ – параметры, которые регулируют относительную значимость отдельных составляющих.

На практике, как правило, используются следующие значения параметров: $\alpha = \beta = \gamma = 1$, $C_3 = C_2/2$, $K_1 = 0,01$, $K_2 = 0,03$.

При вычислении КСП скользящим окном может быть получена карта структурного подобия для заданной пары изображений (эталонное – оцениваемое). Пример такой карты приведен на рис. 1.10. Более светлые участки на карте КСП соответствуют большему структурному подобию и, следовательно, большему качеству изображения.

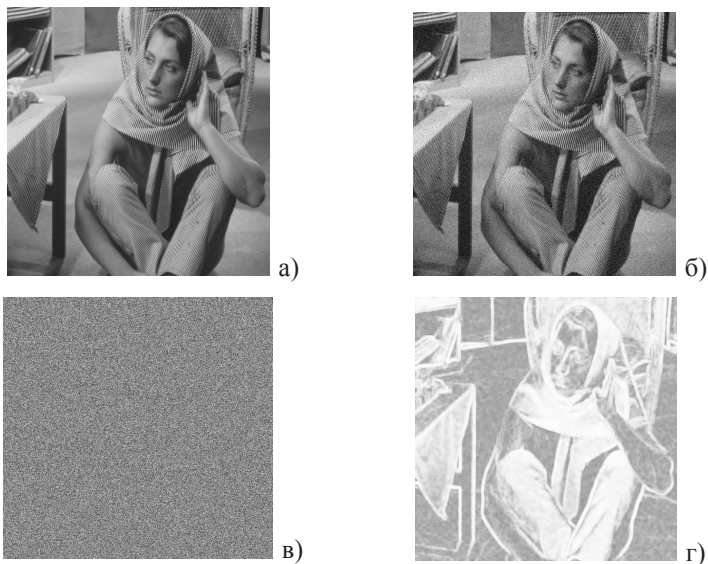


Рис. 1.10. Структурное качество изображения:
а) оригинал; б) зашумленное изображение;
в) карта абсолютных ошибок; г) карта КСП, окно 7×7

На искаженном изображении (рис. 1.10б) шум гораздо лучше виден на однотонных областях, что отражается на карте КСП (рис. 1.10г) в виде более темных областей.

Восприятие изображения зависит от расстояния до наблюдателя и от особенностей его зрительной системы. Что бы учесть этот факт, предложена модификация оценки КСП – многомасштабный коэффициент структурного подобия (МКСП, Multi-Scale SSIM). 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 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, $\alpha > 0$, β , γ – положительные константы, регулирующие вклад параметров алгоритма КСП в результирующий индекс качества на разных масштабах.

1.4. Модификация ПОСШ с учетом характеристик зрительной системы человека

Проведенные исследования показали, что система визуального восприятия человека более восприимчива к искажениям на низких частотах, чем в высокочастотной области. Этот факт, в частности, был учтен при разработке стандарта JPEG. Была спроектирована специальная таблица корректирующих величин, которая применяется для квантования коэффициентов ДКП в блоках 8×8 . Согласно этой таблице высокочастотные коэффициенты кванту-

ются с большим шагом квантования, что приводит к качеству восстановленного изображения, лучшему, чем в случае, если применять для всех коэффициентов равномерное квантование.

Выражение для СКО с плавающим окном 8×8 с ДКП, с учетом визуального восприятия человеком, имеет вид:

$$\text{СКО}_M = \frac{\sum_{i=1}^{I-7} \sum_{j=1}^{J-7} \sum_{m=1}^8 \sum_{n=1}^8 ((X[m,n]_{i,j} - X[m,n]_{i,j}^e) T_c[m,n])^2}{(I-7)(J-7)64},$$

где $X_{i,j}$ – коэффициенты ДКП блока 8×8 , для которого координаты верхнего левого элемента равны i и j , $X_{i,j}^e$ – коэффициенты ДКП соответствующего окна оригинального изображения, T_c – матрица корректирующих факторов. В качестве такой матрицы используется матрица, построенная с использованием таблицы квантования стандарта JPEG.

Матрица T_c обладает следующими двумя свойствами:

1) соотношения ее коэффициентов обратно пропорциональны соответствующим коэффициентам таблицы квантования JPEG;

2) суммарный корректирующий фактор матрицы $\sum_{m=1}^8 \sum_{n=1}^8 \frac{T_c[m,n]^2}{64}$ равен единице. Это означает, что в случае одно-

родного распределения искажений по всем пространственным частотам значения СКО_M совпадают с СКО.

На основе полученного результата можно записать выражение для критерия ПОСШ_M учитывающего особенности системы визуального восприятия человека:

$$\text{ПОСШ}_M = 10 \lg(255^2 / \text{СКО}_M).$$

Использование рассмотренной модификации ПОСШ увеличивает корреляцию с субъективной оценкой экспертов. С другой стороны, возрастают и временные затраты, требуемые на вычисление оценки.

1.5. Сравнительный анализ алгоритмов сжатия изображений относительно критериев ПОСШ и УИК

Проведем пример исследования алгоритмов сжатия изображений JPEG, JPEG2000 и SPIHT относительно объективных эталонных метрик качества – ПОСШ и УИК. Будем использовать три изображения с различной степенью детализации из тестового набора, приведенные на рис. 1.11.

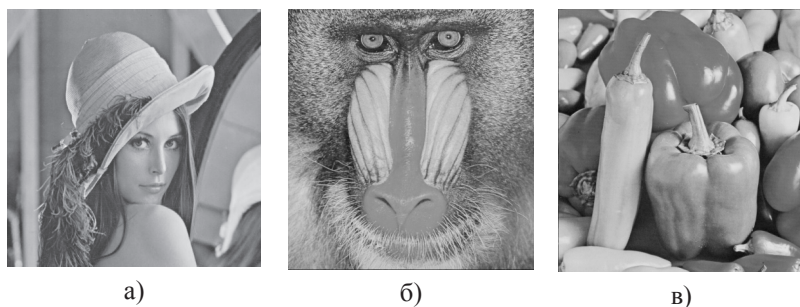


Рис. 1.11. Тестовые изображения:
а) «Лена»; б) «Бабуин»; в) «Перцы»

На рис. 1.12. в качестве примера приведены зависимости оценок качества восстановленного тестового изображения «Лена» от коэффициента сжатия K для различных алгоритмов компрессии. Приведенные зависимости показывают, что:

1) для всех трех типов тестовых изображений наиболее высокие оценки как по шкале ПОСШ, так и по шкале УИК получаются при использовании алгоритма JPEG2000;

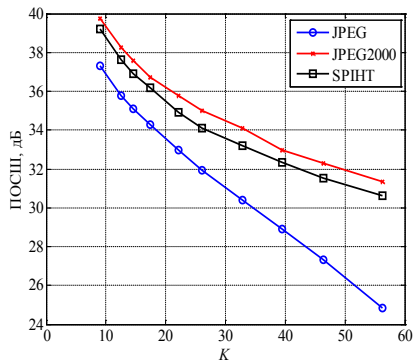
2) при одинаковых коэффициентах сжатия алгоритм SPIHT по качеству сжатого изображения сопоставим с алгоритмом JPEG2000, оценки по ПОСШ и УИК различаются менее чем на 3 %;

3) проигрыш в качестве JPEG-изображений в диапазоне наиболее часто используемых на практике коэффициентов сжатия (2–20) более заметен по шкале ПОСШ, чем по шкале УИК;

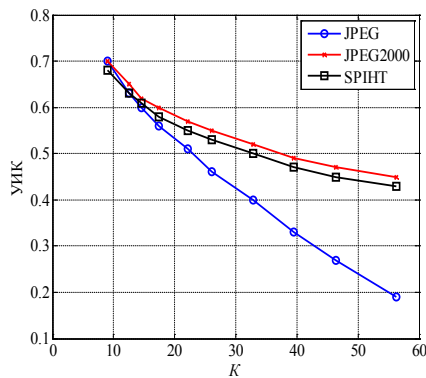
4) с точки зрения оценки УИК при малом сжатии все алгоритмы предоставляют одинаковое качество, однако при высоком

сжатии ($K \geq 20$) стандарт JPEG значительно уступает более современным алгоритмам.

Для дополнительной проверки пункта 4 вышеизложенных выводов на рис. 1.13. представлены визуальные результаты сжатия изображения «Перцы». При малых коэффициентах сжатия разница между декодированными изображениями незаметна. Напротив, при больших коэффициентах сжатия (в данном случае $K = 46,7$) JPEG-изображение обладает ярко выраженной блочной структурой, что свидетельствует о потере большой части информации. Эффект размытия, вносимый в изображения алгоритмами JPEG2000 и SPIHT при данных степенях сжатия, менее заметен.

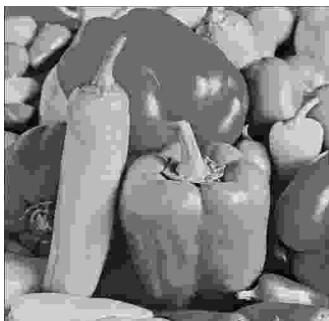


а)



б)

Рис. 1.12. Сравнение эффективности работы алгоритмов сжатия для изображения «Лена»: а) по шкале ПОСШ; б) по шкале УИК



JPEG (K = 46,7), ПОСШ = 27,32 дБ,
УИК = 0,25



JPEG (K = 9), ПОСШ = 36,01 дБ,
УИК = 0,64



JPEG2000 (K = 46,7),
ПОСШ = 32,27 дБ, УИК = 0,42



JPEG2000 (K = 9),
ПОСШ = 38 дБ, УИК = 0,67



SPIHT (K = 46,7),
ПОСШ = 31,46 дБ, УИК = 0,4



SPIHT (K = 9), ПОСШ = 37,23 дБ,
УИК = 0,64

Рис. 1.13. Визуальный анализ алгоритмов сжатия JPEG,
JPEG2000 и SPIHT для тестового изображения «Перцы»

Проведем анализ алгоритмов сжатия изображений для трех типов тестовых изображений. Исследуя графики на рис. 1.14–1.16, можно сделать следующие выводы:

1. Стандарт JPEG сжимает низкочастотные изображения с меньшими потерями качества, чем высокочастотные. Изображение «Бабуин» имеет много резких перепадов, поэтому плохо сжимается алгоритмом JPEG на высоких коэффициентах сжатия. В этом случае при делении на блоки нарушаются естественные границы;

2. При сжатии изображения «Бабуин» коэффициенты, отвечающие за мелкие (особенно диагональные) детали, находятся при работе алгоритма SPIHT в нижнем правом углу плоскости битов. Иными словами, нижние уровни пространственно ориентированного дерева содержат коэффициенты, имеющие достаточно большие значения, что приводит к потере значительной части информации и, следовательно, к уменьшению ПОСШ;

3. На тестовых изображениях «Лена» и «Перцы» алгоритм SPIHT работает так же, как и алгоритм JPEG2000, из-за отсутствия в этих изображениях резких перепадов;

4. Стандарт JPEG2000 хорошо сжимает любые типы изображений, но на высоких коэффициентах сжатия качество лучше для низкочастотных изображений.

