

ISSN 1816-0301 (print)  
УДК 004.932

Поступила в редакцию 12.06.2018  
Received 12.06.2018

**В. В. Старовойтов**

*Объединенный институт проблем информатики  
Национальной академии наук Беларуси, Минск, Беларусь*

## УТОЧНЕНИЕ ИНДЕКСА *SSIM* СТРУКТУРНОГО СХОДСТВА ИЗОБРАЖЕНИЙ

**Аннотация.** Исследуются свойства популярной меры сравнения цифрового изображения с эталоном – индекса структурного сходства, называемого в литературе *SSIM*. Доказывается, что *SSIM* и производные от него функции не являются метриками. Описываются варианты модификации индекса *SSIM*. Показывается, что меры, подобные этому индексу, оценивают не качество изображений, а их пофрагментное сходство. Кроме того, отмечается, что усредненные оценки, выставяемые экспертами и называемые *MOS*, очень субъективны и не могут в точности коррелировать с вычисляемыми количественными оценками сходства сравниваемых изображений. Для подсчета индекса *SSIM* вычисляется матрица локальных оценок. Каждая оценка определяет сходство двух анализируемых пикселей с одинаковыми координатами с учетом соседних пикселей. Обычно в качестве индекса *SSIM* используется средняя арифметическая величина полученной матрицы. Вместо нее для усовершенствования индекса *SSIM* предлагается использовать параметр масштаба распределения Вейбулла, аппроксимирующего гистограмму локальных оценок индекса *SSIM*. На множестве изображений общедоступной базы TID2013 показывается, что предложенный параметр является более точной характеристикой сходства изображений, чем среднее множество локальных оценок.

**Ключевые слова:** цифровое изображение, мера сходства изображений, индекс *SSIM*, распределение Вейбулла  
**Для цитирования.** Старовойтов, В. В. Уточнение индекса *SSIM* структурного сходства изображений / В. В. Старовойтов // Информатика. – 2018. – Т. 15, № 3. – С. 41–55.

**V. V. Starovoitov**

*The United Institute of Informatics Problems of the National Academy of Sciences of Belarus, Minsk, Belarus*

## ENHANCEMENT OF THE STRUCTURAL SIMILARITY INDEX *SSIM*

**Abstract.** Properties of a popular measure of comparing a digital image with a reference – the index of structural similarity, called *SSIM* in the literature – are explored. It is proved that the *SSIM* and its derivative functions are not metrics. Variants of the index modification are described. It is shown that measures similar to this index evaluate not quality of images, but their similarity by fragments. Additionally, it is shown that the averaged expert assessments called *MOS* are very subjective and cannot exact correlate with numerical estimates of similarity of the compared images. To get the *SSIM* index, a matrix of local estimates is calculated. Each evaluation determines similarity of two analyzed pixels with the same coordinates taking into account neighboring pixels. Usually, the average arithmetic value of the obtained matrix is used as the *SSIM* index. Instead, to improve the *SSIM* index, it is proposed to use the scale parameter of the Weibull distribution, which approximates the histogram of the local index estimates. On a set of images from the public database TID2013, it is shown that the proposed parameter is a more accurate characteristic of image similarity than the mean of local estimates.

**Keywords:** digital image, image similarity measure, index *SSIM*, Weibull distribution

**For citation.** Starovoitov V. V. Enhancement of the structural similarity index *SSIM*. *Informatics*, 2018, vol. 15, no. 3, pp. 41–55 (in Russian).

**Введение.** В настоящее время подавляющее большинство глобальных оценок качества цифровых изображений вычисляется как среднее арифметическое множества локальных оценок. Вычисление индекса структурного сходства *SSIM* (от англ. Structural SIMilarity index) служит одним из популярных и широко цитируемых методов измерения сходства между двумя изображениями [1]. Индекс *SSIM* количественно описывает совпадение двух сравниваемых

изображений, хотя в англоязычной литературе его часто называют метрикой оценки качества изображений, имея в виду сравнение изображения с эталоном. Под эталоном подразумевается одно из сравниваемых изображений. Следует отметить, что данная функция метрикой не является, поскольку она не удовлетворяет метрическим аксиомам. Качество изображений этой функцией также не оценивается, поскольку сравнение низкокачественного изображения с самим собой дает максимальное значение индекса  $SSIM = 1$ , что должно свидетельствовать о его высоком качестве.

Отличительным признаком описанного подхода является то, что он учитывает структурные различия в паре изображений и наличие сильной корреляции в цвете соседних пикселей. Вычисление сходства выполняется в небольших окнах. Такие оценки сходства назовем локальными и обозначим  $ssim$ . Каждая из них оценивает яркость, контраст и структуру двух изображений в окне. На заключительном этапе в качестве глобального значения индекса  $SSIM$  берется среднее арифметическое множества локальных оценок  $ssim$ .

Во всех исследованиях индекса  $SSIM$  выполнялись сравнения изображений, представляющих одинаковые сцены с небольшими яркостными искажениями. Никто и никогда не применял этот индекс для сравнения изображений разных сцен или изображений, искаженных геометрически.

Для подобных сравнительных экспериментов используются специально созданные базы изображений. Одна из них, самая большая на данный момент, называется TID2013 [2]. Она содержит 25 эталонных изображений и 3000 их искаженных вариантов. Дополнительно в базе собраны 3000 усредненных субъективных оценок, выставленных экспертами для искаженных изображений, относительно эталонов. Такие оценки называют  $MOS$  (от англ. Mean Opinion Score). Все исследователи сравнивают субъективные оценки  $MOS$  с различными количественными оценками путем вычисления для них коэффициентов корреляции трех типов: Пирсона, Спирмена и Кендела, стремясь получить максимальную корреляцию.

В некоторых публикациях (см., например, [3, 4]) показано, что распределение Вейбулла хорошо описывает такие характеристики изображений естественных сцен, как контраст и текстура. Это может быть использовано для сегментации и классификации данных изображений. В настоящей статье исследуется возможность описания множества локальных оценок  $ssim$  параметрами распределения Вейбулла и замены ими индекса  $SSIM$ , являющегося средним значением локальных оценок  $ssim$ .

### Функция $SSIM$ и ее свойства

**Определение функции  $SSIM$ .** Вычисление локальных оценок  $ssim$  в окнах  $x$  и  $y$  размером  $N \times N$  пикселей выполняется по формуле [1]

$$ssim(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \quad (1)$$

где  $\mu_x$  – среднее значение в окне  $x$ ;  $\mu_y$  – среднее значение в окне  $y$ ;  $\sigma_x^2$  – дисперсия значений в окне  $x$ ;  $\sigma_y^2$  – дисперсия значений в окне  $y$ ;  $\sigma_{xy}$  – ковариация в окнах  $x$  и  $y$ ; константы  $c_1 = (k_1L)^2$  и  $c_2 = (k_2L)^2$ ,  $L = 255$ , – значения динамического диапазона допустимых значений яркости;  $k_1 = 0,01$ ,  $k_2 = 0,03$  – экспериментально определенные константы.

