Signal Processing: *Image Communication* 25 (2010) 517–526

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

Content-partitioned structural similarity index for image quality assessment

Chaofeng Li [a,*], Alan C. Bovik [b,1]

[a] *School of Information Technology, Jiangnan University, Wuxi, Jiangsu 214122, China*
[b] *Department of Electrical & Computer Engineering, The University of Texas at Austin, Austin, TX 78712-1084, USA*

ARTICLE INFO

Article history:
Received 28 April 2009
Accepted 8 March 2010

Keywords:
Four-component image model
Image quality assessment
Structural similarity (SSIM)
Multi-scale structural similarity (MS-SSIM)
Gradient structural similarity (G-SSIM)

ABSTRACT

The assessment of image quality is important in numerous image processing applications. Two prominent examples, the Structural Similarity Image (SSIM) index and Multi-scale Structural Similarity (MS-SSIM) operate under the assumption that human visual perception is highly adapted for extracting structural information from a scene. Results in large human studies have shown that these quality indices perform very well relative to other methods. However, the performance of SSIM and other Image Quality Assessment (IQA) algorithms are less effective when used to rate blurred and noisy images. We address this defect by considering a four-component image model that classifies image local regions according to edge and smoothness properties. In our approach, SSIM scores are weighted by region type, leading to modified versions of (G-)SSIM and MS-(G-)SSIM, called *four-component* (G-)SSIM (4-(G-)SSIM) and *four-component* MS-(G-)SSIM (4-MS-(G-)SSIM). Our experimental results show that our new approach provides results that are highly consistent with human subjective judgment of the quality of blurred and noisy images, and also deliver better overall performance than (G-)SSIM and MS-(G-)SSIM on the LIVE Image Quality Assessment Database.

1. Introduction

Visual images are the most important and data-intensive means for humans to acquire information, and digital image acquisition, communication, storage processing, and display devices have become ubiquitous in daily life. Since digital images are subject to a wide variety of distortions in any of these, and since image traffic has become quite dense, the assessment of digital image quality has become an exceedingly important topic.

Subjective tests have long been considered to be the Final Arbiter of image quality, but such tests are quite time consuming and expensive, and cannot be implemented in systems where real-sensitive quality scoring is needed. Therefore, nowadays there has been an increasing push to develop objective measurement techniques that predict subjective image/video quality automatically. Over the years, numerous objective metrics have been proposed [1–21] to assess image quality. These include many early algorithms that sensibly sought to incorporate perceptual models [10–14,17,19] and distortion models [16,20]; more recent algorithms based on the successful Structural Similarity Image (SSIM) index [7–9] that incorporate saliency [1], compression specificity [2], multi-scales [8], amount and local information [15], and wavelet-domain processing [21], and algorithms using information-theoretic [5], singular-value decomposition [4], the mahalanobis distance [18], and color [20]. A successful visual signal-to-noise ratio (VSNR) has been

* Corresponding author.
E-mail addresses: wxlichaofeng@yahoo.com.cn (C. Li), bovik@ece.utexas.edu (A.C. Bovik).
[1] Tel.: 1 512 471 2887.

0923-5965/$ - see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2010.03.004

proposed for IQA in [3], and the same group has also studied the interesting question of the relationship between image quality assessment with image usability in subsequent image analysis tasks [6].

The simplest and most widely used image quality indices remain the mean square error (MSE), computed by averaging the squared intensity differences of distorted and reference image pixels, and the related peak signal-to-noise ratio (PSNR). However, the MSE and its variants poorly correlate with human visual perception of quality [22–25,34]. There are a number of perceptual factors [26] that influence human perception of visual quality. Successfully incorporating these into objective IQA metrics can lead to improved correlations with visual perception [10].

Two prominent examples, the Structural Similarity Image index (SSIM) [7] and the Multi-scale Structural Similarity index (MS-SSIM) [8] operate under the assumption that visual perception is highly adapted for extracting structural information from a scene. Results in large human studies have shown that these quality indices perform quite well relative against human subjectivity. However, the performance of SSIM and MS-SSIM has been observed to be less competitive when used to assess blurred and noisy images.

It is intuitively obvious that different image regions have different importance for vision perception. Broadly, this suggests that *image content analysis* should be a primary direction of inquiry for improving quality assessment algorithms. Neither SSIM nor MS-SSIM takes into account factors such as the visual importance of image features. A few researchers have explored the possibility of improving the performance of the SSIM indices, by assigning visual importance weights to the SSIM values. In [27], the effect of using different pooling strategies was evaluated. The authors suggest that good performance can be achieved using an information-theoretic approach using "information content-weighted pooling." In [28], a SSIM modification is introduced wherein the SSIM values are weighted by a perceptual importance function, but with desultory results. The authors of [1] deepen this direction of inquiry by using human fixation data to modify the SSIM and MS-SSIM indices, gaining somewhat higher correlations with human subjectivity.

In this paper, we instead take a low-level approach to the problem, where we seek to parse the image to be assessed into regions of different low-level contents. Specifically, we use a popular four-component image model to parse the reference and distorted images, and derive region-weighted versions of SSIM and MS-SSIM, which we call *four-component SSIM* (4-SSIM) and *four-component MS-SSIM* (4-MS-SSIM), respectively. We find that we are able to improve upon the performance of SSIM and MS-SSIM against human subjectivity on the LIVE Image Quality Database.

This paper is organized as follows. Section 2 reviews SSIM, MS-SSIM and a gradient-based SSIM. The 4-SSIM (and 4-MS-SSIM) indices are described in Section 3. Section 4 gives experimental results and comparisons. Finally, future directions are considered in Section 5.