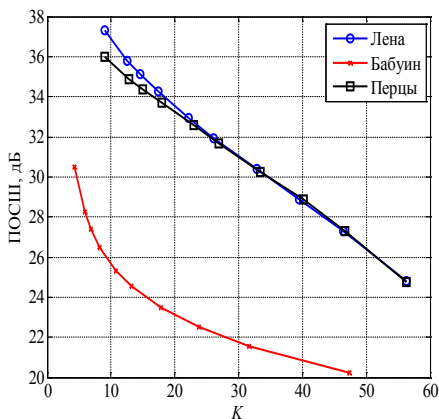


Рис. 1.14. Работа алгоритма JPEG для всех тестовых изображений

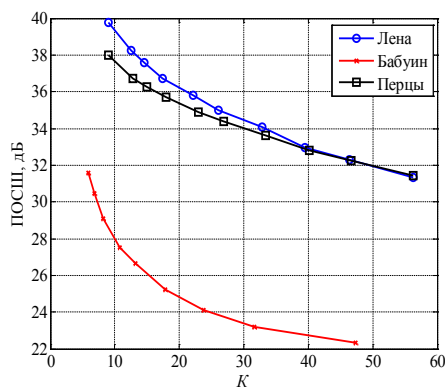


Рис. 1.15. Работа алгоритма JPEG2000 для всех тестовых изображений

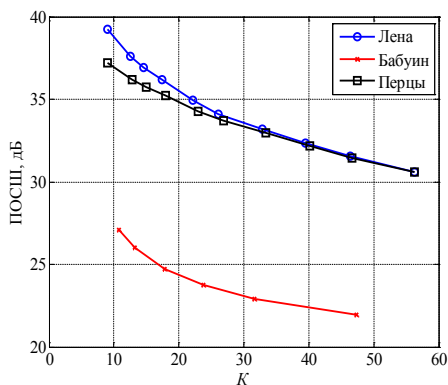


Рис. 1.16. Работа алгоритма SPIHT для всех тестовых изображений

Таким образом, проведенный анализ алгоритмов сжатия изображений относительно критериев ПОСШ и УИК позволяет сделать следующие выводы:

1. Стандарт JPEG проигрывает алгоритмам сжатия на основе вейвлет-преобразования 2–7 дБ по шкале оценок ПОСШ. Эта разница увеличивается с ростом степени компрессии и визуально различима, начиная с $K = 25$;

2. Алгоритм SPIHT целесообразно использовать в случае сжатия малоконтрастных изображений и изображений без резких

границ и мелких деталей (особенно диагональных). Для данного типа изображений можно добиться лучшей работы алгоритма путем совмещения его с энтропийным кодированием;

3. В общем случае предпочтительным для практического использования является стандарт JPEG2000, который позволяет добиваться хороших результатов вне зависимости от особенностей изображений.

1.6. Контрольные вопросы

1. В чем отличие объективных критериев оценки качества от субъективных?

2. Чем эталонные критерии оценки качества отличаются от неэталонных?

3. Что описывается в рекомендации Международного Союза Электросвязи ITU-R BT.500-11?

4. В чем отличие субъективных оценок MOS и DMOS?

5. Запишите формулу для расчета ПОСШ.

6. Перечислите преимущества и недостатки использования оценки ПОСШ?

7. Какие диапазоны значений метрик ПОСШ и УИК встречаются в практических приложениях?

8. В чем причина того, что метрика УИК достигает более точной оценки качества по сравнению с ПОСШ?

9. Почему метрика УИК для изображений вычисляется с использованием скользящего окна, а не для всего изображения сразу?

10. Изменится ли значение метрики УИК, если рассматривать искаженное изображение как эталонное, а эталонное как искаженное?

11. Каково максимальное значение метрики КСП и для каких изображений оно достигается?

12. В чем состоит идея многомасштабного коэффициента структурного подобия?

13. Какая особенность зрительного восприятия человека позволила модифицировать метрику ПОСШ?

14. Почему зависимости оценки качества для тестового изображения «Бабуин» в разделе 1.5. существенно отличаются от аналогичных зависимостей для тестовых изображений «Лена» и «Перцы»?

Глава 2. Неэталонные алгоритмы 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 2.1).



Рис. 2.1. Классификация неэталонных алгоритмов

В первую группу входят алгоритмы, измеряющие «силу» определенного типа искажения: размытия, шума, артефактов при сжатии по стандартам JPEG и JPEG2000 и других 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 использования стандарта

JPEG измеряется величина блочности, которая имеет периодическую структуру на границах изображения.

Особое внимание на современном этапе развития уделяется синтезу неэталонных критериев оценки качества из второй группы, работа которых не ограничивается априори известным видом искажений. Такие алгоритмы в наибольшей степени востребованы на практике. К ним относятся метрики, основанные на статистике естественных изображений. Такая статистика может быть рассчитана в пространственной области, в пространстве коэффициентов дискретного косинусного преобразования или в пространстве вейвлет-преобразования. Другой подход в разработке универсальных алгоритмов неэталонной оценки качества состоит в определении параметров искаженного изображения, которые непосредственно отражаются на его качестве.

Отличительной особенностью алгоритмов, которые не учитывают априорную информацию, является необходимость процедуры предварительного обучения. Обучение универсальных алгоритмов производится на тестовой базе изображений, содержащей изображения с различными типами искажений и соответствующие им субъективные оценки качества. Популярными тестовыми базами изображений являются LIVE, TID2008, TID2013 [21–23], которые будут подробно рассмотрены в главе 3 настоящего пособия.

В работе универсальных алгоритмов можно выделить два этапа: извлечение качественных признаков из изображения и сопоставление этих признаков с соответствующей субъективной оценкой. Обучение необходимо, чтобы найти функцию, связывающую значения признаков и оценку качества на выходе. Отыскание такой функции – задача регрессии, которая решается при помощи методов машинного обучения. Для решения этой задачи в алгоритмах неэталонной оценки качества используют машину опорных векторов, рандомизированные деревья или нейронные сети. После процедуры обучения алгоритм потенциально способен оценить качество абсолютно любого изображения при условии, что рассматриваемый тип искажений присутствовал в базе.

Ниже рассмотрим, каким образом могут быть построены алгоритмы неэталонной оценки качества для самых популярных стандартов сжатия изображений – JPEG и JPEG2000. Эти алгоритмы учитывают априорную информацию о типе искажений в изображении.

2.1. Неэталонный алгоритм оценки качества для изображений, сжатых по стандарту JPEG

Рассмотрим JPEG – изображение при высоких коэффициентах сжатия (рис. 2.2). Невооруженным глазом заметно, оно имеет блочную структуру с размером блока 8х8 пикселей. Поскольку артефакты блочности, возникающие в горизонтальных и вертикальных направлениях, ничем не отличаются друг от друга, предлагаемый алгоритм будет описан только для оценки горизонтальных артефактов блочности.

Каждый блок можно представить как сумму постоянной составляющей и независимого равномерно распределенного белого шума с нулевым математическим ожиданием и неизвестной дисперсией.



Рис. 2.2. Фрагмент тестового изображения «Скарлетт», сжатый алгоритмом JPEG

Рассмотрим два соседних блока 8×8 пикселей b_1 и b_2 со средними значениями μ_1 и μ_2 соответственно, где $\mu_1 \neq \mu_2$. Таким образом, эти блоки можно описать следующими выражениями:

$$b_1 = \mu_1 + \varepsilon_{ij}, \quad b_2 = \mu_2 + \delta_{ij},$$

где ε_{ij} и δ_{ij} – слагаемые, являющиеся независимым равномерно распределенным белым шумом с нулевым математическим ожиданием. Когда коэффициенты ДКП соответствующих блоков b_1 и b_2 квантуются с большим шагом квантования, многие из них становятся равны нулю, что снижает переменные составляющие μ_1 и μ_2 . В результате ступенчатая двумерная функция между b_1 и b_2 может стать видимой (из-за того, что $\mu_1 \neq \mu_2$), создавая артефакт блочности, как показано на рис. 2.2. Основываясь на этом наблюдении, мы формируем новый смещенный блок \hat{b} , составленный из правой половины b_1 и левой половины b_2 , как показано на рис. 2.3.

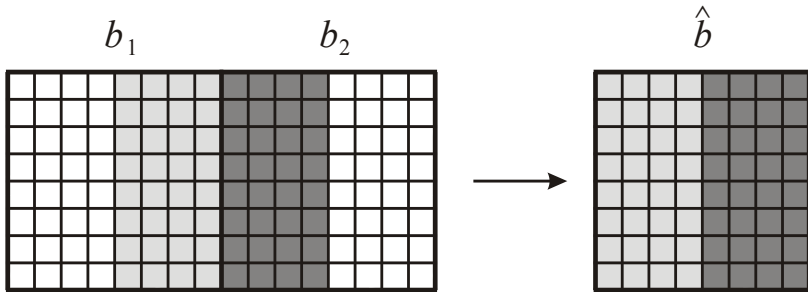


Рис. 2.3. Иллюстрация формирования нового смещенного блока \hat{b} . Артефакт блочности между блоками b_1 и b_2 может быть смоделирован как ступенчатая двумерная функция в блоке \hat{b} . Определим ступенчатый двумерный блок s в новом смещенном блоке \hat{b} как

$$s(i, j) = \begin{cases} -\frac{1}{8}, & \text{при } i = 0, \dots, 7; j = 0, \dots, 3, \\ \frac{1}{8}, & \text{при } i = 0, \dots, 7; j = 4, \dots, 7. \end{cases}$$

Таким образом, смещенный блок можно представить в виде следующего выражения:

$$\hat{b}(i, j) = \beta \cdot s(i, j) + \mu + r(i, j) \quad i = j = 0, \dots, 7, \quad (2.1)$$

где $|\beta|$ – это амплитуда ступенчатой двумерной функции s , μ – среднее значение блока \hat{b} , показывающее локальную яркость фона, а r – остаточный блок, который описывает локальную детализацию на границе исходных блоков b_1 и b_2 . Чем больше значение величины $|\beta|$, тем больше артефакты блочности при неизменных значениях яркости фона и локальной детализации. Далее применяется алгоритм в пространстве трансформант, с помощью которого находятся коэффициенты ДКП блока \hat{b} и величины параметров μ , β и r .