Итоговый индекс  $SSIM$  для изображений  $A$  и  $B$  вычисляется как среднее арифметическое локальных оценок по формуле

$$SSIM(A, B) = \frac{1}{M} \sum_{x,y} ssim(x, y),$$

где  $A$  и  $B$  – два изображения равных размеров,  $M$  – количество окон. Локальные оценки  $ssim(x, y)$  можно вычислять в окрестности каждого пикселя или с некоторым шагом.

Фактически индекс  $SSIM$  сконструирован из трех компонентов: оценки локальной яркости  $l$ , оценки локального контраста  $c$  и оценки локальной структуры  $s$  – посредством вычисления коэффициента корреляции Пирсона:

$$l(x, y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1},$$

$$c(x, y) = \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2},$$

$$s(x, y) = \frac{\sigma_{xy} + c_3}{\sigma_x \sigma_y + c_3},$$

где  $c_3 = c_2/2$ .

В более общем виде локальная оценка сходства изображений в окрестности пиксела вычисляется по формуле

$$ssim(x, y) = l(x, y)^{p_1} \cdot c(x, y)^{p_2} \cdot s(x, y)^{p_3}, \quad (2)$$

где  $p_i$  – весовые коэффициенты. Если эти коэффициенты равны единице, формула (2) превращается в (1). Чаще всего на практике применяется формула (1).

При обобщении *SSIM* для изображений разных размеров можно применять окна  $x$  и  $y$  различных размеров пропорционально размерам изображений сцены, но с привязкой к объективно определенным центрам окон.

В некоторых приложениях требуется использовать неотрицательные локальные оценки. В таких случаях индекс *SSIM* может быть преобразован в нормализованное представление по формуле

$$NSSIM = (SSIM + 1) / 2. \quad (3)$$

Диапазон значений *NSSIM* будет в пределах  $[0, +1]$ .

**Метрические свойства *SSIM*.** В литературе индекс *SSIM* часто называют метрикой, однако он не является метрикой, поскольку  $SSIM(A, A) = 1$ , а не нулю. Кроме того,  $SSIM(A, B)$  может иметь отрицательные значения. Следует отметить, что  $SSIM(A, B) = SSIM(B, A)$ .

Функция *NSSIM* также не является метрикой, так как для нее не выполняется аксиома неравенства треугольника. Это легко доказать через контрпример. Возьмем два изображения  $A = \text{'Im01-01-2'}$  и  $B = \text{'Im01-01-1'}$  из базы TID2013 и построим третье в виде негативной копии  $A$ , т. е.  $C = 255 - \text{'Im01-01-2'}$ , тогда

$$NSSIM(A, B) = 0,9611 > NSSIM(A, C) + NSSIM(C, B) = 0,1045 + 0,1402.$$

В работе [3] утверждается, что компоненты функции *SSIM* можно преобразовать в две метрические функции:

$$S_1(x, y) = \sqrt{1 - l(x, y)}, \quad S_2(x, y) = \sqrt{1 - c(x, y)s(x, y)}.$$

Диапазон значений индекса *SSIM* лежит в пределах от  $-1$  до  $1$ . Значение  $1$  достигается при полном совпадении сравниваемых изображений. Теоретически значение  $-1$  должно достигаться при сравнении изображения  $A$  и его полной противоположности, однако это условие не выполняется. Анализ формул (1) и (2) показывает стремление к неопределенности индекса *SSIM* при сравнении, например, абсолютно белого и черного изображений. Чтобы избежать этого, введены положительные константы  $c_i$ . Функция *ssim*, реализованная в системе Матлаб, для таких изображений выдает ответ  $SSIM(A, B) = 0$ . Сравнение произвольных цветных изображений с их негативными копиями (т. е. пар изображений с меньшими отличиями) дает отрицательные значения *ssim*, но больше  $-1$ . Найти пример пары изображений  $A$  и  $B$ , для которых  $SSIM(A, B) = -1$ , не удалось.

**Связь *SSIM* с другими функциями.** В работе [5] выполнен аналитический анализ индекса *SSIM*. Показано, что при равных значениях  $\mu_x = \mu_y$  и  $\sigma_x = \sigma_y$  индекс  $SSIM = \rho$ , т. е. индекс

упрощается до коэффициента корреляции Пирсона. Отметим, что на практике это маловероятная ситуация. Если один из четырех указанных параметров равен нулю, значение индекса *SSIM* также стремится к нулю, но его поведение нестабильно, поэтому в уравнениях (1), (2) добавлены небольшие константы.

В работе [6] показано, что пиковое отношение сигнала к шуму *PSNR* имеет аналитическую связь с индексом *SSIM*. Напомним, что для двух изображений *A* и *B*

$$PSNR(A, B) = 10 \log_{10} (255^2 / MSE(A, B)), \quad MSE(A, B) = \frac{1}{N} \sum_i^N (a_i - b_i)^2,$$

где *N* – число пикселей в каждом из изображений *A* и *B*; *a<sub>i</sub>* и *b<sub>i</sub>* – значения яркости пикселей.

Отношение *PSNR* можно переписать в виде

$$PSNR(A, B) \sim 10 \log_{10} \left[ \frac{2\sigma_{xy}(l(A, B) - SSIM(A, B))}{255^2 SSIM(A, B)} + \left( \frac{\mu_A - \mu_B}{255} \right)^2 \right].$$

Функция *l(x, y)* описана выше. Выражение *PSNR* может быть упрощено для практических приложений, так как для многих типов искажений изображений значение *l(A, B) > 0,991*, т. е. примерно равно единице. Экспериментально в работе [4] показано, что для изображений, искаженных в результате применения фильтра Гаусса, а также сжатия алгоритмами JPEG и JPEG2000, оценки *PSNR* и *SSIM* при сравнении с оригиналами изображений имеют описанную выше нелинейную зависимость.

Существует класс функций, вычисляющих различия в цифровых изображениях. Методика их построения подробно описана в монографии [7]. Было показано, что оценка *MSE* является частным случаем этой методики. Основным способом анализа различий в 1990-е гг. было сравнение яркостей соответствующих пикселей. Особенностью предложенной методики был учет не только яркостей пикселей с одинаковыми координатами, но и их ближайшей окрестности. К сожалению, данная методика не получила широкого применения в построении мер сравнения изображений.

Индекс *SSIM* можно преобразовать в индекс структурного отличия *DSSIM* (от англ. structural DiSSIMilarity), который выражается через индекс *SSIM*:

$$DSSIM_1 = (1 - SSIM) / 2 \quad \text{либо} \quad DSSIM_2 = 1 - SSIM.$$

*DSSIM<sub>i</sub>(A, B)* равен нулю при полном совпадении сравниваемых изображений (*A=B*) и не имеет отрицательных значений. Диапазон его значений от 0 до +1 в первом случае и до +2 во втором.

