

УДК 004; 004.93

В. В. СТАРОВОЙТОВ

ИНДЕКС SSIM НЕ ЯВЛЯЕТСЯ МЕТРИКОЙ И ПЛОХО ОЦЕНИВАЕТ СХОДСТВО ИЗОБРАЖЕНИЙ

Объединенный институт проблем информатики НАН Беларуси

В статье исследованы некоторые свойства очень популярного показателя структурного сходства изображений, называемого индексом SSIM. Согласно сайту <https://scholar.google.com> на статью [3], где он был впервые описан, за 14 лет сделано более 20000 ссылок. Этот показатель активно используется международным сообществом в научных исследованиях. Он приобрел статус неофициального международного стандарта для оценки качества изображения при наличии эталона, часто называемого метрикой качества изображений. В настоящей статье развенчиваются некоторые мифы, возникшие вокруг этого индекса. Доказана теорема утверждающая, что индекс SSIM и любые его линейные преобразования не являются метрическими функциями. Во многих публикациях и в пакете прикладных программ Матлаб в описании функции `ssim`, сказано, что индекс SSIM используется для измерения качества изображений. Однако этот индекс, а также любая функция сравнения с эталонным изображением (типа `full-reference`) в принципе не могут оценить качество анализируемых изображений. Они оценивают только некоторую степень сходства между изображением и его искаженной копией. В статье также показано, что индекс SSIM не всегда может корректно определить сходство изображений одной и той же сцены, в то время как коэффициент линейной корреляции Пирсона делает это намного быстрее и точнее.

Ключевые слова: оценка качества изображения, индекс SSIM, коэффициент корреляции Пирсона

Введение

За последние десятилетия были разработаны сотни алгоритмов для сравнения изображений [1]. Часть из них ориентирована на оценку качества изображения (IQA). Факторы, влияющие на качество, включают в себя равномерность освещения и яркости, разрешение и геометрию, контрастность и резкость, точность цветопередачи и цветовую дискриминацию наблюдаемого изображения. Алгоритмы были разделены на три группы: полный эталонный (FR), сокращенный эталонный (RR), без эталонный (NR) или слепой [2,3]. Одной из самых популярных FR-мер является структурный индекс сходства изображений SSIM [4]. Он используется во многих приложениях и имеет множество клонов, таких как FSSIM, MS-SSIM, HWSSIM, PSSIM, CW-SSIM и других [5,6]. Все они оценивают яркость, контраст и структуру двух сравниваемых изображений в наборе соответствующих окон. На последнем этапе среднее арифметическое локальных оценок принимается как глобальное значение показателя качества изображения на основе SSIM.

Индекс SSIM был разработан как улучшение традиционных методов оценки сигналов, таких как среднеквадратическая ошибка (MSE) и отношение сигнал / шум (PSNR). Статья [4], где был введен индекс SSIM, цитировалась более 22000 раз, согласно Google Scholar. В работе [7] авторы описали результаты статистического исследования этого индекса. В статье [8] показано, что локальные оценки индекса SSIM имеют четкую аналитическую зависимость с локальными среднеквадратичными оценками MSE. В [9] приведена аналитическая зависимость пикового отношения сигнала к шуму (PSNR) и индекса SSIM. Там же экспериментально показана близость этих оценок в случае сравнения эталонных изображений с их копиями, подвергнутыми гауссовому размытию, сжатию алгоритмами JPEG и JPEG2000. В работе [10] аналитически и экспериментально показано, что для медицинских изображений с фрагментами больших или малых значений средней яркости локальные SSIM-оценки не стабильны, это особенно заметно при сжатии изображений алгоритмом JPEG2000. Тем не ме-

нее, во многих исследованиях индекс SSIM используется как неофициальный стандарт для оценки качества искаженного изображения при наличии эталонного изображения и его часто называют метрикой качества изображений.

Настоящая статья продолжает исследования индекса SSIM, начатые в работе [11].