Для перехода в область ДКП-коэффициентов определим две матрицы q_1 и q_2 следующим образом:

$$q_1 = \begin{bmatrix} O & O_{4 \times 4} \\ I_{4 \times 4} & O \end{bmatrix}, q_2 = \begin{bmatrix} O & I_{4 \times 4} \\ O_{4 \times 4} & O \end{bmatrix},$$

где I – это тождественная (единичная) матрица, а O – нулевая матрица. Представим смещенный блок \hat{b} в виде

$$\hat{b} = b_1 q_1 + b_2 q_2. \quad (2.2)$$

Используя свойства линейности и дистрибутивности ДКП, легко получить отображение в пространстве ДКП:

$$\hat{B} = B_1 Q_1 + B_2 Q_2, \quad (2.3)$$

где \hat{B} , B_1 , B_2 , Q_1 и Q_2 – матрицы коэффициентов ДКП для блоков \hat{b} , b_1 , b_2 , q_1 и q_2 соответственно. Несмотря на то, что матрицы q_1 и q_2 содержат много нулей, соответствующие им в пространстве ДКП матрицы Q_1 и Q_2 не разрежены. Это означает, что для получения \hat{B} потребуется совершить много операций умножения. Однако матрица суммы $F_+ = Q_1 + Q_2$ и матрица разности $F_- = Q_1 - Q_2$ содержат много нулей (более 50 % от всех элементов), поэтому значительная экономия в вычислительных операциях может быть достигнута с помощью следующего выражения:

$$\begin{aligned}
\hat{B} &= B_1 Q_1 + B_2 Q_2 = \\
&= \frac{1}{2}[(B_1 + B_2)(Q_1 + Q_2) + (B_1 - B_2)(Q_1 - Q_2)] = \\
&= \frac{1}{2}(B_+ F_+ + B_- F_-),
\end{aligned}$$

где $B_+ = B_1 + B_2$, а $B_- = B_1 - B_2$. Отметим, что ДКП ступенчатой двумерной функции s имеет лишь четыре ненулевых элемента в первой строке, поскольку s – это матрица, постоянная в вертикальном направлении и антисимметричная в горизонтальном направлении. Пусть вектор $v = [v_0, v_1, \dots, v_7]$ – первая строка матрицы коэффициентов ДКП ступенчатой двумерной функции s . Тогда $v_0 = v_2 = v_4 = v_6 = 0$. Из свойства унитарности ДКП получаем, что

$$\|v\|_2 = \sqrt{\sum_{i=0}^7 v_i^2} = \sqrt{\sum_{m=0}^7 \sum_{n=0}^7 s^2(m, n)} = 1.$$

Таким образом, параметры из (2.1) могут быть посчитаны по формулам:

$$\mu = \hat{B}(0,0)/8, \quad (2.4)$$

$$\beta = \sum_{j=0}^7 v_j \hat{B}(0, j) = v_1 \hat{B}(0,1) + v_3 \hat{B}(0,3) + v_5 \hat{B}(0,5) + v_7 \hat{B}(0,7). \quad (2.5)$$

Пусть R – это матрица коэффициентов ДКП остаточного блока r . Тогда R можно легко получить последовательным выполнением следующих логических операций:

$$R = \hat{B}, \quad R(0,0) = 0, \quad R(0,i) = R(0,i) - \beta \cdot v_i, \quad i = 0, \dots, 7.$$

Благодаря разреженности коэффициентов ДКП, предложенный алгоритм гораздо более эффективен, чем традиционные методы, даже если используются быстрые алгоритмы ДКП.

Теперь, когда мы описали все параметры, необходимые для модели, можно приступить к описанию алгоритма измерения артефактов блочности с учетом известных характеристик зрительной системы человека (ЗСЧ).

В большинстве случаев человек является конечным получателем визуальной информации. Поэтому измерение артефактов

блочности с учетом известных характеристик ЗСЧ является целесообразным для оценки видимости артефактов блочности. Структурная маскировка, возникающая на изображениях с большим количеством мелких деталей, и яркостная маскировка – это две хорошо изученные характеристики ЗСЧ, которые в значительной степени влияют на восприятие артефактов блочности. Рассмотрим сначала структурную маскировку. Для того чтобы использовать эту характеристику человеческой зрительной системы, предположим, что общая детализация каждого блока – это просто сумма детализаций по всем пространственным частотам, причем амплитуда каждой составляющей этой суммы равна соответствующему коэффициенту ДКП. Эффект маскировки также зависит от относительной ориентации сигнала маскировки. Поскольку артефакты блочности возникают только в двух конкретных направлениях (вертикальном и горизонтальном), определим две по-разному ориентированные детализации (вертикальную A_v и горизонтальную A_h) следующим образом:

$$A_v = \sum_{v=1}^7 v \sum_{u=0}^7 |R(u, v)| ,$$

$$A_h = \sum_{u=1}^7 u \sum_{v=0}^7 |R(u, v)| .$$

Для вертикальных артефактов блочности преобладает эффект маскировки за счет горизонтальной детализации, поэтому общая детализация определяется по формуле

$$A_{total}^v = A_h + \alpha A_v .$$

Общая детализация для горизонтальных артефактов блочности определяется аналогично:

$$A_{total}^h = A_v + \alpha A_h ,$$

где $\alpha = 0,8$.

Также необходимо учитывать яркостную маскировку: видимость артефактов блочности зависит от локальной яркости фона. С учетом обоих эффектов маскировки итоговое значение видимости артефактов блочности на границе двух соседних блоков η определяется следующим образом:

$$\eta = \frac{|\beta|}{(1 + A_{total}^h) \left(1 + (\mu/\mu_0)^\gamma \right)},$$

где A_{total}^h – это общая горизонтальная детализация блока \hat{b} , $\mu_0 = 150$ и $\gamma = 2$.

Для вертикальных артефактов блочности может быть применен аналогичный метод. Результатом работы описанного выше алгоритма является вычисление видимости артефактов блочности на границе двух соседних блоков η для каждой соответствующей пары. Локальные значения могут быть объединены в единый неэталонный индекс блочности (НИБ) для всего изображения.

$$\text{НИБ} = \sqrt[p]{\frac{1}{N} \sum_{k=1}^N \eta_k^p},$$

где N – это общее число границ между всеми блоками изображения, а показатель степени – $p = 4$.

Проведем исследование описанного алгоритма для фрагмента тестового изображения «Бабуин» (рис. 2.4). При низких коэффициентах сжатия присутствие блочности практически незаметно. Напротив, при высоких степенях сжатия количество артефактов блочности возрастает, а индекс НИБ соответственно увеличивается.



а) оригинальное изображение



б) JPEG-изображение, $K = 6,3$:
(НИБ = 0,92, ПОСШ = 27,85 дБ)



в) JPEG-изображение, $K = 17,8$:
(НИБ = 8,21, ПОСШ = 23,46 дБ)

г) JPEG-изображение, $K = 31,6$:
(НИБ = 25,36, ПОСШ = 21,54 дБ)

Рис. 2.4. Результаты тестирования НИБ на изображении «Бабуин»

На рис. 2.5 приведен пример использования предложенного алгоритма в задаче измерения артефактов блочности для трех тестовых изображений с различной степенью детализации («Лена», «Бабуин», «Перцы»), сжатых по стандарту JPEG. Отметим тот факт, что существуют различия в приведенных кривых. Действительно, изображение «Бабуин» с точки зрения расчета НИБ является более «блочным». Этот факт объясняется наличием огромного числа естественных границ в данном тестовом изображении.

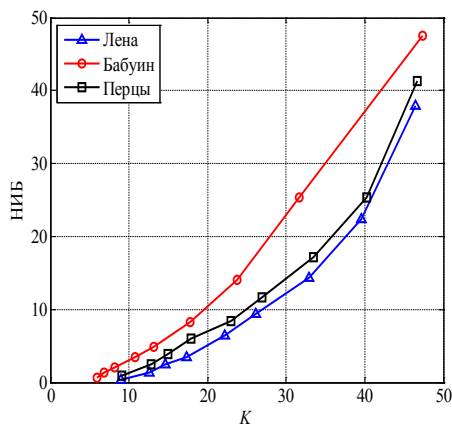


Рис. 2.5. Зависимость оценки НИБ от коэффициента сжатия изображения

2.2. Неэталонный алгоритм оценки качества для изображений, сжатых по стандарту JPEG2000

Нерегулярность, возникающая в изображениях, сжатых по стандарту JPEG2000, связана с тем, что процессы кодирования и декодирования являются итерационными. В итоге декодированное изображение содержит такие артефакты, как размытие и звон. Размытие возникает по причине затухания высоких частот в изображении и характеризуется расплыванием границ и общей потерей детальности. Явление звона вызвано квантованием высокочастотных коэффициентов и проявляется в виде ряби около резких границ на изображении. Данные искажения коррелированы с исходным изображением и плохо детектируются классическими оценками СКО и ПОСШ, т. к. данные метрики изначально разрабатывались для оценки искажений и шумов, некоррелированных с изображением. Опишем процесс построения неэталонного индекса качества для изображений стандарта JPEG2000 (НИК2000). Данная метрика рассчитывается в области вейвлет-коэффициентов, что позволяет учитывать последствия квантования вейвлет-коэффициентов.

Будем использовать статистическую модель изображения (СМИ) в роли эталона, с которым можно сопоставить оцениваемое изображение. Данная модель отображает статистические взаимозависимости вейвлет-коэффициентов изображений в каждом поддиапазоне вейвлет-разложения и их корреляцию с другими вейвлет-коэффициентами аналогичных поддиапазонов в последующих уровнях разложения. Она подходит для измерения эффекта квантования вейвлет-коэффициентов изображений, т. к. квантование зануляет маленькие по величине коэффициенты. Результатом этого процесса является увеличение вероятности нахождения нулевых коэффициентов в любом поддиапазоне в сравнении с ожидаемой вероятностью для обычных изображений.

Статистическая модель оценивает величину вейвлет-коэффициента C , определяемого величиной линейного предсказания коэффициента P :

$$C = MP + N$$

$$P = \sum_{i=1}^n l_i C_i,$$

где M и N – независимые случайные переменные с нулевым средним, C_i – n соседних с C коэффициентов в пространстве, направлении и иерархии и l_i – коэффициенты линейного предсказания.

На рис. 2.6 показаны совместные гистограммы $(\log_2(P), \log_2(C))$ для тестового изображения «Лена» для нескольких поддиапазонов различных уровней вейвлет-разложения. Сильная нелинейная зависимость между C и P лучше всего заметна в логарифмическом масштабе. Как видно из представленных зависимостей, СМИ хорошо описывает статистические взаимозависимости коэффициентов по поддиапазонам и направлениям вейвлет-разложения.

