

УДК 681.325

О.И. ЕРЕМЕЕВ, Н.Н. ПОНОМАРЕНКО, В.В. ЛУКИН*Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ»***ОЦЕНКА ТОЧНОСТНЫХ ХАРАКТЕРИСТИК МЕТОДИКИ ВЕРИФИКАЦИИ
МЕТРИК ВИЗУАЛЬНОГО КАЧЕСТВА НА ОСНОВЕ СПЕЦИАЛИЗИРОВАННЫХ
БАЗ ИЗОБРАЖЕНИЙ**

Рассмотрена задача верификации метрик визуального качества изображений на основе специализированных баз данных. Эффективность такой верификации во многом зависит от методики формирования базы тестовых изображений и проведения экспериментов по получению усредненных экспертных оценок. В работе предлагается математическая модель принятия решений участником экспериментов, по данным базы данных TID2008 вычисляются параметры этой модели. На основе предложенной модели осуществляется моделирование процесса проведения экспериментов для виртуальной базы изображений с заданным числом изображений, количеством участников экспериментов, распределением истинных значений визуального качества изображений. С помощью полученных результатов моделирования оценена эффективность методики верификации метрик визуального качества в целом, а также эффективность различных вариантов вычисления коэффициентов ранговой корреляции между значениями метрик и усредненными экспертными оценками.

Ключевые слова: база тестовых изображений, метрики визуального качества изображений, система визуального восприятия человека, анализ изображений.

Введение

Обеспечение заданного визуального качества изображения и его контроль являются актуальными для многих задач цифровой обработки изображений (ЦОИ), например, при сжатии изображений с потерями, подавлении шума, восстановлении, внедрении цифровых водяных знаков [1-5]. Для вычисления интегральных оценок визуального качества обработанных изображений и верификации методов ЦОИ широко используются метрики визуального качества при наличии эталона [6-12], при этом наиболее часто используемыми метриками, несмотря на их низкое соответствие восприятию человека, являются среднеквадратическая ошибка и пиковое соотношение сигнал/шум.

Метрики визуального качества изображений для обеспечения максимального соответствия вычисляемых оценок восприятию человека должны использовать математические модели, учитывающие особенности системы визуального восприятия человека (СВВЧ). На данный момент хорошо изучено лишь ограниченное количество особенностей СВВЧ [2, 3, 5], что сказывается на точности математических моделей и оценок визуального качества изображений, полученных метриками [3]. Ни одна из множества разработанных метрик пока не обеспечивает приемлемого соответствия СВВЧ [13].

Чтобы оценить эффективность той или иной метрики качества (степень ее соответствия СВВЧ), используются специальные базы тестовых изображений (БТИ), наиболее известными из которых яв-

ляются TID2008 [14], LIVE Database [15], Toyama [16], IVC [17] и CSIQ [18]. Такие БТИ включают набор из нескольких десятков эталонных изображений, эти же изображения с внесенными искажениями различных типов, а также массив значений усредненных экспертных оценок (Mean Opinion Score, MOS) визуального качества искаженных изображений, полученных по результатам психовизуальных экспериментов с добровольцами. В качестве оценки эффективности заданной метрики используется коэффициент корреляции (чаще всего для этой цели используются ранговые корреляции Спирмена и Кендалла [19]) между MOS и массивом значений данной метрики качества, вычисленных для каждого искаженного изображения базы.

Точность верификации метрик визуального качества с помощью БТИ зависит от методики формирования базы изображений и проведения экспериментов. При этом возникают и диалектические противоречия. В частности, чем больше разных изображений включено в базу данных и чем больше учтено различных типов искажений, тем более точными могут быть полученные MOS и тем больше недостатков метрик качества может быть выявлено в результате верификации. Однако увеличение количества тестовых изображений в базе данных приводит к увеличению времени проведения каждого эксперимента с добровольцами, что, вследствие накопления усталости, приводит к падению качества получаемых MOS. Рекомендуемым значением максимального времени проведения одного эксперимента является 30 минут [20]. За это время участник экспери-

мента не может оценить качество всех изображений для таких больших баз изображений, как TID2008 и LIVE, поэтому каждому участнику предлагается оценить качество лишь части тестовых изображений (соответствующих определенному эталонному изображению для TID2008 или соответствующих определенному типу искажений для LIVE). При этом недостаточно исследовано, как такое проведение экспериментов по частям влияет на точность получаемых MOS, каким образом лучше вычислять коэффициенты корреляции между MOS и значениями метрики для такой базы изображений (тоже по частям или же для всего массива MOS).

В данной работе осуществляется моделирование процесса проведения экспериментов в базе TID2008, на основе которого анализируются точностные характеристики получаемых результатов, и выбирается метод вычисления корреляции между MOS и значениями метрики, который обеспечивает наиболее объективные результаты верификации.

В подразделе 1 приводится краткое описание TID2008. Подраздел 2 посвящен оцениванию параметров модели принятия решений человеком при выполнении субъективных экспериментов. В подразделе 3 описывается построение математической модели проведения субъективных экспериментов, позволяющей оценить точность результатов в зависимости от количества тестов и характеристик БТИ. Подраздел 4 посвящен анализу влияния характеристик БТИ на точность MOS. В подразделе 5 проводится моделирование верификации метрик качества и анализ возможных ее вариантов.

1. База тестовых изображений TID2008

База TID2008 является наибольшей из существующих БТИ и позволяет проводить наиболее точную и полную верификацию метрик визуального качества изображений (см. данные Таблицы 1). Главными ее достоинствами являются: 1) количество учтенных типов искажений в несколько раз больше, чем для конкурирующих баз изображений; 2) большее количество проведенных субъективных экспериментов, что наряду с более совершенной методикой проведения экспериментов позволяет получить более точные MOS. Так относительная дисперсия MOS (характеризует "шумность" полученных усредненных мнений людей) для TID2008 является почти в 3 раза меньшей, чем для LIVE Database. Для других баз, из-за существенно меньшего числа проведенных экспериментов, этот показатель еще хуже, чем для LIVE Database. Неточность MOS, в свою очередь, уменьшает достоверность получаемых с помощью таких БТИ оценок эффективности метрик визуального качества.