Определение индекса структурного сходства

Индекс основан на локальных оценках $ssim$ между двумя окнами x и y размером $N \times N$ пикселей, рассчитанных вокруг соответствующих пикселей на изображениях A и B согласно формуле [4]:

$$ssim(x, y) = \frac{(2\mu_x\mu_y + \tilde{n}_1)(2\sigma_{xy} + \tilde{n}_2)}{(\mu_x^2 + \mu_y^2 + \tilde{n}_1)(\sigma_x^2 + \sigma_y^2 + \tilde{n}_2)}, \quad (1)$$

где μ_x – среднее значение в окне x ; μ_y – среднее значение в окне y ; σ_x^2 – дисперсия в окне x ; σ_y^2 – дисперсия в окне y ; σ_{xy} – ковариация окон x и y ; константы $c_1 = (k_1L)^2$ и $c_2 = (k_2L)^2$, $L = 255$, определяют динамический диапазон яркости; $k_1 = 0,01$, $k_2 = 0,03$ – экспериментально определенные константы.

Глобальный индекс SSIM для изображений A и B рассчитывается как среднее арифметическое локальных оценок по формуле:

$$SSIM(A, B) = \frac{1}{M} \sum_{x,y} ssim(x, y), \quad (2)$$

где A и B – два изображения одинакового размера, M – количество окон.

Каждая локальная оценка $ssim(x, y)$ рассчитывается в окрестности пикселя с координатами (x, y) для каждого пикселя. Значения индекса SSIM лежат в диапазоне от -1 до $+1$. Значение 1 достигается, когда сравниваемые изображения полностью идентичны.

Индекс SSIM не может оценить качество изображения: если взять изображение очень низкого качества в качестве эталонного и сравнить его с самим собой, тогда получим $SSIM = 1$. Это означает наилучшее качество исследуемого изображения. Если сравнить некачественное изображение с его слегка искаженной копией, оценка будет близка к единице. Полученные максимальные оценки не являются корректными оценками качества изображений, то есть индекс SSIM не оценивает качество изображений, а только их сходство.

Является ли SSIM метрической функцией?

В литературе индекс SSIM часто называют метрикой, но он является метрикой в математическом смысле. Более того, можно доказать следующую теорему.

Теорема 1. Любая функция, полученная линейным преобразованием индекса SSIM, не является метрикой.

Доказательство. Функция f является метрикой, если для нее выполняются четыре аксиомы:

- 1) $f(x, y) \geq 0$,
- 2) $f(x, y) = 0$ iff $x = y$,
- 3) $f(x, y) = f(y, x)$,
- 4) $f(x, z) \leq f(x, y) + f(y, z)$.

Линейное преобразование функции SSIM может быть описано следующим образом:

$$\text{NewSSIM} = \alpha \text{SSIM} + \beta,$$

где α и β – константы.

Из второй аксиомы и определения индекса SSIM получаем $\max(\text{SSIM}) = 1$, откуда следует $\alpha + \beta = 0$ или $\alpha = -\beta$. Это определяет единственный вариант линейного преобразования индекса SSIM в функцию, которая потенциально может быть метрикой (первые три аксиомы выполняются):

$$\text{NewSSIM} = -\beta \text{SSIM} + \beta,$$

или

$$\text{NewSSIM} = \beta(1 - \text{SSIM}), \quad (3)$$

где $\beta > 0$ соответственно первой аксиоме и $\min(\text{SSIM}) = -1$.

Выполнение третьей аксиомы для функции NewSSIM следует из формул (1–2). Следовательно, первые три метрических аксиомы верны как для индекса SSIM, так и для функции NewSSIM .

Что можно сказать о выполнении четвертой аксиомы? Рассмотрим очень похожие изображения одной и той же сцены из базы данных TID2013 [12]. Возьмем изображение $A = \text{'Im01-01-1'}$, изображение B равное A с одним измененным значением пикселя (вместо $A(1, 1, 1) = 73$ используем $B(1, 1, 1) = 74$) и построим изображение $C = (A + A_2) / 2$, где $A_2 = \text{'Im11-01-2'}$ – изображение из той же базы данных.