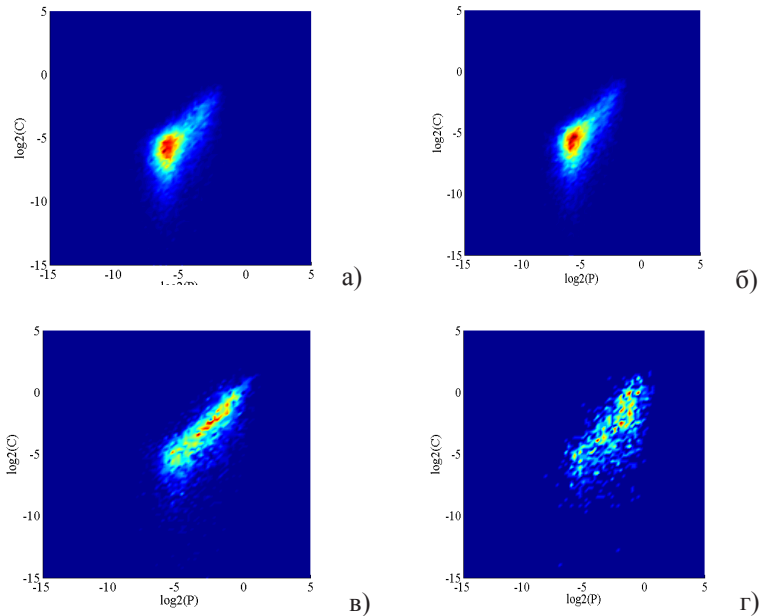
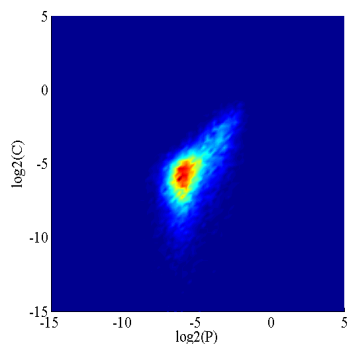


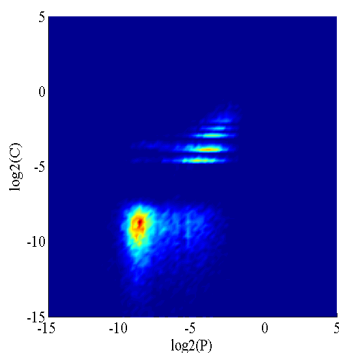
Рис. 2.6. Совместные гистограммы $(\log_2(P), \log_2(C))$ изображения для различных направлений и уровней вейвлет-разложения:

- а) диагональный поддиапазон первого уровня;
- б) горизонтальный поддиапазон первого уровня;
- в) вертикальный поддиапазон второго уровня;
- г) диагональный поддиапазон третьего уровня

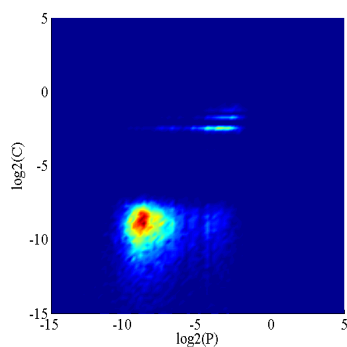
На рис. 2.7 показаны совместные гистограммы одного и того же поддиапазона для исходного изображения «Лена» и сжатых изображений. Влияние эффекта квантования очевидно: происходит сдвиг коэффициентов к началу координат и нарушается зависимость между величинами C и P , что демонстрирует отклонение от естественной СМИ.



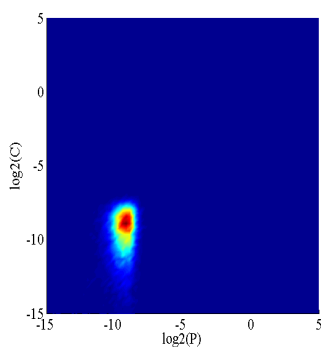
а) оригинальное изображение



б) JPEG2000-изображение,
К = 3,29



в) JPEG2000-изображение,
К = 10,26



г) JPEG2000-изображение,
К = 42,11

Рис. 2.7. Пример изменения совместных гистограмм ($\log_2(P)$, $\log_2(C)$) для одного поддиапазона изображения «Лена» при различных степенях сжатия алгоритмом JPEG2000

На практике можно использовать упрощенную модель двух состояний изображения в вейвлет-области. Эти два состояния соответствуют тому, значителен или незначителен коэффициент или его предсказатель. Коэффициент или его предсказатель считаются значительными, если их значения превышают порог. Совместная модель двух состояний обусловлена тем, что в результате процесса квантования в JPEG2000, который проявляется во всех поддиапазонах, большое количество значений C и P являются менее значительными, чем ожидалось для естественных (без сжатия) изображений. Следовательно, хорошим показателем отклонения изображения от оригинала и наличия визуальных эффектов квантования является малая пропорция значительных C и P .

Два порога (один – для коэффициента C и один – для предсказателя P) выбираются для каждого поддиапазона и зависят от изображения. Подробности выбора порогов будут представлены ниже. В результате мы получаем набор из четырех эмпирических вероятностей: p_{ii} , p_{is} , p_{si} и p_{ss} , соответствующих вероятностям того, что пара предсказатель/коэффициент лежит в одном из четырех квадрантов. Набор соседей, по которым высчитывается P , показан на рис. 2.8.

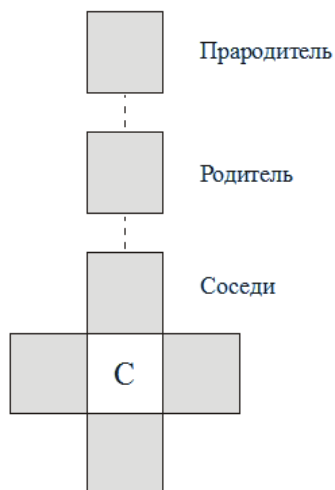


Рис. 2.8. Пример расчета предсказателя

Результаты моделирования показывают, что вероятности p_{ss} являются хорошими показателями потери визуального качества сжатых изображений. Вычисляются соответствующие вероятности для шести поддиапазонов: горизонтального, вертикального и диагонального для первого и второго уровней вейвлет-разложения. Сочетание всех этих параметров должно быть нелинейным, т. к. статистические зависимости отличаются для различных направлений и уровней. Например, вероятность для лучшего поддиапазона убывает быстрее с увеличением степени сжатия, чем вероятность для более грубого поддиапазона. Для того чтобы скомпенсировать эти различия, предлагается независимо друг от друга нелинейно преобразовать каждую вероятность, после чего объединить их в линейную зависимость.

Для преобразования вероятности используется следующее выражение:

$$q_i = K_i \left(1 - \exp \left(- \frac{(p_{ss,i} - u_i)}{T_i} \right) \right),$$

где q_i – преобразованная вероятность (предсказанное качество изображения) для i -го поддиапазона; $p_{ss,i}$ – вероятность p_{ss} для i -го поддиапазона; K_i , T_i и u_i – параметры, аппроксимируемые кривой и получаемые из обучающего набора данных. Параметры подбираются таким образом, чтобы корреляция между результатами работы алгоритма (объективными оценками) и точками кривой – результатом нелинейной регрессии субъективных и объективных оценок для изображений из обучающего набора – была наибольшей.

Взвешенное среднее от преобразованных параметров используется для предсказания качества изображений. По причине схожести статистических зависимостей по горизонтальным и вертикальным поддиапазонам для данной иерархии предполагается, что их веса одинаковы. Таким образом, вектор качества поддиапазона размерности шесть $q = \{q_i | i \in 1 \dots 6\}$ преобразуется в вектор размерности четыре q' путем усреднения предсказаний качества горизонтального и вертикального поддиапазона для данного уровня. Окончательно неэталонным индексом

качества изображений стандарта JPEG2000, лежащим в интервале $[0, 100]$, будет взвешенное среднее \mathbf{q}' :

$$\begin{bmatrix} q'_1 \\ q'_2 \\ q'_3 \\ q'_4 \end{bmatrix} = \begin{bmatrix} \frac{(q_1 + q_2)}{2} \\ q_3 \\ \frac{(q_4 + q_5)}{2} \\ q_6 \end{bmatrix},$$

$$\text{НИК} = \mathbf{q}'^T \mathbf{w},$$

где веса \mathbf{W} могут быть получены путем минимизации ошибки предсказания качества при использовании данных обучающего массива.

Необходимое условие для неэталонной оценки – отражение изменений в визуально воспринимаемом качестве сжатого изображения, которое не должно зависеть от изменений в его содержании. Способ, описанный выше, зависит не только от степени квантования, но и от изменений в содержании изображения. Например, если значения вероятностей сделать постоянными, тогда малое значение p_{ss} может определять как сильно квантованное изображение, так и относительно однородное изображение (например, небо и облака) с невысоким уровнем квантования. Чтобы сделать данную характеристику менее зависимой от содержания изображения, предлагается определять пороги для каждого типа изображений, так чтобы они были ниже для однородных изображений и выше для высоко детализированных изображений. Для решения этой задачи применяется СМИ в комбинации с моделями искажений.

Для тестирования использовались три полутоновых изображения с разрешением 512×512 пикселей и с разной степенью детализации: «Перцы», «Лена» и «Бабуин». Результаты работы алгоритма оценки качества для фрагмента изображения «Лена» приведены на рис. 2.9.



а) оригинальное изображение



б) сжатое изображение,
 $K = 17$: (НИК2000=60,12;
ПОСШ=36,61 дБ)



в) сжатое изображение,
 $K = 50$: (НИК2000=35,62;
ПОСШ=31,96 дБ)



г) сжатое изображение,
 $K = 100$: (НИК2000=31,61;
ПОСШ=28,97 дБ)

Рис. 2.9. Результаты тестирования критериев качества
на изображении «Лена»
с различными коэффициентами сжатия

На рис. 2.10 представлены зависимости оценки НИК2000 от коэффициента сжатия K , полученные для трех тестовых изображений. Отметим, что эти зависимости носят монотонно спадающий характер. Причиной этому служит проявление на изображении искажений, вносимых алгоритмом JPEG2000 с ростом коэффициента сжатия. Таким образом, индекс НИК2000 не только корректно определяет качество изображений, но и находится в хорошей корреляции с традиционно используемыми объективными эталонными оценками.

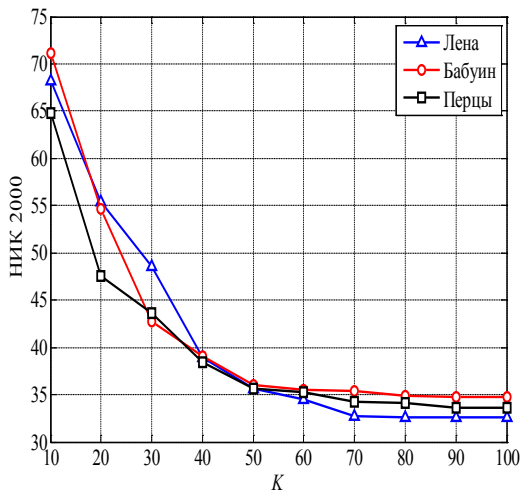


Рис. 2.10. Зависимость оценки НИК2000 от коэффициента сжатия для различных тестовых изображений

2.3. Контрольные вопросы

1. Перечислите основные классы неэталонных алгоритмов оценки качества.
2. Какие два этапа можно выделить в работе универсальных алгоритмов оценки качества?
3. Почему при сжатии по стандарту JPEG возникают артефакты блочности?
4. Каким образом можно смоделировать артефакты блочности JPEG?
5. Какие характеристики зрительной системы человека влияют на восприятие артефактов блочности?
6. Опишите этапы расчета неэталонного индекса блочности JPEG-изображения.
7. В чем отличие вейвлет-преобразования от дискретного косинусного преобразования?