Функции *DSSIM<sub>i</sub>* не являются метриками, так как для них не выполняется аксиома неравенства треугольника. Это легко доказать через контрпример. Возьмем изображение *A = 'Im02-01-2'* из базы TID2013 и построим два других: *B = A-1* (яркость в *A* уменьшаем на единицу) и негативную копию *A*, т. е. *C = 255 - A*. Тогда верно равенство

$$DSSIM_1(A, B) = 0,8605 > DSSIM_1(A, C) + DSSIM_1(C, B) = 0,000\,054\,561 + 0,8539.$$

### Варианты модификации индекса *SSIM*

**Использование разных весовых коэффициентов.** В статье [8] представлен вариант подбора лучшего множества коэффициентов *p<sub>i</sub>* для формулы (2). Автор выполнил достаточно большой перебор вариантов на изображениях одной из стандартных баз TID2008 и предложил использовать следующий набор коэффициентов: *p<sub>1</sub> = 0,61*, *p<sub>2</sub> = 0,077*, *p<sub>3</sub> = 0,241*. Однако применение этого набора коэффициентов в функции *ssim*, реализованной в Матлаб, дает в результате сравнения изображений разных сцен комплексные значения. Скорее всего, для другого набора изображений эти коэффициенты будут иными, т. е. указанные коэффициенты не являются оптимальными.

**Мультимасштабный индекс MSSIM.** Авторы индекса SSIM предложили его усовершенствованный мультимасштабный вариант [9]:

$$MSSIM(x, y) = [I_M(x, y)]^{\alpha_M} \cdot \prod_{j=1}^M [c_j(x, y)]^{\beta_j} [s_j(x, y)]^{\gamma_j}.$$

Для сравниваемых изображений строятся пирамиды Гаусса из  $M$  уровней (обычно  $M = 5$ ) и для каждого уровня вычисляются компоненты  $c_j$  и  $s_j$ . Компонент  $I_M$  вычисляется только для яркости исходных изображений. Для упрощения вычислений авторы предлагают применять равные значения констант  $\alpha_j = \beta_j = \gamma_j$  для всех уровней представления изображений и  $\sum_{j=1}^M \gamma_j$ .

Индекс SSIM вычисляется довольно долго, его модификация в виде индекса MSSIM вычисляется еще дольше при незначительном увеличении ранговых коэффициентов корреляции между величинами MSSIM и MOS, полученными на больших множествах изображений, поэтому индекс MSSIM практически не используется.

**Интеграция разных оценок с помощью нейронных сетей.** В статье [10] предлагается использовать искусственные нейронные сети для повышения количественной оценки с визуальной оценкой качества изображений MOS. Сформулированы требования к набору функций, используемому для вычисления входных данных сети:

- отбираются наиболее точные оценочные функции с максимальными коэффициентами ранговой корреляции относительно субъективных оценок MOS, выставленных экспертами;
- выбранные функции не должны использовать подобные признаки и иметь низкую взаимную корреляцию оценок одинаковых изображений.

Были выполнены эксперименты с шестью типами нейронных сетей и 38 вариантами выбора нейронов. В результате авторы построили интегрированную функцию оценки качества изображений путем отбора пяти лучших оценочных функций на базе четырехуровневой каскадной нейронной сети, что позволило повысить значения ранговых коэффициентов корреляции Спирмена и Кендалла на 0,07 по сравнению с индивидуальными результатами оценочных функций.

Анализ, выполненный в работе [10], показал, что разные функции по-разному оценивают искажения различных типов. Данное направление исследований перспективно, но для дальнейшего повышения точности оценок качества требуется объединять большее число оценочных функций, новые признаки и дополнительно учитывать особенности человеческого восприятия изображений.

**Описание отдельных характеристик параметрами их статистических распределений.** В статьях [3, 4] показано, что некоторые локальные характеристики изображений, например контраст, текстуру и др., можно описывать распределением Вейбулла и обобщенным нормальным распределением. В частности, для 75–90 % реальных изображений их локальный контраст, оцениваемый через статистику градиентов, очень хорошо описывается двухпараметрическим распределением Вейбулла. В работе Шольте [11] утверждается, что реакция человеческого мозга хорошо коррелирует со статистикой Вейбулла при анализе естественных изображений.

Распределение Вейбулла имеет два параметра: форму  $c$  и масштаб  $b$ . Функция плотности распределения Вейбулла для переменных  $x > 0$  записывается в виде

$$f(x) = \frac{c}{b} \left(\frac{x}{b}\right)^{c-1} e^{-\left(\frac{x}{b}\right)^c},$$

где  $0 < b$  – параметр масштаба,  $0 < c$  – параметр формы.

Такие характеристики распределения Вейбулла, как среднее  $\mu_{Weib}$ , медиана  $Median_{Weib}$  и мода  $Mode_{Weib}$ , не совпадают с одноименными характеристиками, вычисляемыми для множества данных. Они определяются по формулам

$$\mu_{Weib} = \lambda \Gamma(1+1/c),$$

$$Median_{Weib} = \lambda (\ln 2)^{1/c},$$

$$Mode_{Weib} = \lambda ((c - 1)/c)^{1/c} \text{ при } c > 1, \quad Mode_{Weib} = 0 \text{ при } c \leq 1,$$

где  $\Gamma(y)$  – гамма-функция от  $y$ .

Как правило, характеристики распределения Вейбулла не равны друг другу, что имеет место для аналогичных характеристик нормального распределения.

Отметим, что для некоторых совокупностей данных  $x$  кривая плотности их распределений близка к кривой нормального распределения. Например, распределение Вейбулла при параметре формы  $b = 3,44$  выглядит как функция плотности нормального распределения за исключением хвостов распределения (рис. 1), а при  $b = 1,0$  распределение Вейбулла идентично экспоненциальному распределению [12].

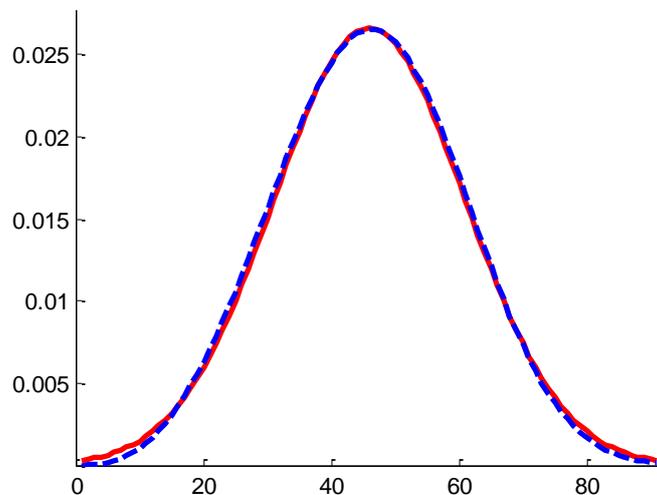


Рис. 1. График нормального распределения данных с параметрами  $\mu = 45$ ,  $\sigma = 15$  (красного цвета) и распределения Вейбулла с параметрами  $c = 50$ ,  $b = 3,44$  (синего цвета)

В рамках настоящего исследования была выдвинута следующая гипотеза: при формировании глобальной (единой) оценки сходства изображений вместо вычисления среднего значения множества локальных оценок лучше использовать параметр формы кривой распределения, описывающего гистограмму этого множества. Поскольку распределение Вейбулла и некоторые другие не работают с нулевыми и отрицательными данными, все локальные оценки *ssim* были нормализованы по формуле (3).