Для этих изображений должно выполняться следующее неравенство:

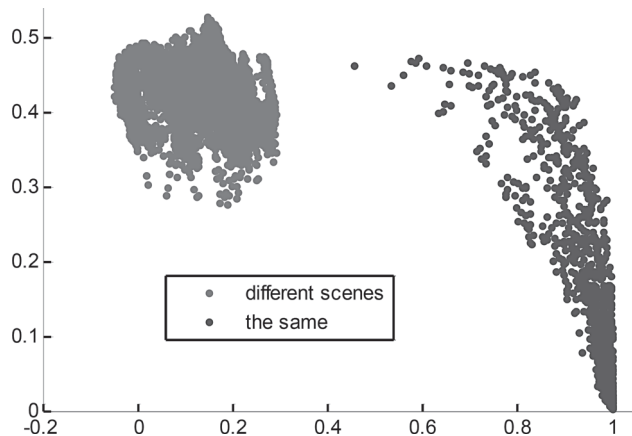
Рис. 1. Изображения A , B и C , использованные в доказательстве теоремы

Рис. 2. Сравнение значений меры NewSSIM и коэффициентов корреляции Пирсона для изображений разных сцен (облако слева) и одинаковых сцен (синие точки справа)

$$\text{NewSSIM}(A, C) \leq \text{NewSSIM}(A, B) + \text{NewSSIM}(B, C)$$

или

$$(1 - \text{SSIM}(A, C)) \leq (1 - \text{SSIM}(A, B)) + (1 - \text{SSIM}(B, C))$$

или

$$-\text{SSIM}(A, C) \leq 1 - \text{SSIM}(A, B) - \text{SSIM}(B, C). \quad (4)$$

Вычислив три значения индекса SSIM для изображений, описанных в уравнении (4), имеем

$$\begin{aligned} -0,97211954683029 &\leq \\ 1 - 0,999999989608151 - 0,972119652499616 &= \\ -0,972119642107767. \end{aligned}$$

Таким образом неравенство треугольника не выполняется для данных изображений. Этот контрпример доказывает **Теорему 1**.

Коэффициенты линейной корреляции Пирсона для пар изображений (A, B) , (A, C) , (B, C) равны 0,9999999994973, 0,9928707064627, 0,9928707094216, соответственно. Можно отметить, что коэффициенты корреляции немного выше, чем соответствующие значения индексов SSIM.

Корреляция или индекс SSIM – что лучше для оценки сходства изображений?

Мы исследовали какой параметр (линейной корреляции Пирсона или индекс SSIM) точнее оценивает сходство изображений. Для экспериментов использовали изображения из двух общедоступных баз данных TID2013 [12] и CCID2014 [13]. Использовался ПК с процессорами Intel(R), Core(TM), процессором i5-4440 с тактовой частотой 3,10 ГГц и 8,00 ГБ Ram, а также пакет Matlab 2014a. Сравнивались два набора из нескольких тысяч пар изображений одинаковых сцен и разных. Для них вычислялись значения индексов SIM и NewSSIM с параметром $\beta = 1$ в уравнении (3), а также коэффициенты корреляции Пирсона. Использовались стандартные функции Matlab corr и ssim.

Из рис. 2 видно, что диапазон значений меры NewSSIM (вертикальная ось) для пар изображений одинаковых сцен пересекается с диапазоном значений NewSSIM, вычисленных для изображений разных сцен. Используя эту меру, невозможно корректно определить сходство анализируемых изображений. Однако, используя коэффициент корреляции (горизонтальная ось), это легко сделать с помощью

