

УДК 681.325

О.И. ЕРЕМЕЕВ, Н.Н. ПОНОМАРЕНКО

*Национальный аэрокосмический университет им. Н.Е. Жуковского «ХАИ», Украина***БАЗА ДАННЫХ ДЛЯ ВЕРИФИКАЦИИ МЕТРИК ВИЗУАЛЬНОГО КАЧЕСТВА
ИЗОБРАЖЕНИЙ ПРИ ОТСУТСТВИИ ЭТАЛОНА**

Рассмотрена задача верификации метрик визуального качества изображений при отсутствии эталона. Сформирована база из 500 изображений в формате JPEG, выбранных из Интернета так, чтобы включать все основные типы искажений, содержать изображения, разные по своим характеристикам, и имеющие разное визуальное качество. Описаны проведенные эксперименты, в ходе которых более 300 добровольцев оценивали визуальное качество этих изображений. По результатам экспериментов сформированы усредненные экспертные оценки качества изображений. С помощью сформированной базы изображений и полученных экспертных оценок проанализировано соответствие восприятию человека некоторых простых метрик качества изображений. Предложенная база данных выложена в Интернете для свободного скачивания.

Ключевые слова: базы тестовых изображений, метрики визуального качества изображений.

Введение

Постановка проблемы. В области цифровой обработки изображений для оценки эффективности методов фильтрации и подавления шумов, сжатия изображений с потерями, и многих других, широкое распространение получили метрики визуального качества изображений. Ряд метрик, для повышения точности оценивания визуального качества, учитывают особенности системы визуального восприятия человека (СВВЧ). Сложность заключается в том, что система зрения человека изучена фрагментарно, и в метриках качества учитываются только некоторые особенности СВВЧ. На данный момент не существует метрик качества, которые точно ей бы соответствовали. Это приводит к необходимости разработки специальных баз тестовых изображений, позволяющих оценить степень соответствия между той или иной метрикой качества изображений и СВВЧ.

Анализ литературы. Большое число разработанных в последнее время метрик визуального качества изображений обуславливает актуальность создания эффективных средств для их верификации.

Большинство таких метрик при анализе изображений учитывают те или иные особенности СВВЧ. Однако не существует строгих математических моделей СВВЧ, что делает невозможной строгую математическую оценку эффективности метрик качества. Поэтому разрабатываются специальные базы тестовых изображений, для которых формируют усредненные экспертные оценки (Mean Opinion Score, MOS), с помощью которых можно верифицировать заданную метрику качества корреляционными методами. Примерами таких баз изображений

являются TID2008 [1] и LIVE Database [2]. Однако их применение ограничивается верификацией метрик, в которых оценивается степень визуального различия между заданными эталонным и искаженным изображениями.

В то же время существует ряд задач, когда эталонное изображение отсутствует, и метрика должна оценивать качество при наличии только одного изображения (искаженного). Примерами областей использования таких метрик является: слепая оценка параметров шума [3]; оценка эффекта размытия и контраста [4 – 7]; слепая оценка параметров сжатия jpeg изображений [8, 9]; оценка искажений и артефактов на изображении [10, 11]. Для объективной оценки эффективности подобных метрик и их сравнения необходима соответствующая база тестовых изображений и оценок MOS.

Цель исследования. Разработать базу тестовых изображений для метрик качества без эталона и сформировать MOS. На основании полученных MOS оценить степень соответствия СВВЧ некоторых простых метрик качества.

1. Описание базы тестовых изображений и субъективных экспериментов

На данный момент метрики визуального качества изображений учитывают только некоторые основные особенности восприятия человека. В то же время, чувствительность человеческого глаза зависит от многих характеристик изображения и присутствующих на нем искажений, наиболее значимыми из которых являются: искажения яркости; искажения контраста, нечеткость изображения, аддитивный

и импульсный шум, блочные артефакты вследствие сжатия с потерями, насыщенность изображения деталями и текстурами, равномерность распределения искажений и шума на изображении. Поэтому к разрабатываемой базе тестовых изображений можно предъявить следующие требования:

1) база должна включать изображения с самыми разными характеристиками: процентным содержанием однородных участков, деталей, текстур, характеристиками текстур и т.д.;

2) в базе должны содержаться, по возможности, все основные типы искажений, типичные для реальных ситуаций: размытие, шум при ночной съемке, искажения яркости и др.

В соответствии с этими требованиями в Интернете было отобрано 500 реальных изображений в формате JPEG, которые соответствуют реальным ситуациям, возникающим при фотосъемке, например изображения с различным коэффициентом сжатия, ночная съемка, макросъемка, а так же компьютерная графика (несколько изображений). При выборе изображений для тестового набора их визуальное качество предварительно оценивалось экспертом по пятибалльной шкале. Для хорошей репрезентативности базы было отобрано 50 изображений «очень плохого» качества, 75 «плохого», 100 «среднего», 125 «хорошего» и 150 «отличного» качества.