Гистограммы локальных оценок были аппроксимированы разными распределениями. Распределение Вейбулла оказалось самым гибким и лучше других описывало асимметричные данные. Далее анализировалось, что следует выбрать в качестве новой оценки совокупности локальных данных: параметр масштаба или формы кривой Вейбулла либо среднее, медиану, моду, коэффициенты асимметрии и эксцесса распределения Вейбулла. Согласно исследованиям У. А. Шухарта [13] и Д. Вилера [14] последние две характеристики бесполезны при анализе реальных данных, особенно имеющих экстремальные значения. В рассматриваемом случае некоторые локальные оценки всегда принимают крайние значения допустимого диапазона, например 1 или 0, поэтому коэффициенты асимметрии и эксцесса были отброшены. После первичных экспериментов параметр формы  $c$  также был отвергнут, поскольку его значение изменялось от 0 и до  $+\infty$  для идентичных изображений, что усложняло получение финальных оценок.

### Индекс SSIM как оценка качества изображений

**База TID2013.** База содержит 25 эталонных изображений, 24 из них взяты из базы изображений Kodak (<http://r0k.us/graphics/kodak/>), а 25-е синтезировано [2]. Ко всем изображениям применены 24 вида искажений (таблица), из них 12 типов искажений могут возникать при ре-

гистрации реальных изображений. Для каждого типа искажений создано пять вариантов изображений, которые имеют значения пикового отношения сигнал-шум  $PSNR$ , примерно равные 21, 24, 27, 30 и 33дБ. Всего в базе содержится 3000 искаженных цветных изображений размером  $512 \times 384$  пиксела. Дополнительно для каждого искаженного изображения были получены субъективные оценки  $MOS$  его сравнения с эталоном в шкале [0, 9]. Для этого привлекали 971 эксперта из четырех стран.

Типы искажений изображений в базе TID2013

Тип искажения	Операции, при которых возникает
1. Аддитивный гауссовский шум	Регистрация изображений
2. Аддитивный шум в цветовых компонентах	Регистрация изображений
3. Пространственно-коррелированный шум	Цифровая фотография, сжатие
4. Маскированный шум	Сжатие
5. Высокочастотный шум	Сжатие
6. Импульсный шум	Регистрация изображений
7. Шум квантования	Регистрация изображений, гамма-коррекция
8. Гауссовское размытие	Регистрация изображений
9. Фильтрация шума	Шумоподавление
10. Сжатие алгоритмом JPEG	Сжатие
11. Сжатие алгоритмом JPEG2000	Сжатие
12. Сбои в канале передачи JPEG-изображений	Передача данных
13. Искажения при передаче JPEG2000-изображений	Передача данных
14. Искажения с низкой степенью оригинальности	Сжатие
15. Локальные блочные искажения	Регистрация изображений
16. Изменения яркости методом mean-shift	Регистрация изображений
17. Изменения контраста	Регистрация изображений, гамма-коррекция
18. Изменения насыщенности цвета	Регистрация изображений, сжатие
19. Мультипликативный гауссовский шум	Регистрация изображений, шумоподавление
20. Комфортный шум	Сжатие
21. Сжатие с потерями зашумленных изображений	Сжатие, шумоподавление
22. Квантование цветного изображения с дизерингом	Сжатие
23. Хроматические аберрации	Регистрация изображений
24. Восстановление изображений по разреженной сетке отсчетов	Сжатие, восстановление

**Анализ корреляции величин  $SSIM$  и  $MOS$ .** Практически все исследователи оценивают корректность оценок сходства или качества изображений, полученных различными количественными функциями, степенью их корреляции с субъективными оценками  $MOS$ , которые обычно прикладываются к базам изображений. База TID2013 не является исключением. Для ее изображений коэффициенты ранговой корреляции Спирмена и Кендалла между величинами  $SSIM$  и  $MOS$  согласно [2] равны 0,6370 и 0,4636, а между  $MSSIM$  и  $MOS$  – 0,7872 и 0,6079 соответственно. Проверка дала немного отличающиеся значения корреляции  $SSIM$  и  $MOS$ . Так, коэффициенты ранговой корреляции Спирмена и Кендалла при использовании стандартных функций системы Матлаб равнялись 0,6194 и 0,4391, а коэффициент корреляции Пирсона равнялся 0,5776.

На рис. 2 показана связь между величинами  $SSIM$  и  $MOS$  для каждого из 3000 изображений базы TID2013. Красная прямая с уравнением  $y = 5x + 0,2$  соответствует линейной аппроксимации этих оценок, зеленая кривая с уравнением  $y = 10x^2 - 10x + 5,7$  – квадратичной аппроксимации. Приведенный график и коэффициенты корреляции указывают на слабую связь индекса  $SSIM$  и субъективных оценок качества изображений  $MOS$  из базы TID2013.

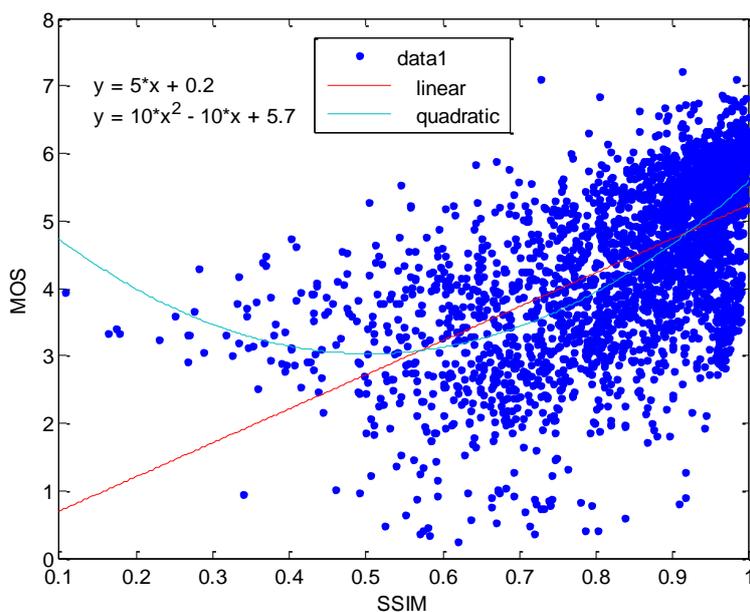


Рис. 2. График соответствия оценок *SSIM* и *MOS*, его аппроксимация прямой и параболой

В базе TID2013 есть изображения с локальными блочными искажениями (15-й тип искажений). Эти искажения формируются в результате наложения на оригинальные изображения небольших квадратных блоков одинакового размера и фиксированного цвета. В пяти вариантах искажений используются 10, 8, 6, 4 и 2 блока. Интересно, что многие эксперты поставили таким изображениям очень низкие оценки – менее четырех, а некоторым – менее трех из девяти возможных. При этом подавляющее большинство различных количественных оценочных функций, включая *SSIM*, подтвердило близость искаженных изображений к эталонным.