Также на точность полученных MOS влияет выбранная методика оценивания визуального качества изображений во время экспериментов (по абсолютной или относительной шкале), проведения субъективных экспериментов (количество этапов сравнений, последовательность отображения изображений или их наборов и т.д.).

Таблица 1

Сравнительный анализ баз тестовых изображений

N	Характеристика	База тестовых изображений				
		LIVE Database	TID2008	Toyama	IVC	CSIQ
1	Число изображений в базе данных	779	1700	196	235	866
2	Число разных типов искажений, учтенных в базе данных	5	17	2	4	6
3	Количество участников экспериментов	161	846	16	15	35
4	Способ оценивания визуального качества	Абсолютная оценка качества по 5-бальной шкале (Excellent, Good, Fair, Poor, Bad)	Относительная оценка (выбор из двух искаженных изображений визуально менее отличающегося от оригинала)	Абсолютная оценка качества по 5-бальной шкале	Абсолютная оценка качества по 5-бальной шкале	Относительная оценка (изображения одного типа искажений упорядочивают на экране по визуальному качеству)
5	Количество оцениваний качества изображений в ходе экспериментов	25000	255816	3136	3525	5000
6	Шкала полученных оценок MOS (Mean Opinion Score, усредненная экспертная оценка)	0..100 (растянута из шкалы 1..5)	0..9	1..5	1..5	0..1
7	Дисперсия оценок MOS	250	0,63	1,5	-	-
8	Относительная дисперсия MOS	0,083	0,031	0,189	-	-

TID2008 содержит 25 эталонных изображений размером 512x384 пикселя. Каждому эталонному соответствует набор из 68 искаженных изображений (17 типов искажений по 4 уровня интенсивности [14]). Субъективные эксперименты проводились для каждого такого набора отдельно. На первом этапе все изображения случайным образом разбивались на пары. На монитор выводились эталонное (снизу) и пара искаженных изображений (см. рис. 1), из которых участник эксперимента выбирал изображение, визуально менее отличающееся от оригинала. Это изображение получало один балл, а второе - ноль баллов.



Рис. 1. Окно программы проведения субъективных экспериментов

Далее на 2-м и всех последующих этапах эксперимента (всего 9 этапов) формировались новые пары изображений. Причем, как в спортивных состязаниях, проводимых по швейцарской системе, в пары объединялись изображения, имеющие на данном этапе близкие по значению суммы баллов.

В результате 9-ти этапов каждое из искаженных изображений получало от 0 до 9 баллов. Усреднением этих баллов по всем участникам экспериментов (с предварительным исключением аномальных оценок) формировались MOS.

В работах [3, 14], посвященных исследованиям метрик качества с помощью TID2008, предлагается вычислять корреляцию между значениями метрик и MOS в целом для всех 1700 искаженных изображений базы или же для ее подмножеств, включающих только часть типов искажений. Однако, возможно, более объективные результаты могли бы быть получены при вычислении коэффициентов корреляции между MOS и значениями метрики отдельно для каждого из 25 эталонных изображений (ведь эксперименты проводились отдельно для каждого эталонного изображения) с последующим усреднением (или же какой-то иной операцией) этих коэффициентов.

Действительно, результаты, полученные для разных эталонных изображений, могут сильно отличаться по статистической однородности, типу распределения и другим характеристикам. Об этом косвенно свидетельствуют результаты некоторых исследований [21-22], в которых использовались эталонные изображения TID2008. В работе [21] в изображения внедрялись цифровые водяные знаки таким образом, чтобы искажения были визуально незаметными и малозаметными. После этого визуальное качество оценивалось известными метриками. Полученные результаты существенно зависели от особенностей эталонных изображений. Работа [22], посвященная анализу влияния сигнально-зависимых помех на визуальное качество, также показала большую зависимость результатов от характеристик изображений и их смыслового содержания.

Рассмотренные работы подтверждают, что характеристики эталонных изображений влияют на визуальное качество и, вследствие независимости экспериментов, потенциально могут приводить к неоднородности получаемых значений MOS. Поэтому необходимо определить величину этого влияния, выбрать оптимальную методику верификации метрик визуального качества.

Необходимо разработать математическую модель принятия решений участником экспериментов и оценить ее параметры.

Далее полученную модель можно использовать для моделирования процесса проведения экспериментов и оценить различные характеристики получаемых MOS в зависимости от исходных данных (количестве изображений в базе, их характеристик, параметров модели принятия решений).

Полученные таким образом данные можно будет использовать и для анализа различных вариантов вычисления коэффициентов корреляции между MOS и значениями верифицируемых метрик визуального качества.

2. Модель принятия решений

Будем считать, что участник эксперимента оценивает визуальное качество изображения с какой-то погрешностью. Для простоты положим, что эта погрешность является случайным числом, соответствующим нормальному закону распределения с нулевым математическим ожиданием и дисперсией $\sigma^2_{\text{МПР}}$. Кроме того, будем считать, что участник эксперимента может сделать полностью случайный выбор (случайное нажатие мышки на одно из изображений) с вероятностью $P_{\text{МПР}}$.

В соответствии с этой моделью значение субъективной оценки качества β , которую участник экс-

перимента приписывает изображению I, можно вычислить в соответствии с выражением:

$$\beta(I) = (1 - \delta)(\vartheta(I) + x) + \delta(\vartheta(I) + \eta), \quad (1)$$

где $\vartheta(I)$ – истинное значение визуального качества изображения; x – случайная величина (имеющая Гауссово распределение) с математическим ожиданием 0 и дисперсией $\sigma^2_{МПР}$; η – случайная величина (имеющая Гауссово распределение) с дисперсией $\sigma^2 = 100\sigma^2_{МПР}$, δ – случайная величина, принимающая значение 1 с вероятностью $P_{МПР}$ и 0 – в остальных случаях.