2. Structural similarity indices

2.1. Single-scale SSIM

Based on the assumption that the HVS is highly adapted to extract structural information from the viewing field, a new philosophy for image quality measurement SSIM was proposed by Wang et al. [7]. It defines the separate functional measures of luminance, contrast and structural similarity between two signals x and y:

$$l(x,y)=\frac{2\mu_x\mu_y+C_1}{\mu_x^2+\mu_y^2+C_1} \tag{1}$$

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

$$s(x,y)=\frac{\sigma_{xy}+C_3}{\sigma_x\sigma_y+C_3} \tag{3}$$

where μ_x and μ_y are the (local) sample means of x and y, respectively, σ_x and σ_y the (local) sample standard deviations of x and y, respectively, and σ_{xy} is the (local) sample correlation coefficient between x and y. Generally, these local sample statistics are computed within overlapping windows, and weighted within each window, e.g., by a Gaussian-like profile. The small constants C_1, C_2, C_3 stabilize the computations of Eqs. (1)–(3) when the denominator(s) become small.

Combining the three comparison functions of Eqs. (1)–(3) yields a general form of the SSIM index:

$$SSIM(x,y)=[l(x,y)]^{\alpha}[c(x,y)]^{\beta}[s(x,y)]^{\gamma} \tag{4}$$

where α, β and γ are parameters that mediate the relative importance of the three components. Usually, $\alpha=\beta=\gamma=1$, $C_2=C_3$, yielding the now-familiar specific form of the SSIM index

$$SSIM(x,y)=\frac{(2\mu_x\mu_y+C_1)(2\sigma_{xy}+C_2)}{(\mu_x^2+\mu_y^2+C_1)(\sigma_x^2+\sigma_y^2+C_2)} \tag{5}$$

In [7], the SSIM index is deployed using an 11×11 sliding window over the entire image space. At each image coordinate, the SSIM index is calculated within the local window; the resulting SSIM index map can be used to visualize the quality map of the distorted images. Finally, the SSIM index values can be spatially pooled, e.g., by taking their sample mean, yielding a single descriptor of the image objective quality.

2.2. Multi-scale SSIM

Wang et al. [8] developed a multi-scale SSIM (MS-SSIM) index. In MS-SSIM, quality assessment is accomplished over multiple scales of the reference and distorted image patches by iteratively low-pass filtering and down-sampling the signals (Fig. 1).

Processing in MS-SSIM is simple: index the original image as Scale 1, the first down-sampled version as Scale 2, and so on. The highest scale M is obtained after $M-1$ iterations. At the jth scale, the contrast comparison (Eq. (2)) and the structure comparison (Eq. (3)) are calculated and denoted as $c_j(x,y)$ and $s_j(x,y)$, respectively.

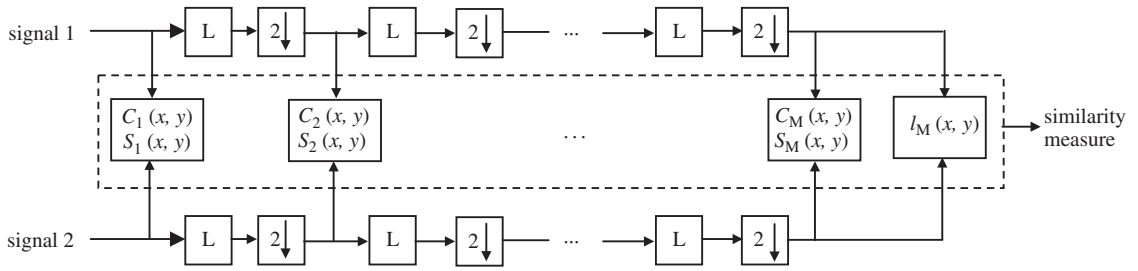

Fig. 1. Multi-scale structural similarity measurement. L denotes low-pass filtering; 2↓ denotes down-sampling by 2.

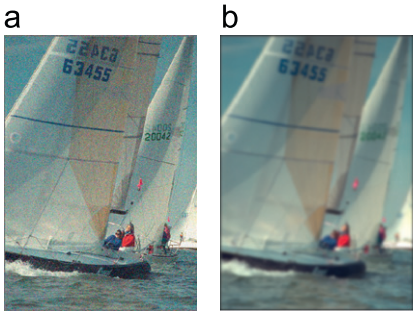

Image	DMOS	SSIM	G-SSIM	MS-SSIM	MS-G-SSIM	4-SSIM	4-G-SSIM	4-MS-SSIM	4-MS-G-SSIM
(a) noise	**44.7534**	0.3491	0.2623	0.8539	0.7292	0.50	**0.3440**	**0.9122**	**0.7856**
(b) blurred	**66.3322**	0.6932	0.3217	0.8825	0.6802	0.5605	**0.2075**	**0.8566**	**0.6384**

Fig. 2. Comparison of IQA algorithms on blurred and noisy images.

The luminance comparison (Eq. (1)) is computed only at Scale M and denoted as $l_M(x, y)$.

The overall MS-SSIM index is obtained by combining the measurement across scales via

$$MS{-}SSIM(x,y) = l_M(x,y) \prod_{j=1}^{M} c_j(x,y) s_j(x,y) \tag{6}$$

In the implementations described here, the highest scale used is $M=5$.

MS-SSIM is able to capture distortions as they occur across scales, as well as better matching the human visual response. MS-SSIM significantly outperforms single-scale SSIM.

2.3. Gradient-based SSIM (G-SSIM) and multi-scale G-SSIM

Chen et al. [29] developed an improved SSIM algorithm, called Gradient-based Structural Similarity (G-SSIM), which compares edge information between the distorted image and the original image. G-SSIM replaces the contrast comparison $c(x, y)$ in Eq. (2), and the structure comparison $s(x, y)$ in Eq. (3), with a gradient contrast comparison $c_g(x, y)$ and a gradient structure comparison $s_g(x, y)$, respectively. The gradient is generated using the Sobel operator.