Можно констатировать, что оценки *MOS* изображений очень субъективны. При изменении группы экспертов такие оценки могут изменяться, и достаточно существенно. При этом процесс их получения достаточно долгий и затратный. Таким образом, они не могут служить в качестве эталонных оценок.

**Экспериментальные исследования.** В экспериментах использовались нормализованные по формуле (3) значения локальных оценок функции *ssim*, а для аппроксимации их распределения применялась функция *fitdist*. Обе функции реализованы в системе Матлаб. Их вычислительная сложность не оценивалась, но время выполнения функции *fitdist* было в 25 раз меньше, чем *ssim*.

Для пары изображений на рис. 3 среднеквадратичные отклонения теоретических кривых Гаусса и Вейбулла равны  $MSE_{\text{Norm}} = 0,014\ 68$  и  $MSE_{\text{Weib}} = 0,002\ 16$  соответственно, из чего следует, что распределение Вейбулла точнее аппроксимирует гистограмму локальных оценок. Это видно на рис. 4: кривой красного цвета показано приближение нормальным распределением, а синим – распределением Вейбулла. Гистограммы локальных оценок (рис. 4, 7, 9, 12) наглядно показывают, что при сравнении эталонных изображений с искаженными версиями многие из локальных оценок оказываются близки к единице. При этом гистограммы и их пики имеют явный сдвиг к единице, т. е. данные не распределены нормально. Такой же вывод подтверждает проверка распределения данных на нормальность, выполненная по критериям  $\chi$ -квадрат, Колмогорова – Смирнова, Андерсона – Дарлинга, Лилиефорса.

Значение нормализованного индекса  $SSIM_{\text{Norm}}$  для пары изображений на рис. 3 равно 0,7991, теоретический параметр масштаба  $b$  кривой Вейбулла равен 0,8386, а мода распределения Вейбулла равна 0,8308, т. е. они ближе к максимальному значению, равному единице, и близки между собой. Однако субъективная оценка сходства этих изображений достаточно низкая:  $MOS = 2,742\ 86$ . Данный факт подтверждает различие между рассматриваемыми субъективными и объективными оценками изображений.

Если сравнивать любое, даже максимально искаженное, изображение с самим собой, все локальные количественные оценки будут равны единице. Субъективная оценка будет зависеть от поставленной задачи – оценить сходство либо качество изображений.



Рис. 3. Изображение Im24 и его искаженный дизерингом вариант пятого уровня Im24-22-5

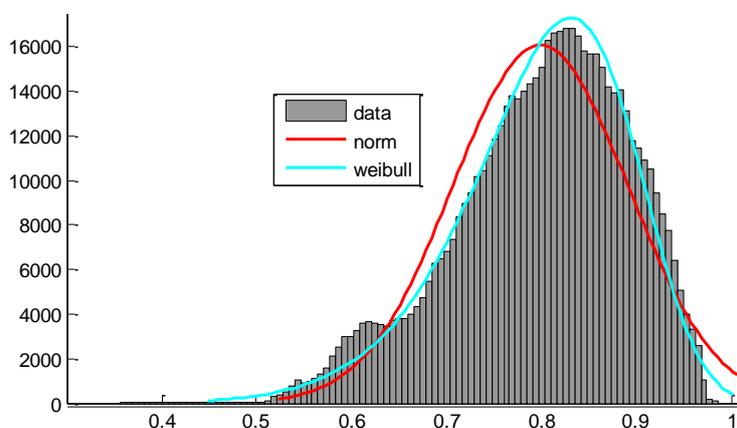


Рис. 4. Гистограмма локальных оценок *ssim* сходства изображений Im24, Im24-22-5 и ее приближение распределениями Гаусса и Вейбулла

В процессе исследования был выполнен анализ точности приближения гистограмм локальных оценок *ssim* распределениями Гаусса и Вейбулла. На рис. 5 показан график отсортированных разностей среднеквадратичного отклонения аппроксимации гистограмм локальных оценок *ssim* между эталонами и их искаженными копиями из базы TID2013 кривыми Гаусса и Вейбулла.

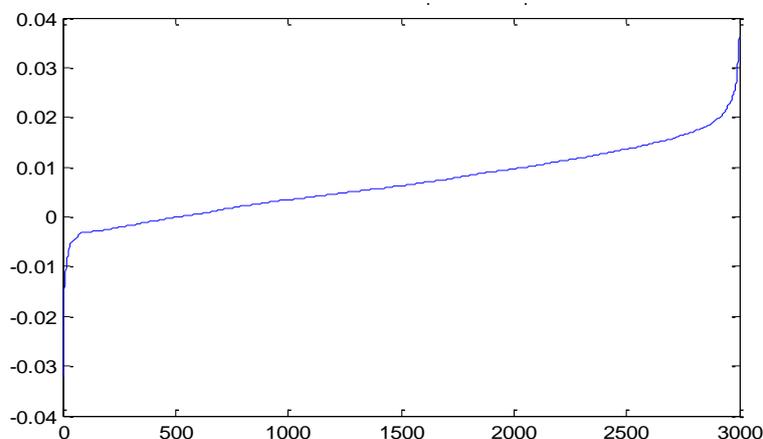


Рис. 5. График разностей среднеквадратичного отклонения аппроксимации гистограмм локальных оценок *ssim* функциями нормального распределения и Вейбулла

Минимальное значение разности равно  $-0,0322$ , максимальное –  $+0,0362$ . Для 514 искаженных изображений из 3000 (17,133 %) эта разность была отрицательной, т. е. для них нормальное распределение давало несколько лучшее приближение. Отсюда следует, что распределение Вейбулла в целом точнее нормального описывает совокупность локальных оценок  $ssim$  и его параметры можно использовать для формирования уточненного индекса  $SSIM$ .

Для полноты анализа было выполнено сравнение всех изображений, представленных в базе TID2013, друг с другом. Приведем несколько примеров (рис. 6–9).