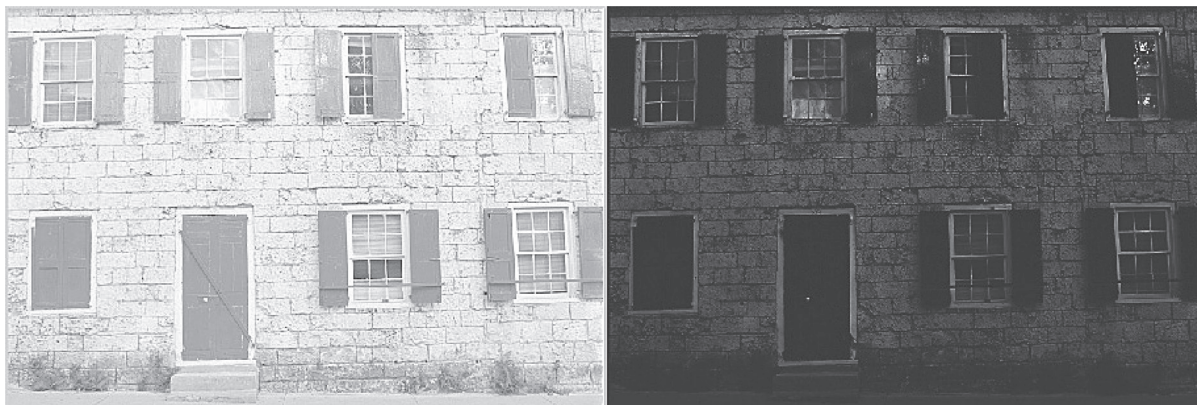


Рис. 3. Изображения img005 (слева) и img045 (справа) из базы данных CCID2014 с характеристиками: $SSIM = 0,1754$, $NewSSIM = 0,4123$ и $corr = 0,7417$



Рис. 4. Изображения img177 (слева) и img187 (справа) из базы данных CCID2014 с характеристиками: $SSIM = 0,3886$, $NewSSIM = 0,3057$ и $corr = 0,1472$

глобального порога в 0,4. Примеры сравнения изображений одной сцены и разных сцен приведены на рис. 3, 4. На рис. 3 приведен пример сравнения двух разноконтрастных изображений одной сцены. Значение индекса SSIM равно $SSIM = 0,1754$, а коэффициента корреляции равно 0,7417. На рис. 4 сравниваются два светлых изображения разных сцен. В этом случае значение индекса SSIM равно 0,3886, а коэффициента корреляции равно 0,1472. Во втором примере значение индекса SSIM гораздо ближе к единице, что означает большее сходство сравниваемых изображений, что не верно. При этом коэффициент корреляции гораздо корректнее оценивает сходство данных изображений.

Можно сделать вывод, что индексы NewSSIM и SSIM могут правильно распознавать только визуально очень близкие изображения и применимы только для оценки сходства изображений одной и той же сцены.

Для цветных изображений из базы данных TID2013 размером 768×512 пикселей истекшее время вычисления корреляции равнялось

0,030 с в Matlab (версия R2014). Время вычисления индекса SSIM на нашем компьютере составило 3,81 с, то есть примерно в 125 раз дольше.

Заключение

В статье исследованы некоторые свойства популярного индекса структурного подобия изображений под названием SSIM. В англоязычной научной литературе он считается неофициальным стандартом для оценки качества изображений при наличии эталона и часто называется метрикой качества. В данной статье продемонстрировано, что 1) индекс SSIM не может точно оценить качество изображения, а только сходство двух изображений; 2) он не является метрической функцией; 3) он не всегда правильно оценивает сходство изображенных сцен, однако коэффициент корреляции Пирсона делает это точнее и быстрее.

Работа частично выполнена в рамках проекта БРФФИ **Ф18МС-028**.