Let X' and Y' denote the gradient maps of the original and the distorted images, respectively, and let x' and y', respectively, be block vectors from X' and Y'. Then:

$$c_g(x,y) = \frac{2\sigma_{x'}\sigma_{y'} + C_2}{\sigma_{x'}^2 + \sigma_{y'}^2 + C_2} \tag{7}$$

$$s_g(x,y) = \frac{\sigma_{x'y'} + C_3}{\sigma_{x'}\sigma_{y'} + C_3} \tag{8}$$

where the sample statistics have the same definition as in Eqs. (1)–(3), but applied on the gradient images. Subsequently setting $C_2=C_3$, G-SSIM is given by

$$G{-}SSIM(x,y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{x'y'} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_{x'}^2 + \sigma_{y'}^2 + C_2)} \tag{9}$$

Applying the identical multi-scale method described in Section 2.2 to G-SSIM results in multi-scale G-SSIM (MS-G-SSIM).

Extensive experimental results have shown that the SSIM and MS-SSIM indices correlate with the perception of visual quality much better than does the PSNR, and remain highly competitive with respect to other IQA algorithms [7,8,30]. However, these studies also suggest that performance might be improved when assessing the quality of blurred and noisy images, as also shown in Fig. 2. Most observers would agree that Fig. 2(a) is of better quality than Fig. 2(b), yet SSIM (G-SSIM) and MS-SSIM (MS-G-SSIM) give contrary assessment results. By comparison, the 4-G-SSIM, 4-MS-SSIM and 4-MS-G-SSIM algorithms, which we will describe, yield

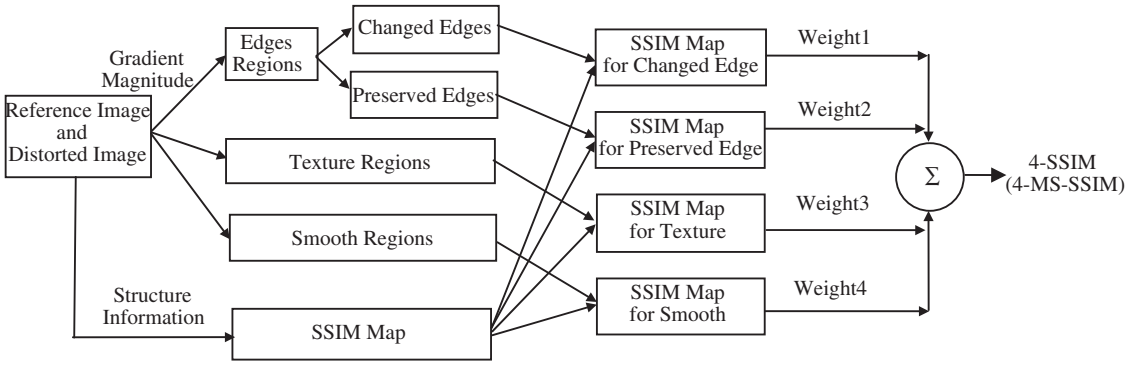

Fig. 3. Diagram for calculating 4-SSIM (or 4-MS-SSIM).

assessments that appear to better agree with visual perception of quality.

3. Modifying SSIM: 4-SSIM and 4-MS-SSIM

IQA algorithms generally operate without attempting to take into account image content. Since algorithms for image content identification remain in a nascent state, IQA algorithms that succeed in assessing quality as a function of content will await developments in that direction. However, low-level content of visual importance, sometimes called salient image features, might be used to improve IQA algorithms. For example, intensity edges certainly contain considerable image information, and are perceptually significant [31]. Using this observation we incorporate a four-component image model into SSIM (or MS-SSIM), and thereby develop four-component weighted SSIM (4-SSIM) and MS-SSIM (4-MS-SSIM) indices. We have also developed four-component indices using G-SSIM (and MS-G-SSIM), which we term the 4-G-SSIM (and 4-MS-G-SSIM) indices.

The development of 4-SSIM and 4-MS-SSIM follows four steps: (1) Calculate the SSIM (or MS-SSIM) map. (2) Independent of the SSIM results, segment the original (reference and distorted) image into four categories of regions (changed edges, preserved edges, textures and smooth regions). Changed and preserved edge regions are found where a gradient magnitude estimate is large, while smooth regions are determined where the gradient magnitude estimate is small. Textured regions are taken to fall between these two thresholds. (3) Apply non-uniform weights to the SSIM (or MS-SSIM) values over the four regions. (4) Pool the weighted SSIM (or MS-SSIM) values, e.g., their weighted average, thus defining a single quality index for the image (4-SSIM or 4-MS-SSIM). A diagram depicting calculation of 4-SSIM (or 4-MS-SSIM) is shown in Fig. 3.

3.1. Definition of 4-SSIM and 4-MS-SSIM

In the 4-SSIM algorithms, an image is partitioned into four parts: changed edges (these include edge pixels that exist in the reference image (or distorted image) but have disappeared from the distorted image (or reference image), preserved edges (these edge pixels coincide in reference and distorted images), textures and smooth regions. We seek to more heavily weight the degradations at the preserved and changed intensity edges in the IQA process. Textured regions, by contrast, often mask degradations. Artifacts in smooth regions may be quite obvious, especially if they are high-frequency or edge-like.

We may partition an image into four components using the computed gradient magnitude. Ref. [32] gives a simple image three-component partition method that we modify it to obtain a partition into four types of regions. The following steps explain the process:

Step 1: Compute the gradient magnitudes using a Sobel operator on the original and the distorted images.