Рис. 6. Эталонное изображение Im11 и его слабоконтрастная копия Im11-17-5

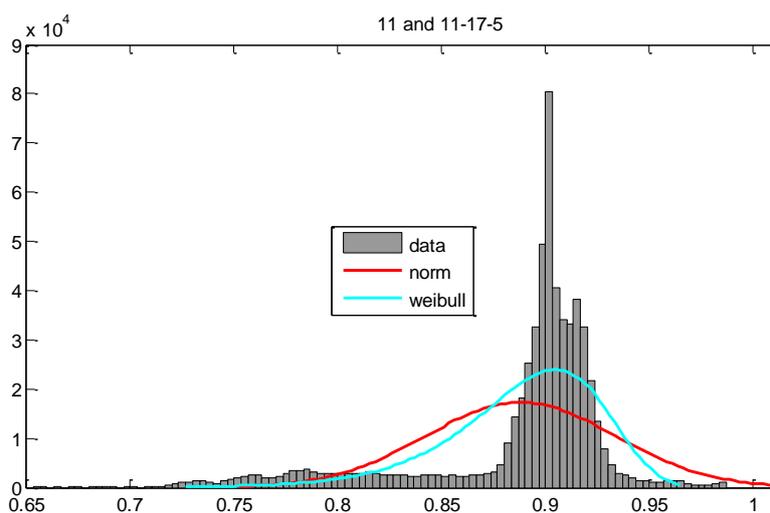


Рис. 7. Гистограмма нормализованных локальных оценок  $ssim(Im11, Im11-17-5)$  и ее аппроксимация функциями Гаусса и Вейбулла

Гистограммы на рис. 5 и 7 показывают, что при сравнении изображений одной сцены большинство локальных оценок близко к единице, т. е. локально изображения похожи. Однако гистограммы не симметричны относительно среднего значения, равного индексу  $NSSIM$ , а сдвинуты вправо. Их форма лучше описывается кривой Вейбулла, при этом среднеквадратичная погрешность такой аппроксимации меньше:  $MSE_{Norm} = 0,052\ 09$ ,  $MSE_{Weib} = 0,025\ 808$ . Приведем оценки изображений, представленных на рис. 5:  $ssim_{norm} = 0,8882$ ,  $b = 0,9060$ ,  $mode_{Weib} = 0,9049$ ,  $MOS = 3,216\ 22$ . Очевидно, что численные оценки больше отражают сходство данных изображений, а субъективная оценка  $MOS$  – качество искаженного изображения относительно эталонного.

Рассмотрим для примера оценки сравнения пары изображений Im01 и Im01-22-5 из базы TID2013 (рис. 8). Второе изображение – это копия первого после квантования цвета методом дизеринга. Усредненная субъективная оценка  $MOS$  искаженного изображения в качественной шкале [0–9] равна 2,1944.

На рис. 9 изображена гистограмма локальных оценок  $ssim$  пофрагментных отличий в данной паре изображений. Вершина красной кривой указывает на среднее значение множе-

ства локальных оценок, которое и является финальной оценкой индекса  $NSSIM = 0,8372$ . Синим цветом показано приближение данных кривой распределения Вейбулла с параметром масштаба  $b = 0,8652$ . Усредненная субъективная оценка  $MOS$  искаженного изображения равна 2,1944.



Рис. 8. Пара тестовых изображений Im01 и Im01-22-5 из базы TID2013

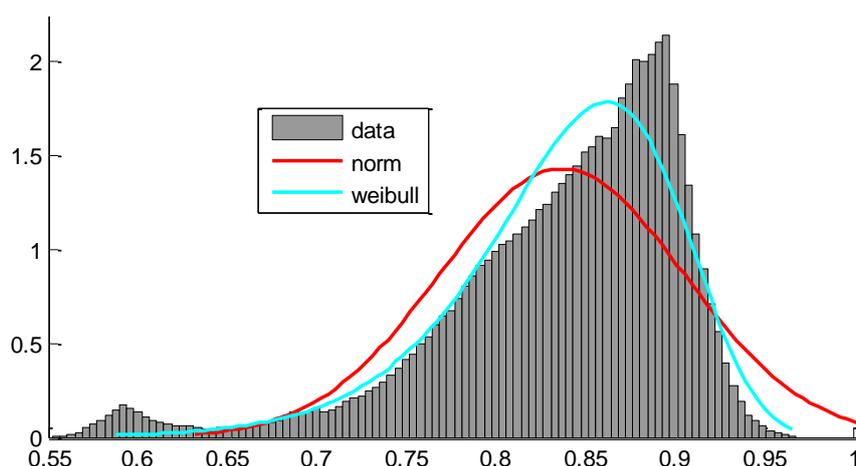


Рис. 9. Гистограмма локальных оценок  $ssim$  пофрагментных отличий в паре изображений Im01 и Im01-22-5

На рис. 10 показаны отклонения теоретических кривых от реальных данных гистограммы, представленной на рис. 9, и среднеквадратичные отклонения для двух рассматриваемых распределений. Графики подтверждают, что кривая Вейбулла в данном примере имеет на порядок лучшее приближение, чем кривая Гаусса ( $MSE_{Weib} = 0,001\ 7271$  против  $MSE_{Gauss} = 0,017\ 174$ ).

Если убрать два десятка выбросов, то отклонения кривой Вейбулла приближенно описываются прямой, чего нельзя сказать о кривой Гаусса. Визуальный анализ пары изображений на рис. 8 показывает, что искаженное изображение очень похоже на оригинал, т. е. это изображение одной и той же сцены, но полного совпадения по цвету, текстуре и геометрическим деталям нет. Можно сделать вывод, что низкое значение  $MOS$  больше отражает субъективную оценку качества правого изображения по сравнению с оригиналом, а количественные оценки определяют сходство изображенной сцены. При этом параметр масштаба  $b = 0,8652$  точнее, чем  $NSSIM = 0,8372$ , в целом характеризует локальные оценки, представленные на рис. 9.

Кроме сравнения искаженных изображений с их эталонами было выполнено выборочное сравнение изображений разных сцен и разного качества из базы TID2013 между собой, а также

этих изображений с их негативными копиями. Количество всех таких пар изображений очень велико ( $6050 \times 3025 = 18\,301\,250$ ), поэтому получить для них оценки *MOS* практически невозможно.