Для удобства размещения пар изображений на экране в базу отбирались только изображения с размерами, не превышающими 500x600 пикселей. При этом минимальный размер изображений составил 300x400 пикселей, то есть площадь изображений в базе может отличаться в 2,5 раза. Это дает возможность эффективно тестировать метрики качества, учитывающие отличия в размере изображений.

Для получения MOS участникам экспериментов предъявляются пары изображений, в которых каждый раз нужно выбирать лучшее по визуальному качеству. Это изображение получает 1 балл, а худшее по качеству - 0 баллов. Каждое изображение в ходе одного эксперимента предъявляется для оценивания 11 раз и, таким образом, может получить от 0 до 11 баллов, что и определяет диапазон возможных значений формируемых оценок визуального качества. Полученные оценки в дальнейшем усредняются по разным участникам экспериментов (от 30 до 50 для разных изображений) и формируются MOS. Всего в экспериментах приняло участие более 300 добровольцев.

Проверка полученных экспериментальных данных осуществлялась в соответствии с рекомендациями стандарта на проведение подобных экспериментов [12]. Согласно этому стандарту полученная оценка считается аномальной, если ее значение отличается от средней оценки больше, чем на $2,33\sigma$,

где σ^2 - дисперсия значений этой оценки по всем экспериментам. Если же в одном эксперименте таких аномальных оценок больше 10%, то весь эксперимент считается неверным. В соответствии с этими требованиями было отбраковано 14 результатов экспериментов, а общий процент отброшенных аномальных оценок составил 6,3 %.

Распределение значений MOS по изображениям показано на рис. 1, а на рис.2 показана гистограмма этих значений.

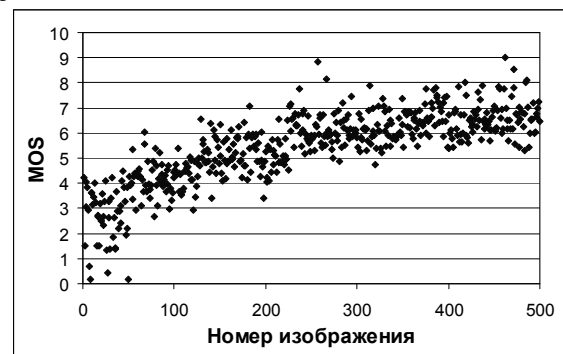


Рис. 1. Значения MOS для всех изображений

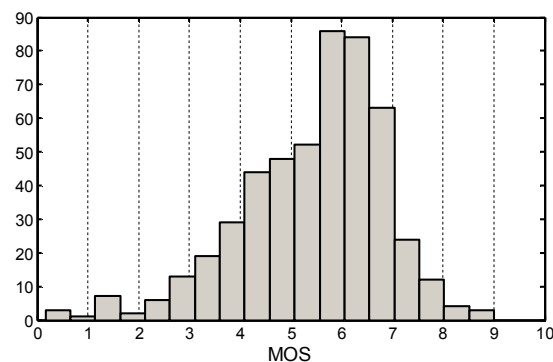


Рис. 2. Гистограмма значений MOS

Видно (рис. 1), что в целом наблюдается высокая корреляция между предварительными оценками качества изображений, сделанными экспертом при отборе изображений для формируемой базы данных, и итоговыми MOS (меньшие номера изображений в базе соответствуют меньшим предварительным оценкам качества изображений). Однако видно, что многие предварительные оценки сильно отличаются от полученных MOS, что еще раз косвенно подтверждает необходимость накопления и усреднения большого количества экспертных оценок для получения достоверных MOS.

Усредненное значение MOS и дисперсии оценок MOS по всем изображениям составили соответственно 5,46 и 1,17, а значение относительной дисперсии - 0,039. Невысокое значение относительной дисперсии полученных MOS (0,0391) свидетельствует об их высокой достоверности и пригодности для использования в процессе верификации метрик качества.

2. Применение базы тестовых изображений для верификации метрик качества

Разработанная база, благодаря разнообразию тестовых изображений, позволяет оценивать метрики качества, учитывающие разные факторы. На рис. 3 приведена структурная схема использования базы для оценки эффективности метрик качества. Как видно из рисунка эффективность метрики качества определяется ранговой корреляцией между значениями MOS и значениями метрики. В качестве корреляции могут использоваться, например, корреляции Спирмена и Кендалла [13].