Step 2: Determine thresholds $TH_1=(0.12)g_{max}$ and $TH_2=(0.06)g_{max}$, where g_{max} is the maximum gradient magnitude value computed over the original image. The threshold values 0.12 and 0.06 affect the image component partition in the following way: the smaller the first value, the more "edgey" the region is. The smaller the second value, the less the smooth region is.

Step 3: Assign pixels as belonging to changed edge, preserved edge, texture, and smooth regions, as follows:

Denoting the gradient at coordinate (i,j) on the original image by $p_0(i,j)$ and the gradient on the distorted image as $p_d(i,j)$, the pixel classification is carried out according to the following rules:

R_1: If $p_0(i,j) > TH_1$ and $p_d(i,j) > TH_1$, then the pixel is considered as a preserved edge pixel.

R_2: If $(p_0(i,j) > TH_1$ and $p_d(i,j) \le TH_1)$ or $(p_d(i,j) > TH_1$ and $p_0(i,j) \le TH_1)$, then the pixel is considered as a changed edge pixel.

R_3: If $p_0(i,j) < TH_2$ and $p_d(i,j) > TH_1$, then the pixel is regarded as part of a smooth region.

R_4: Otherwise, the pixel is regarded as part of a textured region but is not an edge pixel.

Fig. 4 shows a reference image and a compressed version of it. Also shown are the identified changed and preserved edges, the identified smooth regions, and the identified textured regions.

3.2. Determining the weights

Edges play an important role in the perception of images, and edges that are distorted, e.g., by blur, can

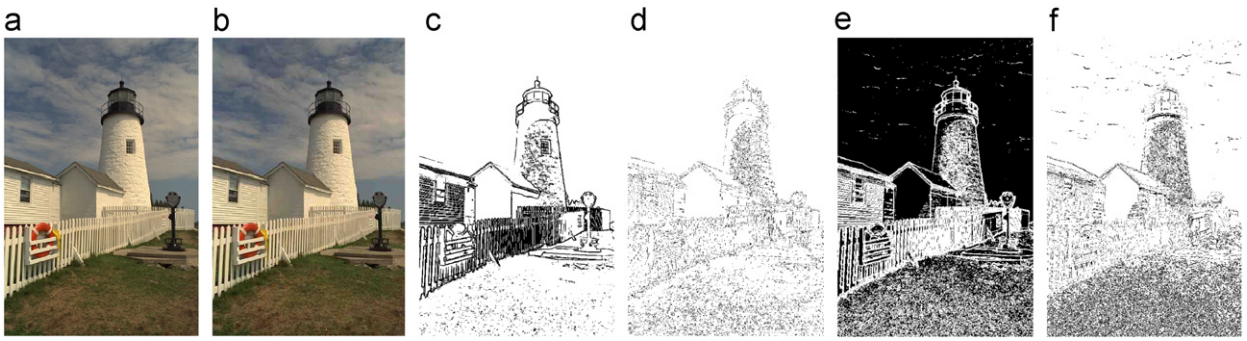

Fig. 4. From left-to-right: (a) original reference image; (b) JPEG compressed image; (c) preserved edge pixel image; (d) changed edge pixel image; (e) smooth pixel image and (f) texture pixel image. In the latter four cases black indicates membership in the classes preserved edge, changed edge, smooth, and texture, respectively.

	(b) Gaussian noise	(c) Speckle noise	(d) Salt-pepper noise	(e) JPEG compressed	(f) Blurred
SSIM	0.64	0.64	0.64	0.64	0.64
4-SSIM	0.74	0.72	0.61	0.59	0.49

Fig. 5. Evaluation of "Lena" image contaminated by different distortions: (a) Original "Lena" image, 512×512, 8 bits/pixel; (b) Gaussian noise contaminated image; (c) speckle noise contaminated image; (d) salt–pepper noise contaminated image; (e) JPEG compressed image and (f) blurred image.

greatly impact the perceived quality of an image. We therefore modify the SSIM and MS-SSIM indices by allocating greater weight to the scores at edge regions than on smooth and textured regions. To keep things simple, we fixed the weight for all edge regions at 0.5, 0.25 for changed edges and preserved edge regions, respectively. When no changed edge region is found (when the distorted image is equal to the reference image) or no preserved edge region (for heavily distortions) exists, the weight 0.5 is set for all edge regions.

Smooth regions are important also, and the eye is sensitive to artifacts such as false contouring, blocking, and high-frequency noise in smooth regions. While the distortion of textures can also be perceptually significant, some distortions can be obscured or masked by the presence of textures. As a simple approach, we apply the same weights on smooth and texture regions, fixing both at 0.25 (hence the sum of all weights is unity).

4. Experimental results and discussion

4.1. Comparisons on 5 distorted images having similar SSIM values

As a first example, we used the reference "Lena" image (shown in Fig. 5(a)) to produce 5 distorted images having nearly identical SSIM values, as shown in Fig. 5(b–f). The human subjective impression of the quality of these

	(a) Salt-pepper noise	(b) Speckle noise	(c) Gaussian noise	(d) JPEG compressed	(e) Blurred
MS-SSIM	0.66	0.66	0.66	0.66	0.66
4-MS-SSIM	0.72	0.71	0.71	0.68	0.66

Fig. 6. Evaluation of "Lena" image contaminated by different distortions: (a) Salt–pepper noise contaminated image, (b) speckle noise contaminated image; (c) Gaussian noise contaminated image; (d) JPEG2000 compressed image and (e) blurred image.