Чтобы оценить параметры $\sigma^2_{МПР}$ и $P_{МПР}$ модели вычисления $\beta(I)$, используем протоколы проведения экспериментов TID2008 (рис. 2). В этих протоколах содержится полная информация о прохождении эксперимента: номер набора, значения оценок качества, подробная информация обо всех сравнениях (номера сравниваемых изображений, итог выбора).

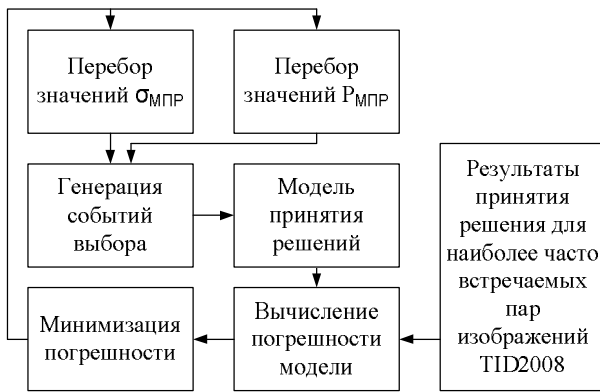


Рис. 2. Блок-схема получения параметров модели принятия решений

По протоколам экспериментов выберем 2500 пар изображений (по 100 для каждого эталонного изображения), для которых наиболее часто принимались решения. Для каждой i -й пары подсчитаем общее количество принятия решений N_i и количество ситуаций K_i , когда участник эксперимента выбрал первое изображение.

Ограничение в 100 пар обусловлено тем, что все они встречаются более 10 раз, что обеспечивает необходимую статистическую достоверность. Как показала практика, дальнейшее увеличение количества пар на итоговый результат не влияет.

Для каждой i -й пары из левого (I^l_i) и правого (I^r_i) изображений в соответствии с их значениями MOS (параметр ϑ из модели (1)) и заданными $\sigma_{МПР}$ и $P_{МПР}$ сгенерируем 5000 событий (выбрано как оп-

тимальное с точки зрения точности результатов и времени вычислений) принятия решений.

В соответствии с моделью (1) подсчитаем число ситуаций, в которых было выбрано первое изображение:

$$K'_i = \sum_{n=1}^{N_i} \tau_n, \quad (2)$$

$$\tau_n = \begin{cases} 1, & \beta(I^l_i) \geq \beta(I^r_i); \\ 0, & \beta(I^l_i) < \beta(I^r_i). \end{cases}$$

Оценим погрешность модели ξ в соответствии с выражением:

$$\xi = \frac{1}{2500} \sum_{i=1}^{2500} \left(\frac{K_i}{N_i} - \frac{K'_i}{5000} \right)^2. \quad (3)$$

Построим зависимость изменения погрешности модели ξ от параметра $\sigma_{МПР}$ при фиксированном значении $P_{МПР}=0,03$ (см. рис. 3). Значение 0,03 соответствует вероятности случайных нажатий (аномальных решений) для протоколов экспериментов базы TID2008.



Рис. 3. Зависимость погрешности модели от значения $\sigma_{МПР}$

Видно, что приведенная кривая имеет минимум в точке $\sigma_{МПР} = 0,75$.

Зафиксируем теперь $\sigma_{МПР} = 0,75$ и будем варьировать параметр $P_{МПР}$ (см. рис. 4).



Рис. 4. Влияние выбора $P_{МПР}$ на погрешность модели

Видно, что влияние параметра $R_{МПР}$ на итоговую погрешность несущественно, поэтому в работе будет использоваться ранее выбранное значение $R_{МПР} = 0,03$, которое хорошо соотносится с практическим опытом проведения экспериментов для баз TID2008 и LIVE.

Полученные параметры модели ($\sigma_{МПР} = 0,75$, $R_{МПР} = 0,03$) характеризуют точность принятия решений участниками экспериментов для TID2008. На практике при создании других баз данных значения этих параметров могут отличаться от полученных в данной работе, так как погрешность решений зависит от многих объективных и субъективных факторов. К объективным факторам можно отнести освещение помещения, характеристики мониторов, которые имеют разные размеры, настройки яркости, контраста, цветопередачи. Субъективные факторы определяют психофизическое состояние добровольцев при прохождении тестов, например, усталость, сосредоточенность, остроту зрения.

Полученное значение СКО принятия решения участником экспериментов $\sigma_{МПР} = 0,75$ (при диапазоне шкалы оценок 0...9) имеет большое самостоятельное значение. В частности, если пользователю предлагается оценивать качество изображений по шкале с диапазоном 0...100, то такая точность не будет оправданной, так как СКО принятия решения для такой шкалы составит $0,75/9 \times 100 = 8,33$, а разброс принятия решений в $\pm 2\sigma$ составит ± 16 единиц. Аналогично для популярной шкалы с диапазоном значений 1...5 разброс принятия решений в $\pm 2\sigma$ составит $\pm 0,66$ единицы. Такая шкала является вполне оправданной, так как погрешность принятия решения не превышает единицы шкалы измерений.

3. Модель проведения субъективных экспериментов

Следующий этап заключается в построении математической модели проведения виртуальных экспериментов на основе полученной модели принятия решений (рис.5).