8. Какие артефакты сжатия характерны для изображений, сжатых по стандарту JPEG2000?

9. Опишите основные этапы алгоритма расчета неэталонного индекса НИК2000.

10. Какой диапазон значений принимает оценка НИК2000?

Глава 3. Базы тестовых изображений с экспертными оценками

Субъективные оценки широко используются как наиболее точная мера качества изображений, поскольку именно визуальная система человека является финальным приемником в большинстве практических приложений. Важным вкладом в развитие рассматриваемой области в последние два десятилетия являются собранные базы данных изображений с субъективными оценками. К ним относятся: LIVE (Laboratory for Image & Video Engineering, база изображений, собранная в лаборатории обработки изображений и видеоданных университета г. Остин, штат Техас, США), TID2008 (Tampere Image Database, база изображений, собранная в университете г. Тампере, Финляндия) и TID2013 (обновленная версия TID2008) [21–23]. Рассмотрим их более подробно.

3.1. База тестовых изображений LIVE

База данных LIVE получена из набора исходных 29 эталонных изображений с различным содержанием. Они включали в себя фотографии лиц, людей, животных, снимки природы, искусственных объектов и изображения без какого-либо конкретного интересующего объекта (рис. 3.1). Изображения подвергались обработке с использованием следующих моделей искажений:

– сжатие JPEG: искаженные изображения были получены путем сжатия эталонных изображений по стандарту JPEG с параметром сжатия в диапазоне от 0,15 бит на пиксель (BPP) до 3,34 BPP;



Рис. 3.1. Эталонные изображения из базы LIVE

– сжатие JPEG2000: искаженные изображения были получены путем сжатия эталонных изображений по стандарту JPEG2000 со значением параметра сжатия в диапазоне от 0,028 BPP до 3,15 BPP;

– аддитивный белый гауссовский шум со значением стандартного отклонения σ добавлялся ко всем компонентам RGB-изображения после их масштабирования. Значения стандартного отклонения σ изменялись от 0,012 до 2,0;

– размытие по Гауссу: три цветовых 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: моделировалась передача потока данных сжатого по стандарту JPEG2000, при этом отношение сигнал/шум 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. Примеры искаженных изображений из базы LIVE показаны на рис. 3.2.

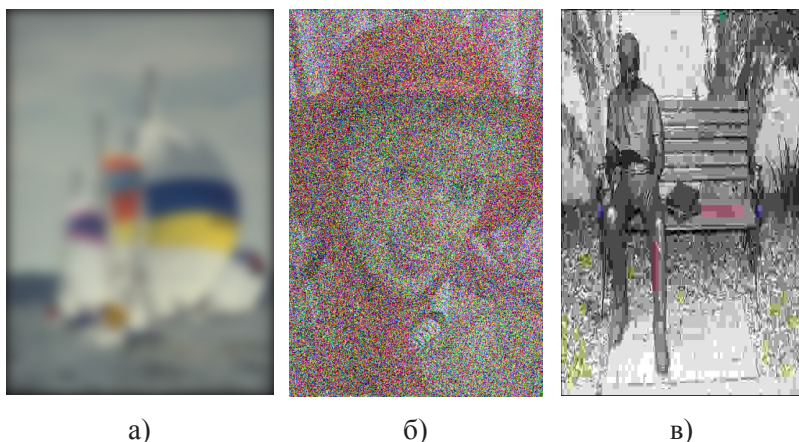


Рис. 3.2. Примеры искаженных изображений из базы LIVE:
а) размытие по Гауссу; б) белый шум; в) JPEG

3.2. База тестовых изображений TID2008

База изображений TID2008 была собрана исследователями из Финляндии, Италии и Украины. Она включает в себя 1700 искаженных изображений, сгенерированных из 25 эталонных, с использованием 17 типов и 4 уровней искажений. Использовались следующие модели искажений:

- 1) аддитивный гауссовский шум;
- 2) аддитивный шум в цветовых компонентах;
- 3) шум с пространственной корреляцией;
- 4) маскирующий шум;
- 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) модель шума, возникающая при преобразовании изображения в карту контуров (Non eccentricity pattern noise);
- 15) локальные внутриблоковые искажения разной интенсивности;
- 16) сдвиг среднего;
- 1) изменение контрастности.

Субъективное качество изображений оценивалось по шкале MOS и изменялось от 0,2 до 7,3.

3.3. База тестовых изображений TID 2013

База тестовых изображений TID2013 создана как расширение базы TID2008. Она содержит 3000 изображений, которые получены искажением 25 эталонных изображений, 24 из которых это изображения из базы Kodak, а 25-е эталонное изображение создано искусственным образом (рис 3.3–3.4).

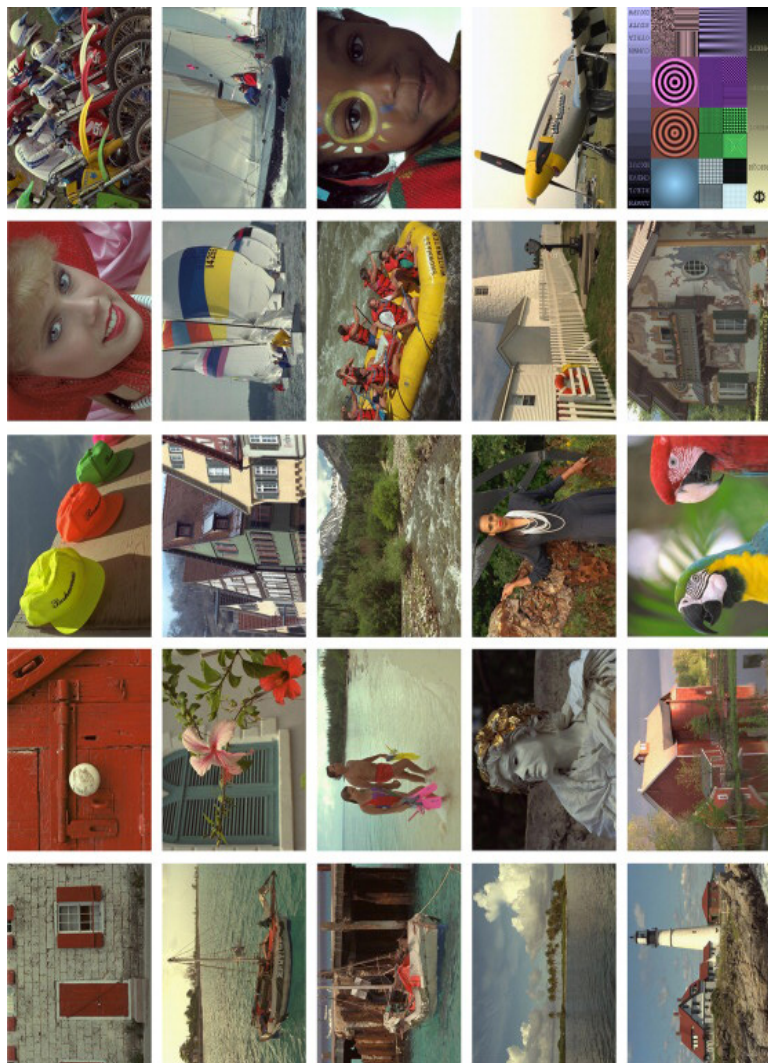


Рис. 3.3. Примеры изображений из базы TID2013



а)



б)



в)



г)



д)



е)

Рис. 3.4. 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 JREG, ПОСШ 24,05 дБ; г) шум квантования, ПОСШ 24,21 дБ; д) импульсный шум, ПОСШ 26,99 дБ; е) ошибки декодирования кодеком JREG2000, ПОСШ 23,98 дБ

Использовалось 24 типа искажения, 17 из которых перечислены выше в описании базы TID2008 и 7 дополнительных:

- 1) изменение насыщенности;
- 2) мультипликативный шум Гаусса;
- 3) комфортный шум (Comfort noise);
- 4) сжатие зашумленных изображений с потерей информации;
- 5) шум квантования для цветных изображений;
- 6) хроматические аберрации;
- 7) пространственное сэмплирование и реконструкция.

Параметры искажений выбирались такими, чтобы после их воздействия качество изображения соответствовало одному из пяти уровней ПОСШ: 21 дБ, 24 дБ, 27 дБ, 30 дБ, 33 дБ. Субъективное качество изображений оценивалось по шкале MOS и изменялось от 0,2 до 7,3.

3.4. Метрики, используемые для тестирования алгоритмов оценки качества изображений

Пусть $\{x_i\}^n = \{x_1, x_2, \dots, x_n\}$ – оценки качества на выходе алгоритма незаталонной оценки качества изображений для n тестовых изображений, 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:

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

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

Ранговая корреляция используется для изучения связи между ординальными (порядковыми) переменными, т. е. переменными, представленными в порядковой шкале. В порядковой шкале для двух значений можно установить лишь порядок, в котором они расположены по степени проявления признака. Например, для субъективной оценки качества изображения мож-

но ввести порядковую шкалу из трёх значений: плохое качество, среднее качество, хорошее качество. Для двух ординальных значений, однако, нельзя установить, во сколько раз или на сколько одно из них больше или меньше другого.

Проблема оценки связи между ординальными оценками разрешима, если упорядочить их по качеству. При этом каждому из них ставится в соответствие специальный номер, называемый рангом.

Коэффициент ранговой корреляции Спирмена находится по формуле:

$$\rho = 1 - \frac{6 \sum_{i=1}^n (r_i - s_i)^2}{n^3 - n},$$

где r_i и s_i – ранги i -й оценки в рядах x и y соответственно, n – число пар наблюдений.

При проверке значимости ρ исходят из того, что в случае справедливости нулевой гипотезы об отсутствии корреляционной связи между переменными статистика

$$t = \frac{\rho \sqrt{n-2}}{\sqrt{1-\rho^2}}$$

имеет t -распределение Стьюдента с степенями свободы. Из этого следует, что ρ значим на уровне α , если фактически наблюдаемое значение t (в абсолютной величине) будет больше критического, или

$$|t| > t_{1-\alpha, n-2},$$

где $t_{1-\alpha, n-2}$ – табличное значение t -критерия Стьюдента, определенное на уровне значимости при числе степеней свободы $n-2$.

Коэффициент корреляции Кендалла рассчитывается по формуле:

$$\tau = 1 - \frac{4}{n(n-1)} K,$$

$$K = \sum_{i=1}^{n-1} \sum_{j=i+1}^n [x_i < x_j] \neq [y_i < y_j].$$

Коэффициент τ принимает значения от $[-1;1]$, равенство $\tau = 1$ указывает на строгую линейную зависимость между x и y , $\tau = -1$ – на обратную. Отметим, что значим на уровне α , если значение статистики

$$t = \tau \sqrt{\frac{9n(n-1)}{2(2n+5)}}$$

по абсолютной величине больше критического $\tau_{1-\alpha}$, где $\Phi(\tau_{1-\alpha}) = 1-\alpha$ и Φ – функция Лапласа.