Table 1
Spearman rank order correlation coefficient (SROCC)

	JPEG2000	JPEG	WN	GBlur	FF	All data
PSNR	0.8898	0.8409	0.9853	0.7816	0.8903	0.9092
SSIM	0.9317	0.9028	0.9629	0.8942	0.9411	0.9250
4-SSIM	**0.9549**	0.9041	**0.9867**	0.9681	**0.9635**	0.9460
G-SSIM	0.9326	0.9038	0.9367	0.9364	0.9451	0.9448
4-G-SSIM	0.9516	0.9061	0.9736	0.9690	**0.9620**	**0.9626**
MS-SSIM	0.9536	**0.9108**	0.9780	0.9539	0.9350	0.9532
4-MS-SSIM	**0.9551**	**0.9109**	**0.9883**	**0.9693**	0.9416	0.9593
MS-G-SSIM	0.9311	0.8804	0.9447	0.9692	0.9381	0.9514
4-MS-G-SSIM	0.9434	0.8969	0.9734	**0.9704**	0.9440	**0.9626**
VIF [5]	**0.9563**	**0.9093**	0.9854	0.9678	0.8662	0.9559
VSNR[3]	0.946	0.908	0.979	0.941	0.906	N/A

Table 2
Linear correlation coefficient (LCC) after nonlinear regression.

	JPEG2000	JPEG	WN	Gblur	FF	All data
PSNR	0.8878	0.8596	0.9862	0.7840	0.8752	0.9293
SSIM	0.9368	0.9297	0.9793	0.8741	0.9452	0.9388
4-SSIM	**0.9601**	0.9321	0.9770	0.9685	**0.9728**	0.9489
G-SSIM	0.9382	0.9343	0.9537	0.9076	0.9479	0.9563
4-G-SSIM	0.9580	0.9416	0.9835	0.9618	0.9584	**0.9710**
MS-SSIM	0.9578	**0.9426**	0.9860	0.9579	0.9346	0.9435
4-MS-SSIM	**0.9602**	**0.9420**	**0.9905**	0.9723	0.9394	0.9440
MS-G-SSIM	0.9378	0.9137	0.9669	0.9731	0.9445	0.9492
4-MS-G-SSIM	0.9510	0.9316	0.9836	**0.9755**	0.9479	0.9555
VIF [5]	**0.9633**	**0.9422**	0.9887	**0.9737**	0.8828	0.9579
VSNR [3]	0.953	**0.943**	0.978	0.934	0.902	N/A

images is markedly different, yet the simple SSIM index does not distinguish them. Conversely, the 4-SSIM scores shown in the caption do distinguish the distorted images, and in a manner that *appears* coincident with human perception of quality. If this proves to be verifiable, then it represents an apparent advantage of four-component weighted SSIM and of the idea of using low-level content in IQA algorithms.

4.2. Comparisons on 5 distorted images having similar MS-SSIM values

We also used the reference "Lena" image (Fig. 5(a)) to produce 5 distorted images having very similar MS-SSIM scores relative to the reference image, as shown in Fig. 6(a–e). Again, these 5 distorted images have apparently different image qualities, yet even the highly regarded MS-SSIM does not distinguish them effectively. In this case, the 4-MS-SSIM index delivers assessment results that appear much more consistent with the human impression of quality. When judging these results, it is important to compare the relative quality scores of each algorithm with itself, rather than between algorithms. Broadly, one seeks a monotonic function of computed quality against a scale of perceptual quality, which is more important than the absolute separation of the scores.

4.3. Experimental comparison on LIVE image database

In order to further provide more extensive and convincing statistical quantitative comparisons, we evaluated the performances of 4-SSIM, 4-MS-SSIM, 4-G-SSIM, 4-MS-G-SSIM, SSIM, G-SSIM, MS-SSIM, MS-G-SSIM and PSNR on the LIVE Image Quality Assessment Database [33]. This database includes five types of distorted images: JPEG2000: 227 images, JPEG: 233 images, white noise (WN): 174 images, Gaussian blur (Gblur): 174 images and Fastfading (FF) noise: 174 images. Difference Mean Opinion Scores (DMOS) are also available for each distorted image.

Three performance metrics were used to evaluate the algorithms. The first is the Spearman rank order correlation coefficient (SROCC). The second is the linear correlation coefficient (LCC) between DMOS and the algorithm scores following nonlinear regression. The third is the RMSE (after nonlinear regression). The nonlinearity chosen for regression was a five-parameter logistic function [30].

Table 3
RMSE after nonlinear regression.

	JPEG2000	JPEG	WN	Gblur	FF	All data
PSNR	7.4764	8.1925	2.6493	9.7945	7.9847	8.1096
SSIM	5.6861	5.9073	3.2403	7.6636	5.3866	7.9653
4-SSIM	4.5448	5.8079	3.4191	3.9301	**3.8205**	7.3012
G-SSIM	5.6240	5.7159	4.8198	6.6222	5.2606	6.7652
4-G-SSIM	4.6575	5.3989	2.9006	4.3171	4.7102	**5.5310**
MS-SSIM	4.6705	**5.3530**	2.6711	4.5281	5.8715	7.6657
4-MS-SSIM	**4.5391**	5.3832	**2.2003**	3.6908	5.6577	7.6319
MS-G-SSIM	5.6407	6.5169	4.0886	3.6363	5.4229	7.2796
4-MS-G-SSIM	5.0256	5.8275	2.8919	**3.4679**	**5.2568**	6.8220
VIF [5]	**4.3634**	**5.3724**	2.3977	3.5946	7.7529	6.6406
VSNR [3]	4.963	**5.339**	3.339	5.692	7.193	N/A

4.3.1. Results on entire LIVE database

The results, presented in Tables 1–3, are reported for the different distortion types as well as on the entire dataset. In the experimental results of the 5 independent datasets, no reference image is included in the presentations to human subjects. In the test results overall data, the hidden reference images are included. This follows the reporting of the LIVE study and other publications.