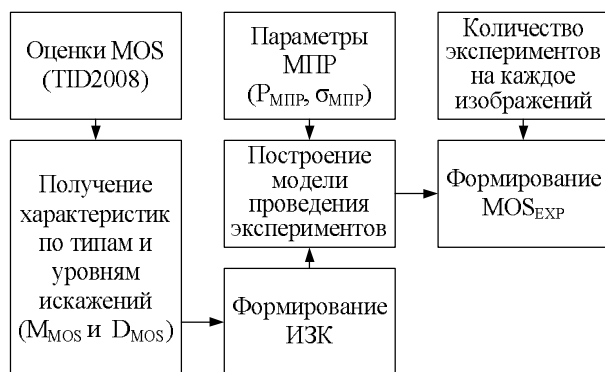


Рис. 5. Блок-схема построения модели проведения экспериментов

На первом шаге по БТИ TID2008 оцениваются математические ожидания M_{MOS} и дисперсии D_{MOS} для MOS каждого типа и уровня искажений.

Далее генерируется массив случайных чисел с математическими ожиданиями M_{MOS} и дисперсиями D_{MOS} для заданного количества изображений.

Этот массив будет использоваться в моделировании экспериментов как истинные значения визуального качества изображений ИЗК. От виртуальных экспериментов будем ожидать формирования усредненных экспертных оценок MOS_{EXP} как можно более близких к ИЗК.

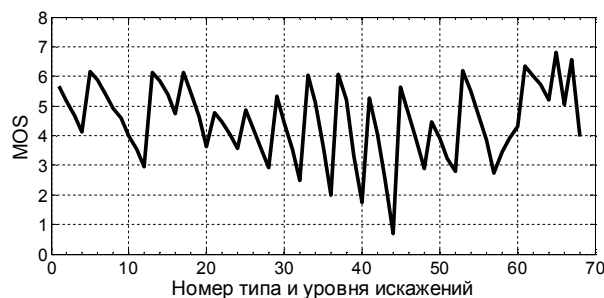


Рис. 6. Зависимость M_{MOS} от типа и уровня искажений

На рис. 6 приведены усредненные по эталонным изображениям значения M_{MOS} для типов и уровней искажений TID2008. Здесь нумерация изображений 1-4 соответствует четырем уровням первого типа искажений, 5-8 - четырем уровням второго типа искажений, и т.д. Номера 65-68 соответствуют уровням 1, 2, 3 и 4 для 17-го типа искажений. Отметим, что для большинства типов искажений с ростом уровня искажения M_{MOS} уменьшается.



Рис. 7. Зависимость D_{MOS} от типа и уровня искажений

На рис. 7 приведены значения D_{MOS} для типов и уровней искажений TID2008. Видно, что для некоторых типов искажений D_{MOS} является очень большим (MOS сильно зависит от содержания эталонного изображения).

Такие зависимости характерны для 11-го, 12-го и 14-го типов искажений (сжатие JPEG2000, ошибки в JPEG при передаче в каналах связи и искажения с

низкой степенью оригинальности), на которых примерно одинаковые по интенсивности искажения на изображениях с разными характеристиками сильно отличаются по визуальному качеству. Объясняется это тем, что искажения проявляются наиболее сильно на определенных участках изображений, причем расположение таких участков случайно. В зависимости от маскирующих свойств (появляется ли локальное искажение на однородном фоне или на текстах) визуальное качество разных искаженных изображений различно, что приводит к большой дисперсии MOS.

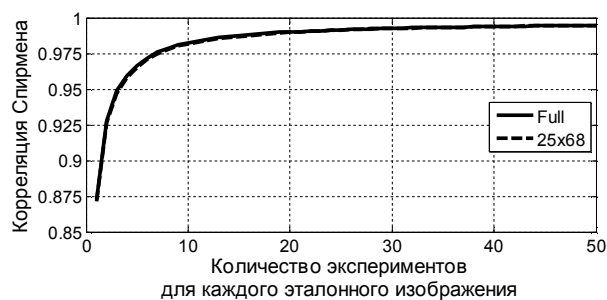
Следующим шагом является генерирование ИЗК. В данной работе будем генерировать ИЗК для 1700 и для 3000 изображений. Число 1700 соответствует количеству изображений в базе TID2008, а число 3000 - количеству изображений в разрабатываемой в данное время базе TID2013 [23].

Виртуальные эксперименты полностью повторяют процесс проведения экспериментов в TID2008 за исключением того, что решения принимаются не человеком, а моделью (1), причем в качестве параметра $\vartheta(I)$ выступает значение ИЗК для данного изображения. Параметрами проведения экспериментов являются число экспериментов K_{EXP} для каждого эталонного изображения, а также параметры модели принятия решений, которые так же могут варьироваться.

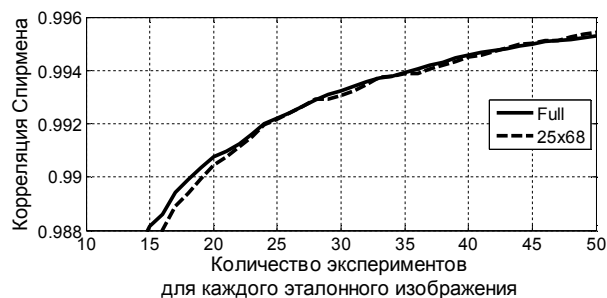
Варьирование количества проведенных экспериментов позволяет определить влияние этого параметра на точность оценок MOS_{EXP} . Зависимость коэффициента ранговой корреляции Спирмена между ИЗК и MOS_{EXP} от K_{EXP} для 1700 изображений приведена на рис. 8а и более подробно (фрагмент кривой) на рис. 8б. Аналогичный график для ранговой корреляции Кендалла показан на рис. 8в. Графики содержат кривые двух видов - для вычисления корреляции в целом по всей базе изображений (Full) и отдельно для каждого эталонного изображения с последующим усреднением (25x68).

Как видно из рис. 8а, приемлемая точность MOS_{EXP} в соответствии с корреляцией Спирмена достигается уже после проведения 20-30 экспериментов для каждого эталонного изображения. При дальнейшем увеличении K_{EXP} корреляция Спирмена растет незначительно.