3.5. Контрольные вопросы

1. Для чего требуются базы данных изображений с субъективными оценками?

2. В каких университетах собраны наиболее популярные тестовые базы изображений?

3. Перечислите основные типы искажений изображений, встречающихся в базах данных.

4. В чем отличие базы TID2013 от базы TID2008?

5. Что показывает оценка корреляции между субъективными и объективными оценками?

6. Как рассчитывается коэффициент корреляции Пирсона?

7. Что такое ранговая корреляция?

8. Запишите формулу для расчета коэффициента ранговой корреляции Спирмена?

9. Запишите формулу для расчета коэффициента корреляции Кендалла?

10. Что означает равенство коэффициента корреляции Кендалла $\tau = -1$?

Глава 4. Оценка качества видеоизображений

4.1. Принципы обработки и кодирования видеоизображений

Качество цифрового видеоизображения определяется следующими тремя основными факторами [7]:

- частотой дискретизации видеопотока во временной области, определяющей количество кадров, формируемых за секунду;
- разрешением, т. е. количеством пикселей в кадре;
- качеством кодирования отдельного кадра. Самой распространенной причиной снижения качества видеоизображения является процедура сжатия видеоданных.

Алгоритмы кодирования видеосигнала используют разные частоты дискретизации. Для естественного восприятия видеоизображения человеку достаточно частоты 25 кадров в секунду. Во многих практических приложениях, например при видеонаблюдении, используются уменьшенные частоты дискретизации от 1 до 10–15 кадров в секунду. В современных мультимедийных сервисах (например, в телевидении высокой чёткости) используются большие частоты дискретизации (до 60 кадров в секунду), что обеспечивает формирование в точке приёма практически натурального видеоизображения.

Существует множество стандартизированных вариантов разрешения кадра. Широкое распространение получил формат CIF (обобщенный промежуточный формат, Common Intermediate Format) и производные от него форматы. В традиционных форматах телевидения обеспечивается соотношение сторон кадра 4:3, в телевидении высокой чёткости – 16:9. В современных системах цифрового телевидения активно используются разрешения 720p (1280x720 пикселей), Full HD (1920x1080 пикселей), 4K (4096x3072 пикселей), 8K (8192x4320 пикселей) и другие.

Сжатие видеоизображений – это технология уменьшения размера видеоданных с сохранением приемлемого уровня качества, с минимизацией видимых искажений. Процедура сжатия

обеспечивает эффективность передачи и хранения видеоданных. Процедура сжатия состоит в сокращении его статистической и психофизической (психовизуальной) избыточности. К ней относятся:

- *пространственная (внутрикадровая) избыточность*, состоящая в том, что значения координат соседних пикселей в цветовом пространстве коррелированы между собой;

- *энтропийная (вероятностная) статистическая избыточность*, заключающаяся в значительной неодинаковости частот встречаемости разных яркостей и цветов в изображении;

- *временная (межкадровая) статистическая избыточность* проявляющаяся в значительной схожести соседних кадров подвижного изображения;

- *психофизическая (психовизуальная) избыточность*, заключающаяся в том, что зрительный аппарат человека имеет различную чувствительность к изменениям яркости и цветов изображения, к крупным и мелким деталям изображения, к качеству изображения в области контуров и вне их.

Для устранения избыточности видеоизображения в алгоритмах сжатия используются следующие процедуры:

- межкадровое предсказание, которое позволяет уменьшить временную избыточность;

- внутрикадровое предсказание и кодирование сигнала изображения на основе дискретного косинусного преобразования, направленные на уменьшение пространственной избыточности;

- энтропийное кодирование на основе статистических зависимостей изображения (например, использование кода Хаффмана).

В большинстве современных кодеков видеосигнала применяются три типа кадров, значительно отличающихся как используемыми принципами кодирования, так и качеством:

- **I-кадры (Intra Coded Pictures, интра-кадры)** – кадры, кодируемые независимо от других. Процедуры их кодирования в основном соответствуют процедурам JPEG. Интра-кадры также зачастую называют ключевыми;

- **P-кадры (Predictive Coded Pictures)** – кадры, кодируемые на основе предсказания (с учётом компенсации движения) по предыдущим I- или P- кадрам. Степень их сжатия выше, чем у интра-кадров;

– **В-кадры (Bidirectionally-predictive Coded Pictures)** – кадры, кодируемые на основе двунаправленного предсказания как по предыдущим, так и по последующим I- и P-кадрам. Они отличаются максимальной степенью сжатия.

Наличие I-кадров является обязательным, их доля в общем количестве кадров определяет скорость передачи данных. Большее количество промежуточных кадров, в свою очередь, увеличивает вычислительные затраты на декодирование сигнала.

Помимо различных алгоритмов кодирования, существуют различные форматы записи и хранения видеофайлов и сопровождающих их аудиофайлов. Применение таких способов на практике приводит к созданию различных типов видеоконтейнеров, определяющих структуру мультимедийных данных. К их числу относятся контейнеры AVI, MKV, MOV, MPEG, WebM и другие.

4.2. Стандарты кодирования видеоизображений

Наиболее популярные на данный момент стандарты компактного цифрового представления видеоизображений подготовлены специалистами двух организаций: **MPEG** и **ITU-T** [1, 7, 10].

Группа экспертов по кодированию подвижных изображений (MPEG – Moving Picture Coding Experts Group) образована в 1988 г. международной организацией по стандартизации (ISO – International Organization for Standardization) и международной электротехнической комиссией (International Electro-Technical Commission – IEC). Соответствующие алгоритмы составляют линейку стандартов MPEG. Эксперты для группы MPEG собираются из профильных коммерческих компаний, исследовательских организаций, институтов, имеющих мировую известность. Группа MPEG организует свою работу во время регулярных встреч, которые проходят четыре раза в год. Результаты встреч оформляются в виде комплекта входных и выходных документов. Входные документы включают в себя результаты работы подгрупп на предварительных встречах, предложения по добавлениям

или изменениям в разрабатываемые стандарты, а также предложения от других корпораций. Выходные документы включают отчеты конференций и черновые версии стандартов.

Сектор стандартизации в области телекоммуникаций Международного союза электросвязи (Telecommunication Standardization Sector of International Telecommunications Union – ITU-T) разрабатывает стандарты и рекомендации для телекоммуникационной области. Им сформирована группа экспертов по видеокодированию (VCEG – Video Coding Experts Group). Членство в VCEG открыто для всех заинтересованных сторон (при условии одобрения председателем). Члены группы VCEG встречаются каждые три месяца для рассмотрения новых начинаний и предложений, некоторые из них включаются в черновые проекты стандартов. Соответствующие алгоритмы составляют линейку стандартов H.26х.

В 2001 г. для улучшения координации работ группы MPEG и VCEG создали объединенную видеокomанду JVT (Joint Video Team). Рассмотрим существующие на сегодняшний день стандарты видеокодирования более подробно.

MOTION JPEG формируется из серии отдельных JPEG-изображений. Все изображения в видеопоследовательности имеют одно и то же качество, определяемое уровнем сжатия.

H.120 – первый стандарт видеокodeка, разработанный в 1984 г. Первая версия кодека использовала скалярное квантование и кодирование переменной длины. Вторая версия была создана в 1988 году. По сравнению с первой версией добавлены алгоритмы компенсации движения и фонового предсказания. Стандарт поддерживал битрейты между 1544 и 2048 Кбит/с.

Стандарт H.261 стал первым распространенным стандартом, разработанным ITU-T для поддержки видеотелефонии и видеоконференций по сетям ISDN с коммутацией каналов. Эти сети работали на скорости 64 кбит/с. Ключевыми особенностями этого стандарта являлись скалярное квантование, сканирование зигзагом и кодирование переменной длины. Частота кадров кодируемой видеопоследовательности ограничена 30 кадрами в секунду. Стандарт поддерживает CIF и QCIF размеры кадров с разрешением 352×288 и 176×144 пикселей соответственно. Вторая версия

стандарта от 1993 г. обладала обратно-совместимым режимом высокого разрешения с битрейтами между 64 и 2048 Кбит/с.

Стандарт MPEG-1 принят в 1993 г. Это результат дальнейшей разработки H.261, позволивший дополнительно повысить качество видеоизображений. Важными нововведениями в стандарте явились алгоритмы двустороннего предсказания движения, введение понятия «полупиксельного движения», кодирование, основанное на «слайсах», и использование весовых матриц квантования. Максимальное допустимое разрешение кадра стандарта составляет 768×576 пикселей. Стандарт MPEG-1 был разработан для приложений просмотра и хранения видео на компакт-дисках CD. На такой диск помещалось около 70 минут аудио+видео при скорости записи 1,4 Мбит/с. Соответствующий формат Video CD должен был конкурировать с видео на аналоговых VHS-кассетах. Однако стандарт не получил коммерческого развития, потому что качество видеоизображения было ненамного выше, чем у аналоговых видеокассет, что не побудило потребителей массово перейти на цифровую технологию.

В 1994 г. разработан **стандарт MPEG-2**, который широко используется для хранения видеоданных на дисках DVD. Стандарт был заточен под диапазон используемых битрейтов от 2 до 5 Мбит/с. В стандарте MPEG-2 были впервые реализованы концепции профилей и уровней (задающие точки соответствия и пределы производительности) как способ поощрения взаимодействия разработчиков приложений без ущемления гибкости стандарта. MPEG-2 имел большой коммерческий успех. Он обеспечил кодирование фильмов в формате DVD-Video, который смог вытеснить аналоговый формат VHS, что не смогли сделать MPEG-1 и Video-CD.

Стандарт H.263 разработан в 1995 г. Он имел лучшее качество для всех битрейтов по сравнению с предыдущими стандартами. Кроме того, его отличали полупиксельная компенсация движения, медианное предсказание вектора движения, увеличенный диапазон векторов движения с экстраполяцией изображения, улучшенная эффективность сжатия и выбор референсного изображения на блочном и макро-блочном уровнях.

Стандарт MPEG-4 принят в 1999 г. и поддерживал как приложения с низкой пропускной способностью, так и приложения, очень требовательные к качеству изображения. В стандарте не накладывалось ограничений на частоту кадров, и теоретически он поддерживает любую пропускную способность. MPEG-4 обеспечивал свое функционирование посредством набора инструментов кодирования, организованных в «профили», которые состоят из рекомендуемых групп модулей для конкретных приложений. Классы профилей включают «простые» профили (кодирующие прямоугольные видеокadres), профили на основе объектов (кодирующие визуальные объекты прямоугольной формы), текстурные профили (кодирующие неподвижные изображения или текстуры), масштабируемые профили (кодирующие многократные разрешения или качественные уровни) и студийные профили (кодирующие высококачественные студийные приложения).