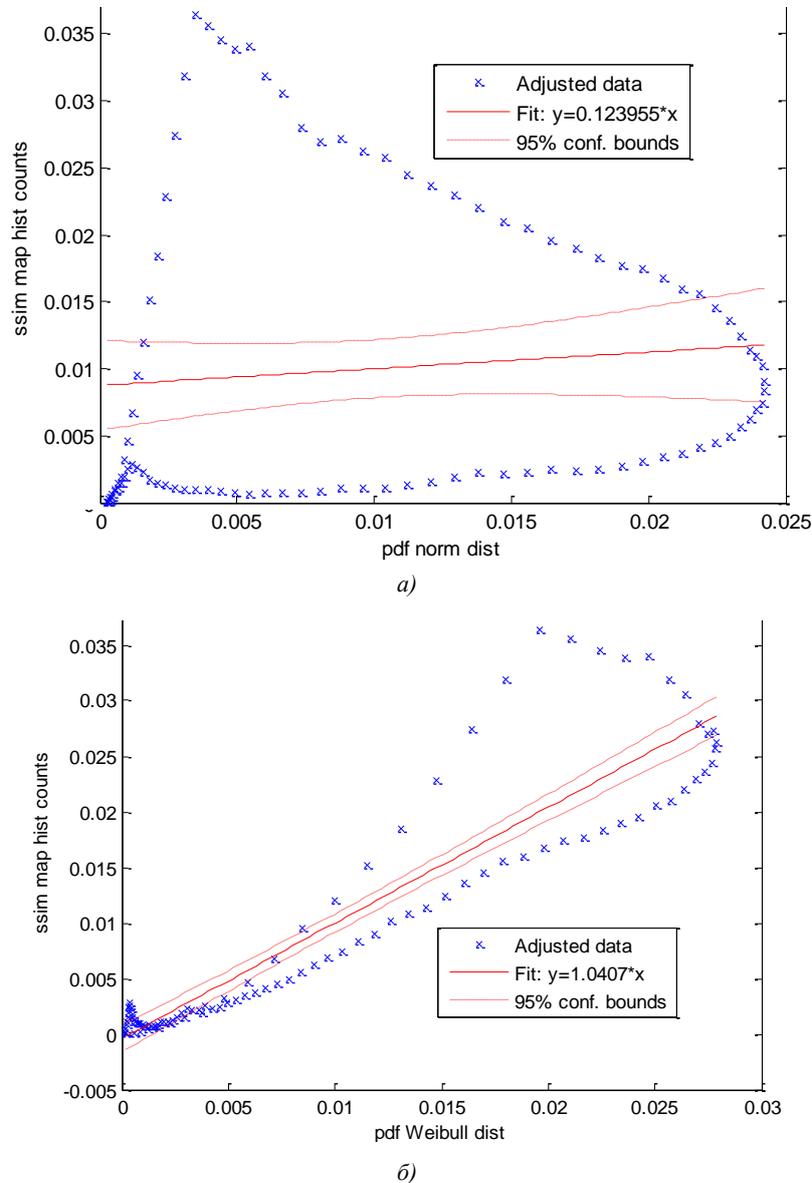


Рис. 10. Отклонения теоретических кривых Гаусса (а) и Вейбулла (б) от реального распределения локальных оценок *ssim*

Анализ результатов сравнения ряда пар разнообразных изображений подтвердил вывод о более точной аппроксимации гистограммы локальных оценок *ssim* распределением Вейбулла. На рис. 11 приведен пример пары изображений с разными типами гауссовского шума. Визуально они практически не отличимы, но все локальные оценки *ssim* (рис. 12) несколько меньше единицы, что означает несовпадение на локальном уровне, а их гистограмма не имеет нормального распределения. Глобальная оценка сходства, полученная из распределения Вейбулла, равна 0,9815, индекс  $SSIM_{\text{Norm}}=0,9732$ . Среднеквадратичное отклонение кривой Вейбулла от графика гистограммы локальных оценок меньше, чем у кривой Гаусса: 0,0540 против 0,0676. Отметим, что коэффициент корреляции Пирсона между оценками  $SSIM_{\text{Norm}}$  и параметром масштаба  $b$  распределения Вейбулла, полученный для 3000 изображений базы TID2013, равен 0,9923. Это говорит о их близости.

Значение параметра моды распределения Вейбулла во всех экспериментах было очень близко к значению параметра масштаба  $b$  этого распределения. Для уменьшения вычислений в качестве глобальной меры сходства двух изображений вместо среднего от множества локальных оценок  $ssim$  предлагается использовать параметр масштаба  $b$  распределения Вейбулла, которое аппроксимирует гистограмму этих локальных оценок.



а)

б)

Рис. 11. Изображение Im01-01-1 с аддитивным гауссовским шумом (а) и оно же с мультипликативным гауссовским шумом (б)

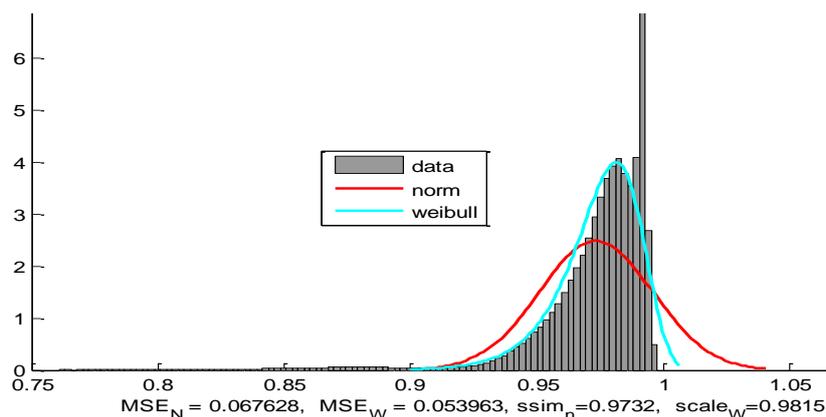


Рис. 12. Гистограмма локальных оценок  $ssim$  пофрагментных отличий в паре изображений Im01-01-1 и Im01-19-1

**Заключение.** В работе исследованы свойства широко применяемого индекса структурного подобия изображений  $SSIM$ . Показано, что данная мера и варианты ее нормализации не являются метрическими функциями. По результатам исследований можно сделать вывод, что локальные количественные оценки  $ssim$  характеризуют сходство изображений, а субъективная оценка  $MOS$  делает акцент на визуальные отличия искаженного изображения.

Были исследованы также варианты аппроксимации гистограммы локальных оценок сходства разными распределениями. Распределение Вейбулла оказалось самым гибким. Из трех характеристик этого распределения: параметра масштаба  $b$ , параметра формы  $c$  и медианы теоретической кривой Вейбулла – предложено в качестве глобальной меры сходства двух изображений вместо популярного индекса  $SSIM$  использовать параметр масштаба  $b$  распределения Вейбулла. Показано, что значение этого параметра, вычисленное для множества локальных оценок  $ssim$ , точнее, чем их среднее арифметическое, отражает сходство или различие двух изображений. Для множества нормализованных локальных оценок  $ssim$  данный параметр всегда будет находиться в диапазоне  $[0, 1]$  и равняться единице, если изображения идентичны.