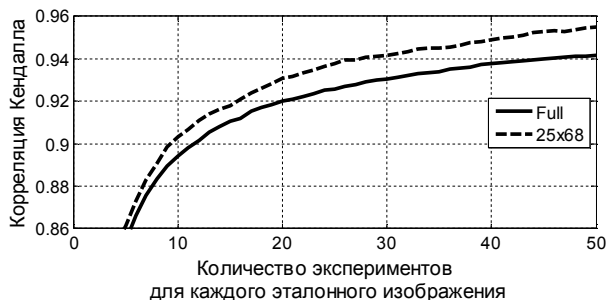
Результаты на рис. 8б свидетельствуют, что проведение 20 экспериментов позволяет достичь значения корреляции Спирмена на уровне 0,991, 30 экспериментов – 0,993, 50 – 0,995. На практике, для БТИ с большим количеством изображений сложно провести большее количество экспериментов.



а



б



в

Рис. 8. Соответствие между ИЗК и MOS_{EXP} : корреляция Спирмена (а) и увеличенный фрагмент кривой (б), корреляция Кендалла (в)

Например, для TID2008 за несколько лет было проведено 846 экспериментов, что для 25 эталонных изображений соответствует, в среднем, 34 экспериментам на каждое искаженное изображение.

При этом, как видно из рис. 8б, увеличение числа экспериментов (30 и более) увеличивает значение корреляции Спирмена на 0,003, но практически в два раза растет необходимое количество тестов. Поэтому в работе граничным параметром моделирования числа экспериментов выбрано значение $N = 30$.

Увеличенный фрагмент кривой на рис. 8б показывает, что, в соответствии со значением корреляции Спирмена, несущественно какой из вариантов ее вычисления используется. В то же время графики на рис. 8в показывают, что раздельное вычисление коэффициента корреляции Кендалла обеспечивает более высокие его значения. Одновременно с тем, насыщение кривой для корреляции Кендалла происходит позже - в районе 40-50 эксперимента.

На рис. 9 приведены аналогичные кривые для количества изображений в базе, равного 3000 (120 искаженных изображений для каждого эталонного).

Видно, что получаемые оценки точности сравнимы с приведенными на рис. 8 за исключением значений корреляции Кендалла, которые сравнивались для обоих случаев.

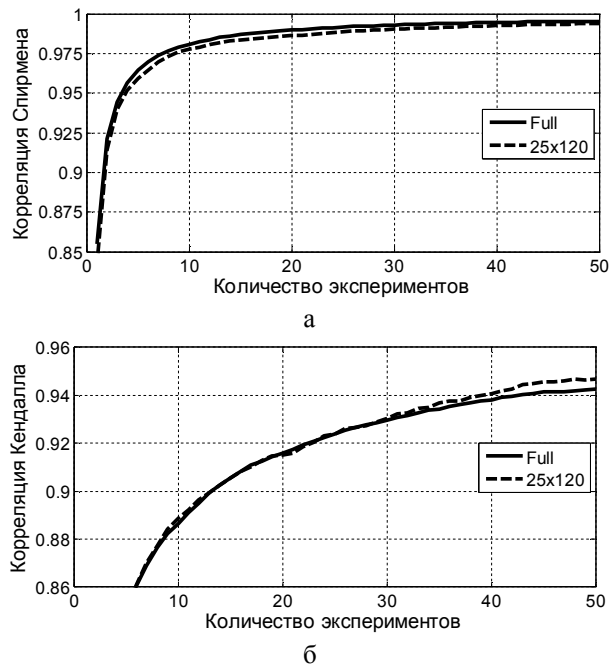


Рис. 9. Зависимость корреляций Спирмена (а) и Кендалла (б) между ИЗК и MOS_{EXP} для базы, включающей 3000 искаженных изображений и его оценкой от количества экспериментов

Таким образом, можно сделать вывод, что для больших баз изображений достаточным для получения достоверных MOS является проведение 30-50 экспериментов для каждого эталонного изображения.

4. Оценка эффективности методики проведения экспериментов по формированию MOS

Полученные в предыдущем разделе данные свидетельствуют, что точность результатов при проведении 30 субъективных экспериментов для TID2008 может достигать 0,993-0,994 по корреляции Спирмена. Но для других БТИ параметры модели принятия решений в зависимости от условий проведения экспериментов могут изменяться, поэтому необходимо проанализировать устойчивость получаемых MOS_{EXP} к изменениям этих значений.

Все графики и зависимости получены в предположении, что истинные оценки качества точно соответствуют параметрам M_{MOS} и D_{MOS} . Но эти характеристики являются оценками, полученными по результатам экспериментов. В зависимости от условий проведения, возраста, психофизического состояния добровольцев действительные значения

могут отличаться. Чтобы оценить влияние такой погрешности при создании ИЗК варьировалось их СКО в диапазоне 0,2...3 от $\sqrt{D_{MOS}}$ (рис. 10). Результаты показывают, что корреляция Спирмена (рис. 10а) при значениях СКО от 0,2 до $2\sqrt{D_{MOS}}$ не выходит за пределы диапазона $0,993 \pm 0,001$, при более высоких – начинает уменьшаться. Аналогичная ситуация получена и для корреляции Кендалла (рис. 10б).