As can be seen, 4-SSIM outperforms SSIM and PSNR on all types of distorted images, and also performs better

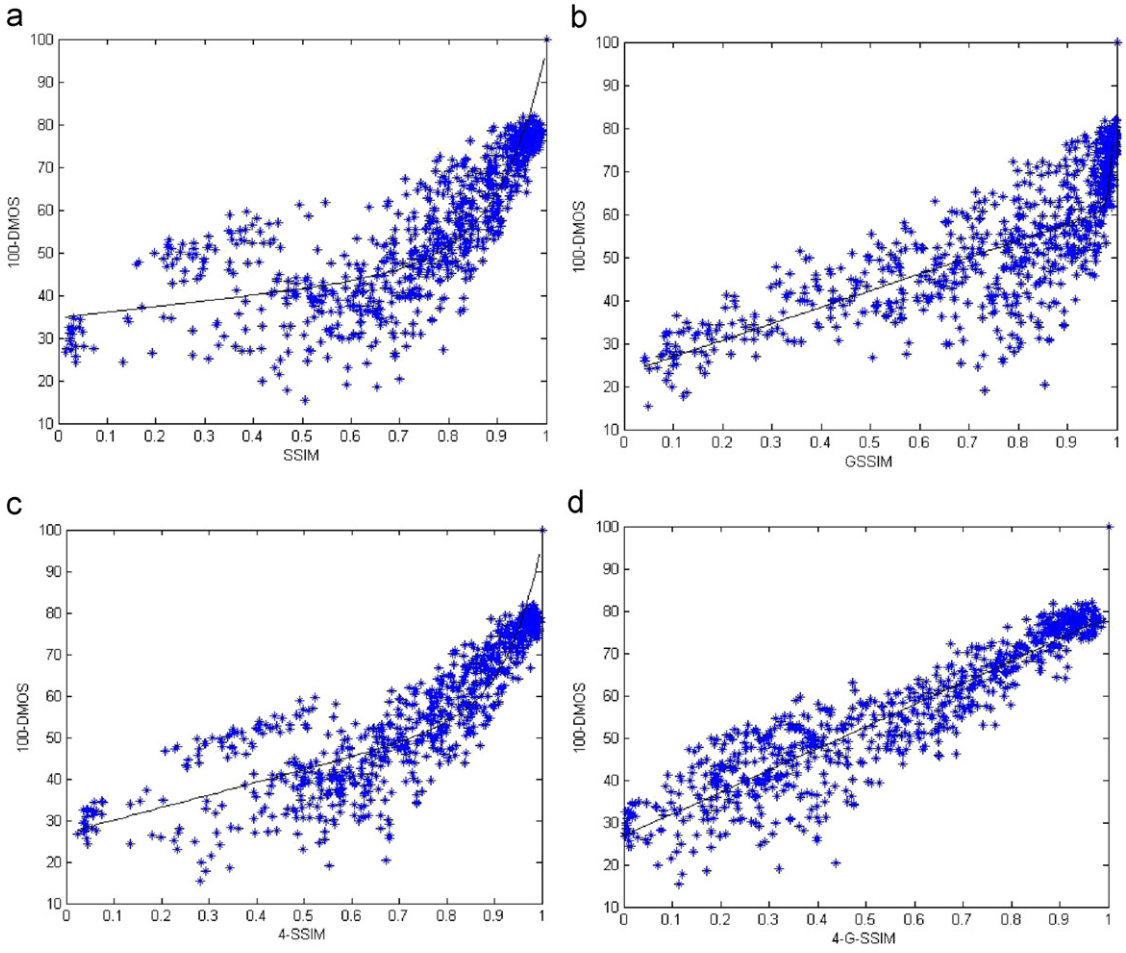

Fig. 7. Scatter plots for DMOS versus single-scale QA score for all images.