REFERENCES

1. Pedersen, M., Hardeberg, J. Y. Full-reference image quality metrics: Classification and evaluation // Foundations and Trends® in Computer Graphics and Vision, 2012), V. 7, № 1. – pp. 1–80.
2. Lin W., Kuo C. C. J. Perceptual visual quality metrics: A survey // Journal of Visual Communication and Image Representation. – 2011. – Т. 22. – №. 4. – pp. 297–312.
3. Wang, Z., Bovik, A. C. Modern image quality assessment // Synthesis Lectures on Image, Video, and Multimedia Processing, 2006, V. 2. – № 1. – pp. 1–156.
4. Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P. Image quality assessment: from error visibility to structural similarity // IEEE Trans. Image Processing, 2004, V. 13. – № 4. – pp. 600–612.
5. Chandler, D. M. Seven Challenges in Image Quality Assessment: Past, Present, and Future Research', ISRN Signal Processing, 2013, V. 2013. – C. 1–53.
6. Renieblas G. P. et al. Structural similarity index family for image quality assessment in radiological images // Journal of Medical Imaging. – 2017. – Т. 4. – №. 3. – C. 035501.
7. Sheikh H. R., Sabir M. F., Bovik A. C. A statistical evaluation of recent full reference image quality assessment algorithms // IEEE Transactions on image processing. – 2006. – Т. 15. – №. 11. – pp. 3440–3451.
8. Dosselmann R., Yang X. D. A comprehensive assessment of the structural similarity index // Signal, Image and Video Processing. – 2011. – Т. 5. – №. 1. – pp. 81–91.
9. Hore A., Ziou D. Image quality metrics: PSNR vs. SSIM // Proc. Of the 20-th Int. Conf. on Pattern Recognition, 2010. – pp. 2366–2369.
10. Pambrun J. F., Noumeir R. Limitations of the SSIM quality metric in the context of diagnostic imaging // Proc. of the IEEE International Conference on Image Processing, 2015. – pp. 2960–2963.
11. Старовойтов В. В. Уточнение индекса SSIM структурного сходства изображений // Информатика. – 2018. – Т. 15. – №. 3. – С. 41–55.
12. Ponomarenko, N, Jin, L, Ieremeiev, O, Lukin, et.al. Image database TID2013: Peculiarities, results and perspectives // Signal Processing: Image Communication, V. 30. С. 57–77.
13. Gu, K., Zhai, G., Lin, W., Liu, M. The analysis of image contrast: From quality assessment to automatic enhancement' // IEEE Transaction on Cybernetics, 2016. – V. 46. – № 1. – С. 284–297.

Поступила
04.04.2019

После доработки
20.06.2019

Принята к печати
01.07.2019

STAROVOITOV V. V.

THE SSIM INDEX IS NOT A METRIC AND IT IS BADLY EVALUATE THE SIMILITY OF IMAGES

The article explored some properties of a very popular feature of image structural similarity, called the SSIM index. According to <https://scholar.google.com>, the article [3], where it was first described, has made more than 20,800 citations during the last 14 years. This indicator is actively used by the scientific community in imaging research. It acquired the status of an unofficial international standard for assessing image quality in the presence of a template, often referred to as the image quality metric. This article debunks some of the myths that have arisen around this index. A theorem is proved which states that the SSIM index and any of its linear transformations are not metric functions. In many publications and in the Matlab application software package in the description of the ssim function, it is said that the SSIM index is used to measure the image quality. However, this index, as well as any comparison function with a reference image (such as full-reference), in principle, cannot assess the quality of the analyzed images. They estimate only a certain degree of similarity between the template image and its distorted copy. The article also shows that the SSIM index cannot always correctly determine the similarity of images of the same scene, while the Pearson linear correlation coefficient makes it much faster and more accurate.

Keywords. image quality assessment, the SSIM index, Pearson's correlation coefficient.



Старовойтов Валерий Васильевич, доктор технических наук, профессор. Главный научный сотрудник ОИПИ НАН Беларуси. Лауреат премии Ленинского комсомола БССР (1990 г.) и Государственной премии Республики Беларусь (2003 г.). Сфера научных интересов: обработка и анализ цифровых изображений, полученных в разных участках электромагнитного спектра.

Starovoitov Valery, Doctor of Sciences and professor of computer science. He is a Principal research fellow at the United Institute of Informatics Problems, National Academy of Sciences of Belarus (UIIP NAN Belarus). Awards: the Belarus Lenin Komsomol Prize and the State Prize of the Republic of Belarus in science. Research interests of professor Starovoitov are processing and analysis of digital images obtained in different parts of the electromagnetic spectrum.