Стандарт H.264 позволяет добиться уменьшения видеозображения примерно на 80 % по сравнению с MOTION JPEG и на 50 % по сравнению с MPEG-4, сохраняя исходное качество изображения. Этот стандарт разработан для эффективного кодирования и помехоустойчивой передачи видео прямоугольного формата. Его первоначальной задачей было обеспечение функциональности, схожей с H.263+ и MPEG-4 (простой профиль), но с лучшим сжатием и с более надежным механизмом передачи видеоданных. Целевые приложения стандарта включают в себя двусторонние видеокommunikации (видеоконференции и видеотелефонию), видео высокого качества и организацию видеопотоков через сети пакетной передачи данных. Кодер традиционно использует режимы внутрикадрового и межкадрового предсказаний, включая процедуры компенсации движения. Кодируемый кадр сигнала подвижного изображения разбивается на микроблоки размером 16x16 пикселей.

Стандарт H.265 (High Efficiency Video Coding, HEVC) принят в 2013 г. Он разрабатывался с прицелом на увеличение разрешения видео и увеличение использования архитектур параллельных вычислений. К числу основных особенностей H.265 следует отнести:

- возможность увеличения размеров макроблоков с 16x16 до 64x64 пикселей;

- возможность параллельного декодирования различных областей текущего кадра;

- произвольный доступ к изображениям в процессе декодирования (отсутствие необходимости наличия информации о предыдущих кадрах при декодировании произвольно выбранного кадра).

Эти и другие новшества H.265 обуславливают повышение вычислительной сложности процедур кодирования и декодирования, однако соотношение скорости и качества кодирования цифрового подвижного изображения при этом улучшается.

4.3. Обзор методов оценки качества видеоизображений

Как и метрики оценки качества изображений, метрики оценки качества видеоданных можно разделить на две категории: объективные и субъективные. Объективные метрики качества делятся на эталонные метрики (full reference, FR), сокращенные эталонные (reduced reference, RR) и неэталонные (no-reference, NR). Оценка, основанная на эталонной метрике, вычисляется с учётом оригинального видеофрагмента. Сокращенная эталонная метрика использует для сравнения не исходные видеоданные, а некий набор его априорных характеристик, как, например, информацию о его текстуре. Неэталонные метрики не требуют исходного видеоизображения. Они используют только информацию из битового потока и оценивают её на уровне пикселей. Именно такие метрики наиболее востребованы в практических приложениях.

Неэталонные метрики можно разделить на три категории: основанные на пикселях, на битовом потоке и гибридные метрики (рис. 4.1).

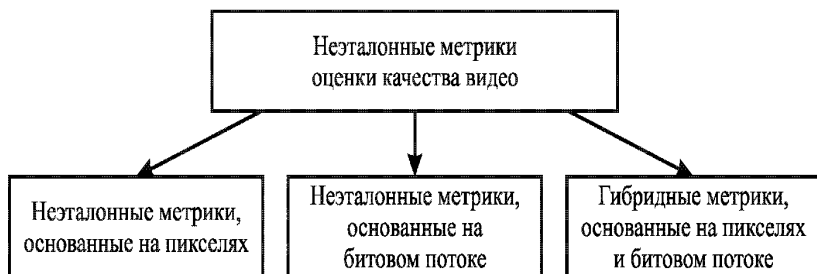


Рис. 4.1. Неэталонные метрики оценки качества видеоизображений

Неэталонные метрики, основанные на пикселях, используют измерение артефактов (размытие, блочность, звон, шум, временные нарушения) и признаков (статистика естественных сцен и другие пространственные признаки).

Неэталонные метрики оценки качества видео, основанные на битовом потоке, используют модели с соответствующими параметрами. Для параметрической модели планирования такими параметрами являются битрейт и частота потери пакетов, которые используются и в двух других моделях. Для модели уровня битового потока к ним добавляется вектор движения, а для параметрической модели пакетного уровня – кадровая частота и другие параметры.

Гибридные метрики, основанные на пикселях и битовом потоке, используют артефакты пиксельного и поточного уровней (пространственные признаки и параметры битового потока), а также статистику коэффициентов преобразования (статистику, основанную на коэффициентах ДПФ/вейвлет-преобразования).

При использовании методов, основанных на признаках, оценка качества зависит от наличия на видеоизображении артефактов, отрицательно влияющих на визуальное восприятие.

Неэталонные оценки, основанные на битовом потоке, проще в вычислении. Алгоритм вычисления оценок такого типа зависит от конкретного метода кодирования видео и формата битового потока. Такой подход удобен для контроля потоковой передачи видеоданных (рис. 4.2).

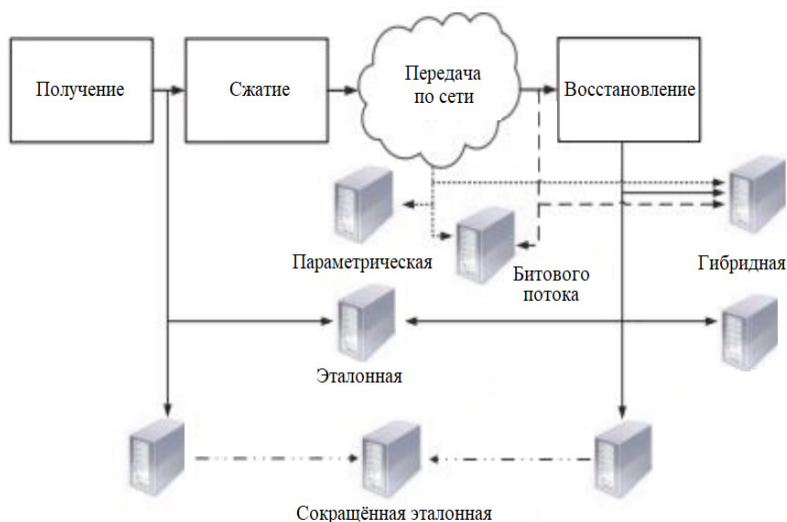


Рис. 4.2. Источники данных для различных метрик качества

Неэталонные метрики уровня битового потока делятся на три категории: параметрическая модель планирования, параметрическая модель пакетного уровня и модель уровня битового потока.

Модель планирования не имеет доступа к самому битовому потоку, но производит оценку, основываясь на битрейте, типе кодека и частоте потери пакетов. Такая модель менее сложна, чем модель пакетного уровня. Модель пакетного уровня имеет доступ к битовому потоку и может извлечь такие параметры, как частоту и типы кадров, частоту потери пакетов, текущий битрейт. Модель пакетного уровня также известна, как модель качества сервиса (QoS).

Модели уровня битового потока имеют доступ к большей части данных, используемых для оценки качества видео, кроме информации пиксельного уровня. На вход модели подаются параметры из заголовка и полезной нагрузки пакета.

Гибридный подход является комбинацией двух предыдущих подходов. Он преследует цель достичь как простоты вычисления, свойственной основанной на битовом потоке оценке, так и робастности оценки пиксельного уровня.

4.4. Неэталонные методы оценки качества, основанные на влиянии одного артефакта

Размытие. Под размытием понимается артефакт, проявляющийся в потере пространственных деталей и смягчении резкости переходов. Размытие может возникнуть по многим причинам как при получении, так и при передаче и сжатии видеоданных. Основная причина размытия при сжатии заключается в отбрасывании высокочастотных компонент сигнала. Другими возможными причинами могут быть расфокусированность камеры при съемке, движение объекта относительно камеры или технические ограничения в оптической системе.

Звон. Артефакт звона связан с феноменом Гиббса и имеет вид ряби вдоль резких краев между сглаженными участками изображения. Такое ухудшение вызывается грубым квантованием высокочастотных компонент.

Шум. К наиболее частым видам шума относятся импульсный шум модели соль-и-перец, шум квантования, Гауссов шум, а также пятна при ярком освещении. Шум может возникнуть при получении, записи, обработке и передаче видеоизображений.

Поскольку вышеописанные артефакты встречаются как на видеоданных, так и на статических изображениях, оценка их влияния на качество видеоданных отличается от оценки качества изображений только необходимостью каким-то образом объединять результаты оценок для отдельных кадров.

4.4.1. Временные повреждения

Временные (темпоральные) повреждения делятся на две основные категории: повреждения при кодировании и повреждения при передаче по сети. Наиболее часто это отображается на видеоизображении в виде следующих артефактов:

– **Толчки.** Представляют собой неплавное воспроизведение видеоизображения в результате децимации по времени;

– **Замораживание кадра.** Возникает из-за отсутствия нового кадра для показа, по причине потери пакетов и т. п.;

– **Дрожание изображения.** Воспринимается как неестественное движение по причине вариации задержки передачи;

– **Мерцание.** Заметный разрыв между последовательными кадрами, вызванный слишком низкой кадровой частотой вместе с артефактами кодирования и движением содержимого;

– **Москитный шум.** Выглядит как временное мерцание, видимое в основном в областях равномерной интенсивности. Возникает вследствие звона и ошибки предсказания, вызванной несоответствиями при компенсации движения.

На рис. 4.3 приведена общая схема оценки наличия временных повреждений. Сначала вычисляется разница между кадрами

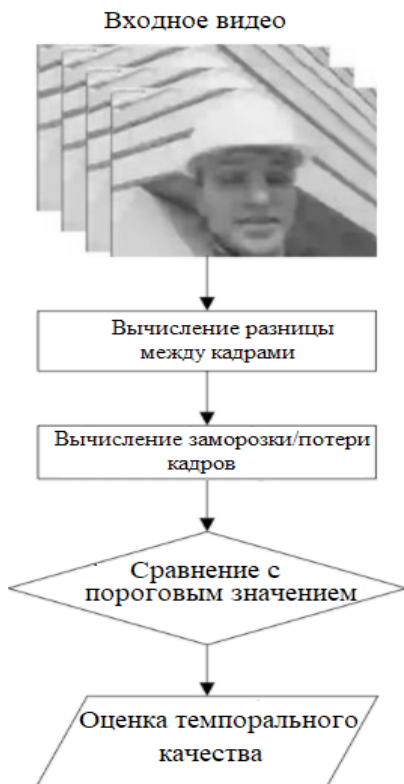


Рис. 4.3. Общая схема вычисления временных повреждений

(обычно используется только яркостный канал). Затем для определения места и вероятности заморозки/потери кадров используются различные подходы. К ним относятся, например, вычисление среднеквадратической разницы между кадрами. Далее имеет смысл применить сравнение с каким-то пороговым значением. Оно должно быть адаптивным, чтобы отличить сцены с большим количеством движения от сцен с малым количеством. Наконец, для получения общей оценки требуется механизм объединения результатов по всему видеофрагменту.