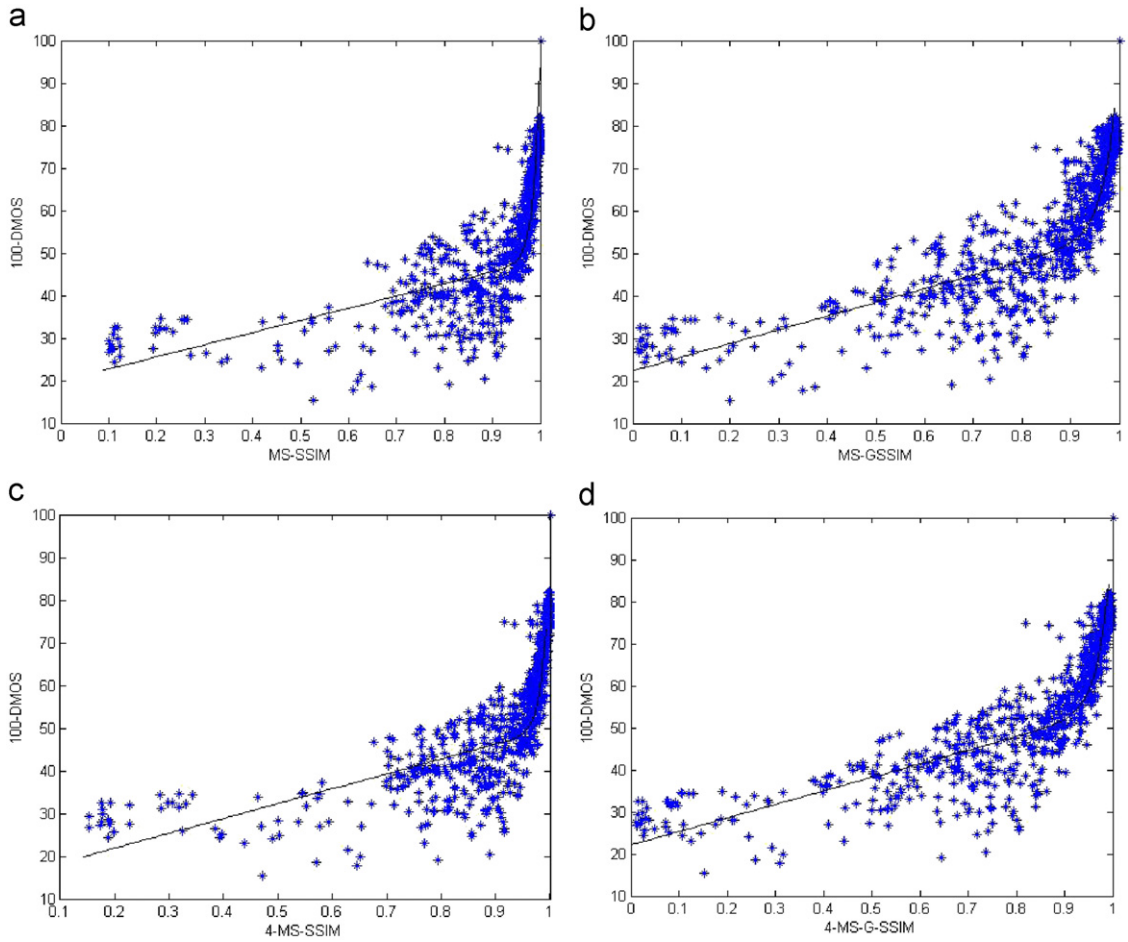

Fig. 8. Scatter plots for DMOS versus multi-scale QA scores for all images.

than MS-SSIM on JPEG2000, WN, GBlur and FF distortions. Overall, 4-MS-SSIM outperforms MS-SSIM. G-SSIM outperforms SSIM, and 4-G-SSIM far outperforms G-SSIM. Interestingly, G-SSIM is not as effective in multi-scale form, and MS-G-SSIM proved somewhat inferior to MS-SSIM. Yet the numbers are so close as to be statistically insignificant. However, the four-component weighted method effective improves MS-G-SSIM, yielding 4-MS-G-SSIM, which overall outperforms MS-G-SSIM in whole dataset. According to the three performance metrics overall data, 4-G-SSIM yields the best IQA performance on the LIVE database. Figs. 7 and 8 depict scatter plots of DMOS versus all above-described SSIM and MS-SSIM variant indices for all images, which also suggest that 4-G-SSIM has the best prediction performance relative to DMOS.

4.3.2. Results on LIVE noise and blurred database

In order to show the advantages of the four-component weight method in detail relative to the prior SSIM algorithms and their variants, we also studied the experimental results to rate algorithm performance specifically on blurred and noisy images, as shown in Table 4. As can be seen, four-component weighted 4-SSIM (4-G-SSIM, 4-MS-SSIM, 4-MS-G-SSIM) *greatly* improves the performance of each SSIM index (G-SSIM, MS-SSIM, MS-G-SSIM), which makes them more competitive, especially in applications where images are noisy or blurred, relative to other IQA methods.

4.3.3. The discussion of weight

In any case, in order to explore the relevance of the various weights of 4-SSIM, we varied the weight for edge regions from 0 to 1.0, in steps of 0.1. To simplify things, we applied the same weight to changed edge and preserved edge regions. Equal weights were also allocated to textured and smooth regions, with the total weight summing to unity. Using this protocol, the plot of SROCC for 4-SSIM when tested on the LIVE Image Quality Assessment Database [33], as a function of the weight on the edge regions, is shown in Fig. 9. As can be seen, when the total edge weight (changed and preserved edges) takes the value 0.6, the SROCC for 4-SSIM reaches the highest (discrete) score, suggesting that the 4-SSIM QA index should weight edge regions more heavily than other regions, but that we should not eliminate the evaluation of quality on textured and smooth regions.

Table 4
Assessment results on blurred and noisy images.

	SROCC	LCC	RMS
PSNR	0.8621	0.8483	8.4058
SSIM	0.8323	0.7472	10.5488
4-SSIM	0.9218	0.9095	6.5983
G-SSIM	0.8834	0.8691	7.8521
4-G-SSIM	**0.9594**	**0.9491**	**5.0016**
MS-SSIM	0.9097	0.9083	6.6389
4-MS-SSIM	0.9476	0.9375	5.5242
MS-G-SSIM	0.8930	0.8960	7.0484
4-MS-G-SSIM	0.9412	0.9386	5.4765

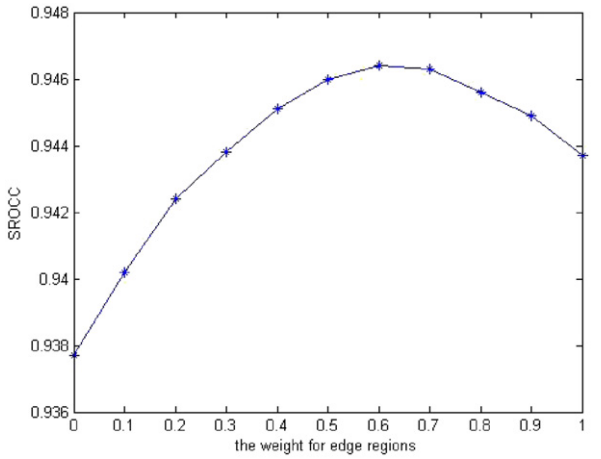

Fig. 9. Plot of SROCC of 4-SSIM against edge region weight.