Работа частично выполнена в рамках проекта БРФФИ Ф18МС-028.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Image quality assessment: from error visibility to structural similarity / Z. Wang [et al.] // *IEEE Transactions on Image Processing*. – 2004. – Vol. 13, no. 4. – P. 600–612.
2. Image database TID2013: peculiarities, results and perspectives / N. Ponomarenko [et al.] // *Signal Processing: Image Communication*. – 2015. – Vol. 30. – P. 57–77.
3. Geusebroek, J. A six-stimulus theory for stochastic texture / J. Geusebroek, A. W. M. Smeulders // *International Journal of Computer Vision*. – 2005. – Vol. 62, no. 1–2. – P. 7–16.
4. Xue, W. Reduced reference image quality assessment based on Weibull statistics / W. Xue, X. Mou // *Proc. of the 2nd Intern. Workshop on Quality of Multimedia Experience*. – N.Y., 2010. – P. 11–16.
5. Brunet, D. On the mathematical properties of the structural similarity index / D. Brunet, E. R. Vrscay, Z. Wang // *IEEE Transactions on Image Processing*. – 2012. – Vol. 21, no. 4. – P. 1488–1499.
6. Hore, A. Image quality metrics: PSNR vs. SSIM / A. Hore, D. Ziou // *Proc. of the 20th Intern. Conf. on Pattern Recognition*. – Washington, 2010. – P. 2366–2369.
7. Старовойтов, В. В. Локальные геометрические методы цифровой обработки и анализа изображений / В. В. Старовойтов. – Минск : Ин-т техн. кибернетики НАН Беларуси, 1997. – 284 с.
8. Сидоров, Д. В. Модификация алгоритма SSIM / Д. В. Сидоров // *Прикладная информатика*. – 2010. – № 4. – С. 123–125.
9. Wang, Z. Multiscale structural similarity for image quality assessment / Z. Wang, E. P. Simoncelli, A. C. Bovik // *Proc. of the 37th Asilomar Conf. on Signals, Systems and Computers*. – USA, CA, 2004. – Vol. 2. – P. 1398–1402.
10. Еремеев, О. И. Интегрированная метрика визуального качества изображений при наличии эталона / О. И. Еремеев // *Системы обработки информации*. – 2014. – № 5. – С. 35–42.
11. Brain responses strongly correlate with Weibull image statistics when processing natural images / H. S. Scholte [et al.] // *Journal of Vision*. – 2009. – Vol. 9, no. 29. – С. 11–25.
12. Статистические методы. Распределение Вейбулла. Анализ данных : ГОСТ Р 50779.27–2017. – Введ. 10.08.2017. – М. : Госстандарт России : Изд-во стандартов, 2017. – 62 с.
13. Shewhart, W. A. *Statistical Method from the Viewpoint of Quality Control* / W. A. Shewhart. – Washington : Courier Corporation, 1939. – 155 p.
14. Wheeler, D. J. Problems with Skewness and Kurtosis. Part II. What do the shape parameters do? [Electronic resource] / D. J. Wheeler // *Quality Digest Daily*. – 2011. – Mode of access: <https://www.spcpress.com/pdf/DJW231.pdf>. – Date of access: 20.05.2018.

## References

1. Wang Z., Bovik A. C., Sheikh H. R., Simoncelli E. P. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, vol. 13, no. 4, pp. 600–612.
2. Ponomarenko N., Jin L., Ieremeiev O., Lukin V., Egiazarian K., Kuo C. Image database TID2013: peculiarities, results and perspectives. *Signal Processing: Image Communication*, 2015, vol. 30, pp. 57–77.
3. Geusebroek J., Smeulders A. W. M. A six-stimulus theory for stochastic texture. *International Journal of Computer Vision*, 2005, vol. 62, no. 1–2, pp. 7–16.
4. Xue W., Mou X. Reduced reference image quality assessment based on Weibull statistics. *Proceedings of the 2nd International Workshop on Quality of Multimedia Experience*. N.Y., 2010, pp. 11–16.
5. Brunet D., Vrscay E. R., Wang Z. On the mathematical properties of the structural similarity index. *IEEE Transactions on Image Processing*, 2012, vol. 21, no. 4, pp. 1488–1499.
6. Hore A., Ziou D. Image quality metrics: PSNR vs. SSIM. *Proceedings of the 20th International Conference on Pattern Recognition*. Washington, 2010, pp. 2366–2369.
7. Starovoitov V. V. Lokal'nye geometricheskie metody cifrovoj obrabotki i analiza izobrazhenij. *Local Geometric Methods of Digital Image Processing and Analysis*. Minsk, In-t teh. kibernetiki NAN Belarusi Publ., 1997, 284 p. (in Russian).
8. Sidorov D. V. Modifikacija algoritma SSIM [SSIM algorithm modification]. *Prikladnaja informatika [Applied Informatics]*, 2010, no. 4, pp. 123–125 (in Russian).
9. Wang Z., Simoncelli E. P., Bovik A. C. Multiscale structural similarity for image quality assessment. *Proceedings of the 37th Asilomar Conference on Signals, Systems and Computers*. USA, CA, 2004, vol. 2, pp. 1398–1402.
10. Eremeev O. I. Integrirovannaja metrika vizual'nogo kachestva izobrazhenij pri nalichii jetalona [Integrated metric of visual quality of images in the presence of a standard]. *Sistemy obrabki informacii [Information Processing Systems]*, 2014, no. 5, pp. 35–42 (in Russian).
11. Scholte H. S., Ghebreab S., Waldorp L., Smeulders A. W., Lamme V. A. Brain responses strongly correlate with Weibull image statistics when processing natural images. *Journal of Vision*, 2009, vol. 9, no. 29, pp. 11–25.
12. Statisticheskie metody. Raspreделение Vejbullla. Analiz dannyh. *State Standart R 50779.27–2017. Statistical methods. Weibull distribution. Data analysis*. Moscow, Gosstandart Rossii, Izd-vo standartov Publ., 2017, 62 p. (in Russian).
13. Shewhart W. A. *Statistical Method from the Viewpoint of Quality Control*. Washington, Courier Corporation, 1939, 155 p.
14. Wheeler D. J. Problems with Skewness and Kurtosis. Part II. What do the shape parameters do? *Quality Digest Daily*, 2011. Available at: <https://www.spcpress.com/pdf/DJW231.pdf> (accessed 20.05.2018).

---

---

**Информация об авторе**

*Старовойтов Валерий Васильевич* – доктор технических наук, профессор, главный научный сотрудник, Объединенный институт проблем информатики Национальной академии наук Беларуси (ул. Сурганова, 6, 220012, Минск, Республика Беларусь).

E-mail: [valerys@newman.bas-net.by](mailto:valerys@newman.bas-net.by)

**Information about the author**

*Valery V. Starovoitov* – D. Sc. (Engineering), Chief Researcher, The United Institute of Informatics Problems of the National Academy of Sciences of Belarus (6, Sarganova Str., 220012, Minsk, Republic of Belarus).

E-mail: [valerys@newman.bas-net.by](mailto:valerys@newman.bas-net.by)