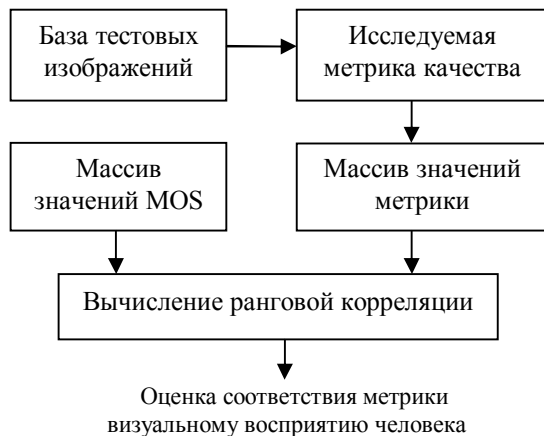


Рис. 3. Структурная схема процесса верификации метрики визуального качества изображений

В данной работе была осуществлена оценка эффективности следующих метрик.

Метрика, определяющая визуальное качество изображения по размеру сжатого jpeg изображения. Для данной простой метрики выполнялось вычисление ранговой корреляции между значениями MOS и размерами тестовых jpeg изображений. Полученные значения ранговой корреляции Спирмена и Кендалла равны соответственно 0,568 и 0,405.

Метрика, определяющая качество изображения по суммарной энергии его цветовых компонент в цветовом пространстве RGB. В скользящем окне 8x8 пикселей для каждого положения окна вычисляется локальная дисперсия:

$$\sigma_{\text{loc}}^2 = \frac{1}{64} \sum_{i=1}^8 \sum_{j=1}^8 (I_{ij} - \bar{I})^2, \quad (1)$$

где I_{ij} – значение пикселя с индексом i, j в текущем окне, \bar{I} – среднее значение пикселя в окне.

Далее значения локальных дисперсий по всему изображению усредняются и суммируются для всех трех цветовых компонент. Полученные значения для корреляции Спирмена равны 0,459, а для корреляции Кендалла – 0,315.

Метрика, учитывающая распределение энергии по пространственным частотам. Данная метрика вычисляет соотношение высоких и низких пространственных частот, что учитывает такие искажения, как размытие или завышение контраста. Метрика вычисляет энергию отдельно для высоких и низких частот в соответствии с выражением:

$$S1 = \frac{1}{N} \sum_{k \in U} \sum_{i=0}^7 \sum_{j=0}^7 \delta(i, j) D_k(i, j)^2,$$

$$S2 = \frac{1}{N} \sum_{k \in U} \sum_{i=0}^7 \sum_{j=0}^7 \lambda(i, j) D_k(i, j)^2, \quad (2)$$

$$\delta(i, j) = \begin{cases} 0, & (i + j) = 0; \\ 1, & 0 < (i + j) < 3; \\ 0, & (i + j) > 2; \end{cases} \quad \lambda(i, j) = \begin{cases} 0, & 0 < (i + j) < 3; \\ 1, & (i + j) > 2; \end{cases}$$

где $S1$ – энергия низких частот, $S2$ – энергия высоких частот, N – число проанализированных блоков 8x8 пикселей изображения, D_k – коэффициенты дискретного косинусного преобразования k -го блока, U – анализируемая область (в данной работе – все изображение).

Значения корреляции для предложенной метрики составили 0,449 (корреляция Спирмена) и 0,310 (корреляция Кендалла).

Как видно из полученных результатов, все метрики качества плохо соответствуют восприятию человека. Полученные результаты показывают, что данная база изображений может эффективно использоваться для выявления недостатков и совершенствования существующих метрик.

Предложенная база изображений выложена в Интернете [14] для свободного доступа.

Заключение

В данной работе описана новая база тестовых изображений, позволяющая оценить эффективность метрик визуального качества без наличия эталона. Полученные MOS использованы для верификации некоторых простых метрик, не учитывающих особенности СВВЧ. Низкие значения корреляции значений этих метрик и MOS свидетельствуют, во-первых, о плохом соответствии метрик СВВЧ, и, во-вторых, о пригодности разработанной базы для выявления недостатков тестируемых метрик.

Литература

1. TID2008 – A Database for Evaluation of Full-Reference Visual Quality Assessment Metrics / N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli, F. Battisti // Успехи современной радиоэлектроники. – 2009. – № 10. – С. 30-45.
2. Sheikh H.R. A Statistical Evaluation of Recent Full Reference Image Quality Assessment Algo-

rithms / H.R. Sheikh, M.F. Sabir, F.C. Bovik // IEEE Trans. on Image Proc. – 2006. – Vol. 15, no. 11. – P. 3441-3452.

3. Еремеев О.И. Мера информативности участков изображений / О.И. Еремеев, Н.Н. Пономаренко // *Авиационно-космическая техника и технология*. – 2008. – № 6 (53). – С. 105-108.