4.5. Неэталонные пиксельные метрики, основанные на нескольких артефактах

Оценки различных артефактов, обнаруживаемых на видеоданных, могут быть скомбинированы для предсказания общего воспринимаемого качества видеофрагмента (рис. 4.4).

Важность таких метрик подчеркивает тот факт, что некоторые преобразования над изображениями, например сжатие по стандарту JPEG2000, приводят к появлению сразу нескольких видов артефактов. Другое приложение таких метрик – случай, когда обработка изображения с целью удалить один артефакт (например, шумоподавление) приводит к возникновению другого артефакта.

Существуют разные подходы к тому, как следует комбинировать эффекты от различных типов повреждений. Согласно рекомендации ITU-T P.1202.1, четыре типа повреждений предлагается отсортировать согласно их степени влияния. Значения первых двух наиболее важных типов объединяются при помощи линейной комбинации с большим коэффициентом у первого. Поскольку измерения различных артефактов проводится в различных диапазонах значений, важно привести их к одному масштабу. Помимо использования линейной комбинации, в некоторых статьях предлагается использовать для объединения метрику Минковского.

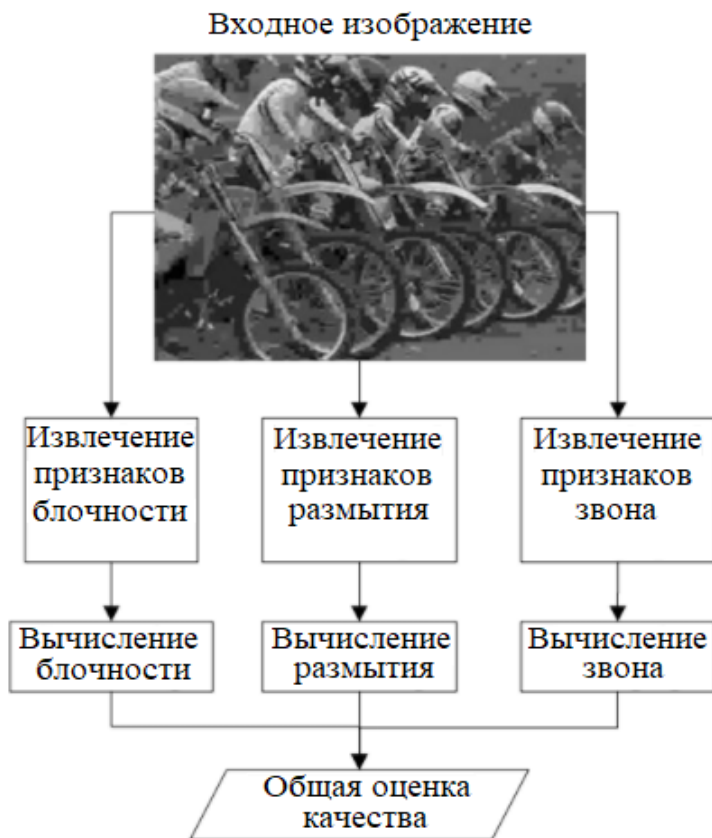


Рис. 4.4. Схема оценки качества для различных артефактов

Выбор конкретной стратегии объединения может зависеть от многих факторов, включая относительную критичность появления артефакта, пространственно-временные характеристики содержимого и наличие маскирующих эффектов. Линейная комбинация лучше подходит, если её составляющие могут быть соотнесены при помощи линейной регрессии. С другой стороны, метрика Минковского исходит из аддитивных свойств зрительной системы человека.

4.6. Контрольные вопросы

1. Какими факторами определяется качество цифрового видеоизображения?
2. Какие типовые разрешения видеоизображений характерны для практических приложений?
3. Какие виды статистической и психовизуальной избыточности присущи видеоданным?
4. Что такое I-, P- и B-кадры?
5. Какие стандарты сжатия видеоизображений нашли широкое практическое применение?
6. Каковы отличительные особенности формата видеосжатия H.265?
7. На какие категории можно разделить неэталонные методы оценки качества видео?
8. На какие категории делятся неэталонные метрики уровня битового потока?
9. Перечислите характерные типы темпоральных повреждений видеоизображений.
10. Что описывает рекомендация ITU-T P.1202.1?

Литература

1. Дворкович, В. П. Цифровые видеоинформационные системы (теория и практика) / В. П. Дворкович, А. В. Дворкович. – М. : Техносфера, 2012.
2. Методы сжатия данных. Устройство архиваторов, сжатие изображений и видео / Д. Ватолин [и др.] – М. : Диалог-МИФИ, 2002.
3. Методы компьютерной обработки изображений / под ред. В. А. Сойфера. – М. : Физматлит, 2003.
4. Сэломон, Д. Сжатие данных, изображений и звука / Д. Сэломон. – М. : Техносфера, 2004.
5. Гонсалес, Р., Цифровая обработка изображений / Р. Гонсалес, Р. Вудс. – М. : Техносфера, 2012.
6. Гонсалес, Р. Цифровая обработка изображений в среде MATLAB / Р. Гонсалес, Р. Вудс, С. Эддинс. – М. : Техносфера, 2006.
7. Афанасьев, А. А. Цифровая обработка сигналов / А. А. Афанасьев, А. А. Рыболовлев, А. П. Рыжков. – М. : Горячая линия – Телеком, 2017.
8. Приоров, А. Л. Цифровая обработка изображений : учеб. пособие / А. Л. Приоров, А. Н. Ганин, В. В. Хрящев ; Яросл. гос. ун-т. им. П. Г. Демидова. – Ярославль : ЯрГУ, 2001.
9. Приоров, А. Л. Цифровая обработка изображений : учеб. пособие / А. Л. Приоров, В. В. Хрящев, И. В. Апальков ; Яросл. гос. ун-т. им. П. Г. Демидова – Ярославль : ЯрГУ, 2007.
10. Приоров, А. Л., Хрящев В.В. Обработка и передача мультимедийной информации : учеб. пособие / А. Л. Приоров, В. В. Хрящев ; Яросл. гос. ун-тим. П. Г. Демидова. – Ярославль : ЯрГУ, 2010.
11. Simoncelli, E. Natural image statistics and neural representation / E Simoncelli, B. Olshausen // Annual Review of Neuroscience. – 2001. – V. 24, № 1. – P. 1193–1216.
12. Motoyoshi, I. Image statistics and the perception of surface qualities / I. Motoyoshi [et all] // Nature. – 2007, May. – V. 447. – P. 206–209.

13. Wang, Z. A universal image quality index / Z. Wang, A Bovik // IEEE Signal Processing Letters. – 2002. – V. 9, № 3, Mar. – P. 81–84.

14. Wang, Z., Image quality assessment: From error visibility to structural similarity / Z Wang [et. all] // IEEE Trans. Image Processing. – 2004. –V. 13, № 4, April. – P. 600–612.

15. Bovik, A. The Handbook of Image and Video Processing (Communication, Networking and Multimedia) / A. Bovik // Academic Press, 2005.

16. Wang, Z. Modern Image Quality Assessment / Z. Wang, A. Bovik. Morgan and Claypool Publishing Company, 2006.

17. Wang, Z. Mean squared error: Love it or leave it? – A new look at signal fidelity measures / Z. Wang, A. Bovik / IEEE Signal Processing Magazine. – 2009. – V. 26, № 1, January. – P. 98–117.

18. Xu, L. Visual Quality Assessment by Machine Learning / L. Xu, W. Lin, C. Kuo. Springer, 2015.

19. Deng, C. Visual Signal Quality Assessment. Quality of Experience (QoE) / C. Deng [et. all]. Springer, 2015.

20. Recommendation ITU-R BT.500-11. Methodology for the subjective assessment of the quality of television pictures, 2002.

21. Sheikh, H. LIVE image quality assessment database release 2. 2006 / H. Sheikh [et. all].

– URL: [http://live.ece.utexas.edu/research/ quality/subjective.htm](http://live.ece.utexas.edu/research/quality/subjective.htm).

22. Ponomarenko, N. TID2008 – a database for evaluation of full-reference visual quality assessment metrics / N. Ponomarenko [et. all] // Advances of Modern Radioelectronics. – 2009. – V. 10. – P. 30–45.

23. Image database TID2013: Peculiarities, results and perspectives / N. Ponomarenko [et. all] // Signal Processing: Image Communication. – 2015. –V. 30. – P. 57–77.

Оглавление

Введение.....	3
Глава 1. Эталонные алгоритмы оценки качества изображений.....	7
1.1. Пиковое отношение сигнал/шум.....	7
1.2. Универсальный индекс качества.....	12
1.2.1. Определение УИК для одномерных сигналов.....	13
1.2.2. Применение УИК для оценки качества изображений.....	15
1.3. Коэффициент структурного подобия.....	16
1.4. Модификация ПОСШ с учетом характеристик зрительной системы человека.....	24
1.5. Сравнительный анализ алгоритмов сжатия изображений относительно критериев ПОСШ и УИК.....	26
1.6. Контрольные вопросы.....	31
Глава 2. Неэталонные алгоритмы оценки качества изображений.....	32
2.1. Неэталонный алгоритм оценки качества для изображений, сжатых по стандарту JPEG.....	34
2.2. Неэталонный алгоритм оценки качества для изображений, сжатых по стандарту JPEG2000.....	41
2.3. Контрольные вопросы.....	48
Глава 3. Базы тестовых изображений с экспертными оценками.....	49
3.1. База тестовых изображений LIVE.....	49
3.2. База тестовых изображений TID2008.....	52

3.3. База тестовых изображений TID 2013.....	52
3.4. Метрики, используемые для тестирования алгоритмов оценки качества изображений.....	55
3.5. Контрольные вопросы.....	57
Глава 4. Оценка качества видеоизображений.....	58
4.1. Принципы обработки и кодирования видеоизображений.....	58
4.2. Стандарты кодирования видеоизображений.....	60
4.3. Обзор методов оценки качества видеоизображений.....	64
4.4. Неэталонные методы оценки качества, основанные на влиянии одного артефакта.....	67
4.4.1. Временные повреждения.....	67
4.5. Неэталонные пиксельные метрики, основанные на нескольких артефактах.....	69
4.6. Контрольные вопросы.....	71
Литература.....	72

Учебное издание

Матвеев Дмитрий Вячеславович
Седов Анатолий Геннадьевич
Хрящев Владимир Вячеславович
Приоров Андрей Леонидович

ОЦЕНКА КАЧЕСТВА ЦИФРОВЫХ ИЗОБРАЖЕНИЙ И ВИДЕОДАННЫХ

Учебно-методическое пособие

Верстка Е. Б. Половкова
Редактор, корректор Л. Н. Селиванова

Подписано в печать 28.09.18. Формат 60×84 ¹/₁₆.

Усл. печ. л. 4,42. Уч.-изд. л. 4,6.

Тираж 5 экз. Заказ

Оригинал-макет подготовлен
в редакционно-издательском отделе ЯрГУ.

Ярославский государственный университет
им. П. Г. Демидова.
150003, Ярославль, ул. Советская, 14.