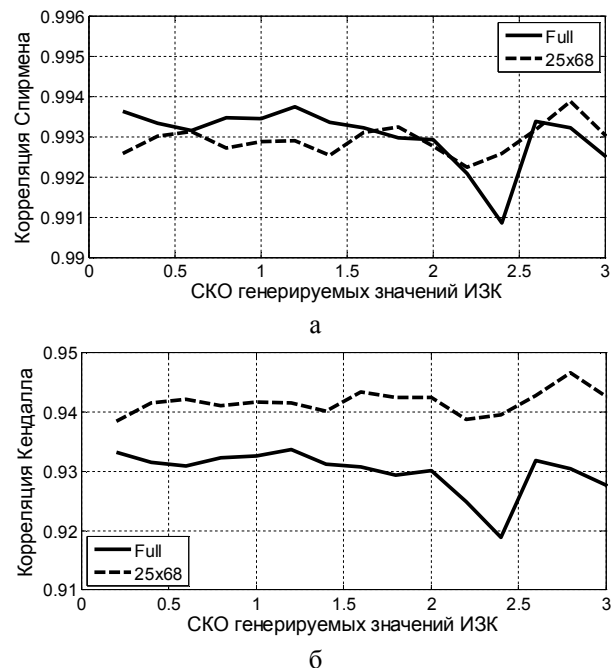


Рис. 10. Зависимость корреляции Спирмена (а) и Кендалла (б) от СКО истинных значений качества

Как видно из графиков на рис. 10, вариативность ИЗК не слишком влияет на точность получаемых MOS_{EXP} , особенно в соответствии с коэффициентом корреляции Спирмена, вычисленным отдельно.

В зависимости от условий прохождения, психофизического состояния участников экспериментов параметры полученной модели принятия решения могут тоже меняться. И если погрешность выбора практически не зависит от вероятности случайного нажатия, то СКО принятия решения существенно влияет на точность результатов, что подтверждается полученными графиками на рис. 11а и 11б. Значение $\sigma = 0,75$, полученное при обработке реальных данных, приводит к незначительному уменьшению корреляции Спирмена и Кендалла по сравнению с идеальным случаем (σ_{MIP} равно нулю). Максимум корреляции Спирмена для всех 1700 искаженных изображений находится на уровне 0,995, Кендалла – 0,942.

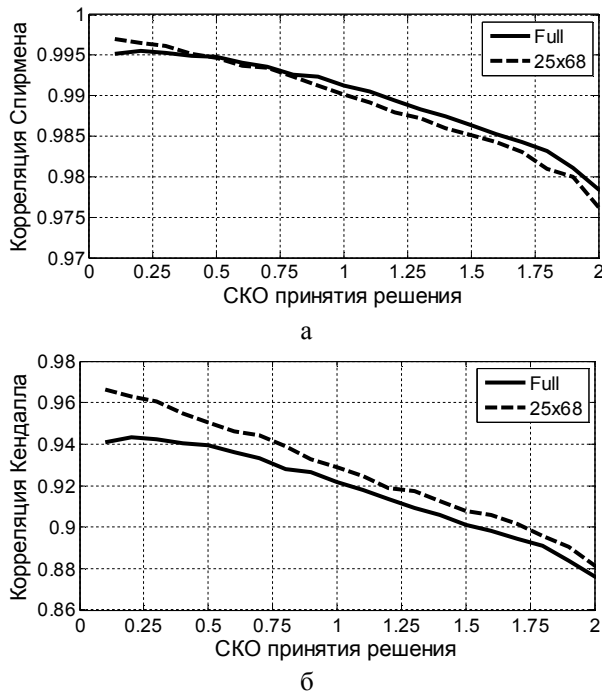


Рис. 11. Зависимость значений корреляции Спирмена (а) и Кендалла (б) от σ_{MIP}

Усреднение значений корреляций по 25 наборам обеспечивает более высокую точность (максимум по корреляции Спирмена равен 0,997, по Кендалла – 0,965). Это отличие (особенно при значениях σ_{MIP} , близких к нулю) показывает, что разделение всех искаженных изображений на 25 независимых наборов вносит определенную погрешность в методику подсчета значений оценок MOS. Но ряд ограничений и особенностей СВВЧ, например, конечное количество воспринимаемых яркостей и цветов, маскирующие эффекты, вносят более сильные погрешности и нивелируют методологические ошибки.

5. Оценка эффективности верификации метрик визуального качества изображений

Последний этап верификации БТИ – моделирование процесса верификации метрик визуального качества. Исследование заключается в получении значений метрик для всех тестовых изображений с последующим вычислением ранговых корреляций Спирмена и Кендалла. В основе метрик лежат разные подходы к оцениванию визуального качества, а итоговые значения сильно отличаются по статистическим характеристикам. Поэтому предполагается, что параметры модели неизвестны и оценки визуального качества, генерируемые метрикой визуального качества, искажены аддитивным нормальным шумом.

Зависимость значения корреляции между значениями метрики и ИЗК от СКО значений метрики

качества показана на рис. 12а (корреляция Спирмена) и на рис. 12б (корреляция Кендалла).

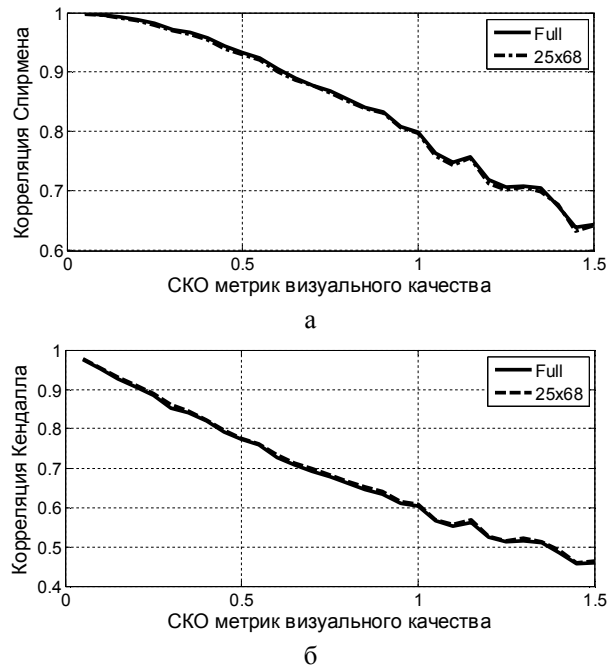


Рис. 12. Зависимость значений корреляции Спирмена (а) и Кендалла (б) от СКО значений метрик визуального качества

Разница между значениями корреляции для ИРК и MOS_{30} показана на рис. 13.