4. Marziliano P. A no-reference perceptual blur metric / P. Marziliano, F. Dufaux, S. Winkler // IEEE 2002 International Conference on Image Processing. – 2002. – Vol. 3. – P. 57-60.

5. Marziliano P. Perceptual Blur and Ringing Metrics: Application to JPEG2000 / P. Marziliano, F. Dufaux, S. Winkler // *Signal Processing: Image Communication*. – February. – 2004. – Vol. 19, №. 2. – P. 163-172.

6. Caviedes J.E. No-reference sharpness metric based on local edge kurtosis / J. E. Caviedes, S. Gurbuz // *Proc. of ICIP2002*. – Sept., 2002. – Vol. 3. – P. 53-56.

7. Ferzli R. A no-reference objective image sharpness metric based on the notion of just noticeable blur (JNB) / R. Ferzli, L.J. Karam // *IEEE Trans. on Image Proc.* – April, 2009. – Vol. 18, Issue 4. – P. 717-728.

8. Horita Y. No-reference image quality assessment for JPEG/JPEG2000 coding / Yuukou Horita,

Shinji Arata, Tadakuni Murai // *Proc. of EUSIPCO-2004*. – September 2004. – Vol. 2. – P.1301-1304.

9. Hontsch I. Adaptive image coding with perceptual distortion control / I. Hontsch, L.J. Karam // *IEEE transactions on image processing*. – March, 2002. – Vol.11, № 3. – P. 213-222.

10. Farias M.C.Q. No-reference video quality metric based on artifact measurements / M.C.Q. Farias, S.K.Mitra // *IEEE International Conference on Image Processing*. – 2005. – Vol. 3. – P. 141-144.

11. Bovik A. DCT-domain blind measurement of blocking artifacts in DCT-coded images / A. C. Bovik, S. Liu // *Proceedings of the Acoustics, Speech, and Signal Proc.* – May, 2001. – Vol. 3 – P. 1725-1728.

12. *Methodology for Subjective Assessment of the Quality of Television Pictures Recommendation BT.500-1*. – Geneva: ITU, 2002. – 48 p.

13. Kendall M.G. *The advanced theory of statistics. Vol. 1* / M.G. Kendall – London: Charles Griffin & Company limited, 1945. – 457 p.

14. База даних для верифікації метрик візуального качества изображений при отсутствиі эталона [Электронный ресурс]. – 2010. – Режим доступа: <http://ponomarenko.info/nrtid.htm>.

Поступила в редакцию 15.02.2010

Рецензент: д-р техн. наук, проф., проф. кафедри проектування радіоелектронних систем летальних апаратів В.К. Волосюк, Національний аерокосмічний університет ім. Н.Е. Жуковського "ХАІ", Харків.

БАЗА ДАНИХ ДЛЯ ВЕРИФІКАЦІЇ МЕТРИК ВІЗУАЛЬНОЇ ЯКОСТІ ЗОБРАЖЕНЬ ЗА ВІДСУТНІСТЮ ЕТАЛОНА

О.І. Еремеев, М.М. Пономаренко

Розглянуто задачу верифікації метрик візуальної якості зображень за відсутністю еталона. Сформовано базу з 500 зображень у форматі JPEG, вибраних з Інтернету так, щоб включати всі основні типи спотворень, містити зображення, різні за своїми характеристиками та візуальною якістю. Описано проведені експерименти, в ході яких понад 300 добровольців оцінювали візуальну якість цих зображень. За результатами експериментів сформовані усереднені експертні оцінки якості зображень. За допомогою сформованої бази зображень та отриманих експертних оцінок проаналізовано відповідність сприйняттю людини деяких простих метрик якості зображень. Запропоновану базу даних викладено в Інтернеті для вільного доступу.

Ключові слова: бази тестових зображень, метрики візуальної якості зображень.

TEST IMAGE DATABASE FOR VERIFICATION OF NON-REFERENCE IMAGE VISUAL QUALITY METRICS

O.I. Eremeev, N.N. Ponomarenko

The task of verification of non-reference image quality metrics is considered. A test image database including 500 JPEG images is proposed. The images are selected to provide the presence of images with different characteristics and visual quality. More than 300 observers took part in experiments by obtaining of Mean Opinion Scores. Correspondence of some simple quality metrics to human perception is estimated using the proposed database. The database is available for downloading for Internet now.

Key words: test image databases, image visual quality metrics.

Еремеев Олег Игоревич - аспірант кафедри приєма, передачі і обробки сигналів, Національний аерокосмічний університет ім. Н.Е. Жуковського «ХАІ», Харків, Україна.

Пономаренко Николай Николаевич - канд. техн. наук, докторант кафедри приєма, передачі і обробки сигналів, Національний аерокосмічний університет ім. Н.Е. Жуковського «ХАІ», Харків, Україна, e-mail: nikolay@ponomarenko.info.