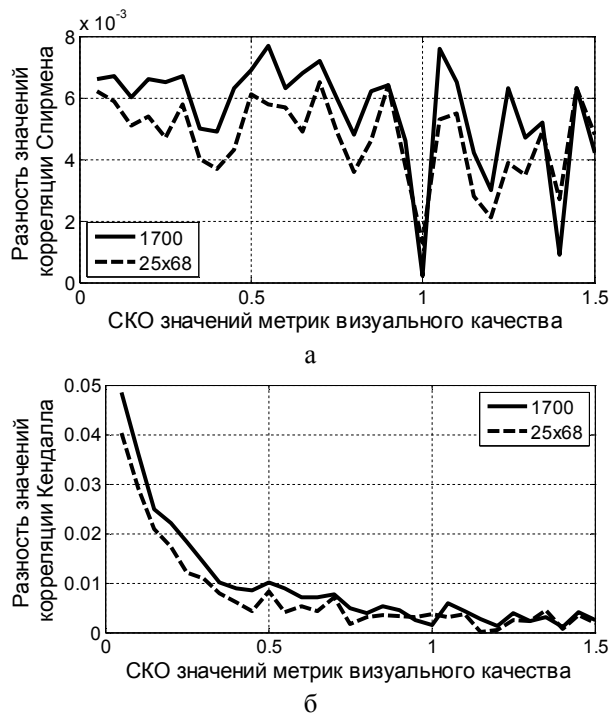


Рис. 13. Разница в верификации метрики с помощью ИЗК и MOS_{EXP} : по корреляции Спирмена (а), по корреляции Кендалла (б)

Чем меньше эта разница, тем более высокую точность верификации метрик визуального качества обеспечивает данная методика. Видно, что получаемые MOS_{EXP} обеспечивают точность верификации метрик качества, близкую к той, которая могла бы быть достигнута в идеальном случае (MOS_{EXP} совпадают с ИЗК). Максимальное отличие по значению коэффициента корреляции Спирмена не превышает 0,008, а по коэффициенту корреляции Кендалла - не превышает 0,05.

При этом раздельное вычисление коэффициентов корреляции (вариант 25x68) в большинстве ситуаций обеспечивает чуть меньшее отличие от верификации с помощью ИЗК, поэтому на практике следует предпочитать именно этот вариант вычисления коэффициентов корреляций.

Заключение

В работе рассмотрена задача верификации метрик визуального качества изображений с помощью специальных БТИ на примере TID2008 и проведена оценка точностных характеристик такой верификации. Основываясь на данных субъективных тестов, построена модель проведения экспериментов. Результаты показали, что оценки визуального качества зависят от характеристик эталонных изображений и методики проведения экспериментов. Также для получения корректных оценок MOS желательно обеспечить идентичные условия для всех участников, что проявилось при построении модели принятия решений участниками экспериментов. Для TID2008 влияние этих характеристик не превышает значения 0,005 и 0,03 по корреляции Спирмена и Кендалла между визуальным качеством изображений и его оценками по результатам тестов.

Основной фактор, определяющий точность оценок MOS – это количество проведенных экспериментов. Показано, что для повышения точности для каждого изображения должно быть получено более 30 оценок (в случае относительного оценивания). Для других БТИ с абсолютным способом оценивания это значение должно быть большим из-за меньшей шкалы и более высокой чувствительности зрения к относительному сравнению, нежели абсолютной оценке. Показано, что при раздельном проведении экспериментов для каждого эталонного изображения, коэффициенты корреляции целесообразно вычислять отдельно для каждого эталонного изображения с последующим их усреднением.

Литература

1. Keelan, B.W. *Handbook of Image Quality [Text]* / B.W. Keelan. - New York, USA, 2002. - 516 p.
2. Pratt, W.K. *Digital Image Processing [Text]* / W.K. Pratt. - Wiley-Interscience, 2007. - 782 p.

3. *Metrics performance comparison for color image database [Text]* / N. Ponomarenko, F. Battisti, K. Egiazarian, J. Astola, V. Lukin // *Proc. of VPQM. - USA, 2009. - 4 p.*

4. Гонсалес, Р. *Цифровая обработка изображений [Текст]* / Р. Гонсалес, Р. Вудс. - М.: Техносфера, 2005. - 1072 с.

5. Chandler, D.M. *Seven Challenges in Image Quality Assessment: Past, Present, and Future Research [Text]* / D.M. Chandler // *ISRN Signal Processing. - 2013. - P. 1-53.*

6. Wang, Z. *Multi-scale structural similarity for image quality assessment [Text]* / Z. Wang, E.P. Simoncelli, A.C. Bovik // *IEEE Asilomar Conf. on Signals, Systems and Computers. - Vol.2. - USA, 9-12 November 2003. - P. 1398-1402.*

7. *On between-coefficient contrast masking of DCT basis functions [Електронний ресурс]* / N. Ponomarenko, F. Silvestri, K. Egiazarian, M. Carli, J. Astola, V. Lukin - 80 Min / 700 MB. / *Proc. of the Third Int. Workshop on Video Proc. and Quality Metrics. - USA, 2007. - 4 p. 1 електрон. опт. диск (CD-ROM), 12 см.*

8. Chandler, D.M. *VSNR: A Wavelet-Based Visual Signal-to-Noise Ratio for Natural Images [Text]* / D.M. Chandler, S.S. Hemami // *IEEE Transactions on Image Proc. - 2007. - Vol. 16 (9). - P. 2284-2298.*

9. Wang, Z. *A universal image quality index [Text]* / Z. Wang, A. Bovik // *IEEE Signal Processing Letters. - 2000. - Vol. 9. - P. 81-84.*

10. Zhang, L. *FSIM: a feature similarity index for image quality assessment [Text]* / L. Zhang, X. Zhang, D. Zhang // *IEEE Transactions on Image Processing. - 2011. - Vol. 20, No 5. - P. 2378-2386.*

11. *Modified image visual quality metrics for contrast change and mean shift accounting [Text]* / N. Ponomarenko, O. Ereemeev, V. Lukin, K. Egiazarian and M. Carli // *Proceedings of CADSM. - Ukraine, 23-25 February 2011. - P. 305-311.*

12. Jin, L. *Perceptual Image Quality Assessment Using Block-Based Multi-Metric Fusion (BMMF) [Text]* / L. Jin, K. Egiazarian, C-C. Jay Kuo // *Proceedings of ICASSP. - Japan, 25-30 March 2012. - P. 1145-1148.*

13. Пономаренко, Н.Н. *Метрика визуального качества изображений для корректного учета искажений яркости и контраста [Текст]* / Н.Н. Пономаренко, О.И. Еремеев, В.В. Лукин // *Радиоелектронні і комп'ютерні системи. - 2009. - № 1(35). - С. 36-41.*

14. TID2008 – *A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics [Text]* / N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti // *Успехи современной радиоэлектроники. - 2009. - № 10. - С. 30-45.*

15. *LIVE Image Quality Assessment Database Release 2 [Электронный ресурс]* / H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik. - Режим доступа: <http://live.ece.utexas.edu/research/quality/subjective.htm>. - 23.12.2012.

16. Horita, Y. *Subjective quality assessment to*

yama database [Электронный ресурс] / Y. Horita, K. Shibata, Z.M. Parvez Saddam. – Режим доступа: <http://mict.eng.u-toyama.ac.jp/mict/>. - 15.01.2013.

17. Le Callet, P. Subjective quality assessment IRCCyN/IVC database [Электронный ресурс] / P. Le Callet, F. Atrousseau. – Режим доступа: <http://www.irccyn.ec-nantes.fr/ivcdb/>. - 17.01.2013.

18. Larson, E. C. Most apparent distortion: full-reference image quality assessment and the role of strategy [Text] / E. C. Larson, D. M. Chandler // *Journal of Electronic Imaging*. – 2010. – 19 (1). – P. 011006-1 – 011006-21.

19. Kendall, M.G. *The advanced theory of statistics. Volume 1* [Text] / M.G. Kendall. - Charley Griffin & company united, London, 1945. – 457 p.

20. ITU (2002). *Methodology for Subjective Assessment of the Quality of Television Pictures Recom-*

mendation BT.500-11 [Text]. - International Telecommunication Union, Geneva, Switzerland, 2002. – 48 p.

21. Оценка визуальных искажений при внедрении в изображения цифровых водяных знаков [Текст] / О.И. Еремеев, Н.Н. Пономаренко, В.В. Лукин, А.А. Зеленский // *Материали конференції ДУІКТ*. – Харьков, 25-26 ноября 2010. - С. 20-24.

22. Визуальное качество изображений при различных типах помех [Текст] / О.И. Еремеев, Д.В. Февралев, Н.Н. Пономаренко, В.В. Лукин // *Радиоелектронні і комп'ютерні системи*. – 2012. - № 2(54). – С. 49-57.

23. *Color image database TID2013: peculiarities and preliminary results* [Text] / N. Ponomarenko, O. Ieremeiev, V. Lukin, K. Egiazarian and others // *Proceedings of EUVIP*. – Paris, 2013. – 7 p.

Поступила в редакцію 12.04.2013, рассмотрена на редколлегии 15.05.2013

Рецензент: д-р техн. наук, проф., проф. кафедры «Проектирования радиоэлектронных систем летательных аппаратов» Э.Н. Хомяков, Национальный аэрокосмический университет им. Н.Е. Жуковского "ХАИ", г. Харьков.

ОЦІНКА ТОЧІСНИХ ХАРАКТЕРИСТИК МЕТОДИКИ ВЕРИФІКАЦІЇ МЕТРИК ВІЗУАЛЬНОЇ ЯКОСТІ НА ОСНОВІ СПЕЦІАЛІЗОВАНИХ БАЗ ЗОБРАЖЕНЬ

О.І. Єремеев, М.М. Пономаренко, В.В. Лукин

Розглянуто задачу верифікації метрик візуальної якості зображень на основі спеціалізованих баз даних. Запропоновано математичну модель прийняття рішень учасником експериментів. Здійснено моделювання процесу проведення експериментів для віртуальної бази зображень. Оцінено ефективність методики верифікації метрик візуальної якості в цілому, а також ефективність різних варіантів обчислення коефіцієнтів рангової кореляції між значеннями метрик і усередненими експертними оцінками.

Ключові слова: бази тестових зображень, метрики візуальної якості, система візуального сприйняття людини, аналіз зображень.

ESTIMATION OF ACCURACY OF EVALUATION OF IMAGE VISUAL QUALITY METRICS BY USING OF TEST IMAGES DATABASES

O.I. Ieremeiev, N.N. Ponomarenko, V.V. Lukin

The problem of image visual quality metrics evaluation based on test images databases is considered. A mathematical model of decision-making by experiments participants is proposed. Simulations of experiments for a virtual data bases are carried out. Effectiveness of the method of quality metrics evaluation using test image databases is estimated as well as effectiveness of different ways of calculating correlation coefficients between MOS and metrics values.

Keywords: test image databases, visual quality metrics, human visual system, image analysis.

Єремеев Олег Игоревич – мл. науч. сотр. кафедры «Приема, передачи и обработки сигналов», Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: ol.eremeiev@mail.ru.

Пonomаренко Николай Николаевич – д-р техн. наук, доц., доц. кафедры «Приема, передачи и обработки сигналов», Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: nikolay@ponomarenko.info.

Лукин Владимир Васильевич – д-р техн. наук, проф., проф. кафедры «Приема, передачи и обработки сигналов», Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Харьков, Украина, e-mail: lukin@xai.kharkov.ua