5. Concluding remarks

In this paper, we presented a new four-component weighted structural similarity metric that improves upon the well-known SSIM and multi-scale SSIM indices and some of their popular derivatives. Using the idea that different image regions have different perceptual significances relative to quality, we defined different weights on SSIM scores according to a low-level content-based segmentation of images into homogeneous regions. Our experimental results show that 4-(G-)SSIM (and 4-MS-(G-)SSIM) provide results that are more consistent with human subjectivity, when comparing the quality of blurred and noisy images, and also deliver better performance than (G-)SSIM (and MS-(G-)SSIM) on five types of distorted images from the LIVE Image Quality Assessment Database. Interestingly, 4-(G-)SSIM performed better than either MS-(G-)SSIM or 4-MS-(G-)SSIM, suggesting that edge relevance might be of greater significance than multi-scale.

Acknowledgement

This research is supported by the Self-determined Research Program of Jiangnan University (No. JUSRP20915).

References

[1] A.K. Moorthy, A.C. Bovik, Visual importance pooling for image quality assessment, IEEE Journal of Selected Topics in Signal Processing 3 (2) (2009) 193–200.
[2] A.C. Brooks, X.N. Zhao, T.N. Pappas, Structural similarity quality metrics in a coding context: exploring the space of realistic distortions, IEEE Transactions on Image Processing 17 (8) (2008) 1–12.
[3] M.C. Damon, S.H. Hemami, VSNR: a wavelet-based visual signal-to-noise ratio for natural images, IEEE Transactions on Image Processing 16 (9) (2007) 2284–2298.
[4] A. Shnayderman, A. Gusev, A.M. Eskicioglu, An SVD-based grayscale image quality measure for local and global assessment, IEEE Transactions on Image Processing 15 (2) (2006) 422–429.
[5] H.R. Sheikh, A.C. Bovik, G.D.E. Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, IEEE Transactions on Image Processing 14 (12) (2005) 2117–2128.
[6] D.M. Rouse, R. Pepion, S.S. Hemami, P. Le Callet, Image utility assessment and a relationship with image quality assessment, SPIE: Human Vision and Electronic Imaging 7240 (2009) 724010–724010-14.
[7] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Transactions on Image Processing 13 (4) (2004) 600–612.
[8] Z. Wang, E.P. Simoncelli, A.C. Bovik, Multi-scale structural similarity for image quality assessment, Asilomar Conference on Signals, Systems, and Computers (2003) 1398–1402.
[9] Z. Wang, A.C. Bovik, A universal image quality index, IEEE Signal Processing Letters 9 (2002) 81–84.
[10] A.M. Eskicioglu, P.S. Fisher, Image quality measures and their performance, IEEE Transactions on Communications 43 (12) (1995) 2959–2965.
[11] S.A. Karunasekera, N.G. Kingsbury, A distortion measure for blocking artifacts in images based on human visual sensitivity, IEEE Transactions on Image Processing 4 (6) (1995) 713–724.
[12] N.B. Mill, A visual model weighted cosine transform for image compression and quality assessment, IEEE Transactions on Communications 33 (6) (1985) 551–557.
[13] J.A. Saghri, Image quality measure based on a human visual system model, Optical Engineering 28 (7) (1989) 813–818.
[14] A.P. Bradley, A wavelet visible difference predictor, IEEE Transactions on Image Processing 8 (1999) 717–730.
[15] X.B. Gao, T. Wang, J. Li, A Content-based image quality metric, Proceedings of LNAI 3642 (2005) 231–240.
[16] N. Damera-Venkata, T.D. Kite, W.S. Geisler, B.L. Evans, A.C. Bovik, Image quality assessment based on a degradation model, IEEE Transactions on Image Processing 9 (2000) 636–650.
[17] M.P. Eckert, A.P. Bradley, Perceptual quality metrics applied to still image compression, Signal Processing 70 (3) (1998) 177–200.
[18] F. Imai, N. Tsumura, Y. MIYAKE, Perceptual color difference metric for complex images based on mahalanobis distance, Journal of Electronic Imaging 10 (2001) 385–393.
[19] M. Miyahara, K. Kotani, R. Algazi, Objective picture quality scale (PQS) for image coding, IEEE Transactions on Communications 46 (9) (1996) 1215–1226.
[20] A. Toet, M. Lucassen, A new universal colour image fidelity metric, Displays 24 (2003) 197–204.
[21] Z. Wang, E. Simoncelli, Translation insensitive image similarity in complex wavelet domain, IEEE International Conference on Acoustics, Speech and Signal Processing 2 (2005) 573–576.
[22] B. Girod, What's wrong with mean-squared error, in: A.B. Watson (Ed.), Digital Images and Human Vision, MIT Press, Cambridge, MA, 1993, pp. 207–220.
[23] P.C. Teo, D.J. Heeger, Perceptual image distortion, Proceedings of SPIE 2179 (1994) 127–141.
[24] M.P. Eckert, A.P. Bradley, Perceptual quality metrics applied to still image compression, Signal Processing 70 (1998) 177–200.
[25] Z. Wang, A.C. Bovik, L. Lu, Why is image quality assessment so difficult, IEEE International Conference on Acoustics, Speech and Signal Processing 4 (2002) 3313–3316.
[26] B.A. Wandell, Foundations of Vision, Sinauer Associates Inc., 1995.
[27] Z. Wang, X. Shang, Spatial pooling strategies for perceptual image quality assessment, IEEE International Conference on Image Processing (Sep. 2006) 2945–2948.
[28] D.V. RAO, L.P. REDDY, Image Quality Assessment Based on Perceptual Structural Similarity, LNCS, 4815, Springer-Verlag, Berlin Heidelberg, 2007, pp. 87–94.

[29] G.H. CHEN, C.L. YANG, S.L. XIE, Gradient-based structural similarity for image quality assessment, IEEE International Conference on Image Processing (2006) 2929–2932.
[30] H.R. Sheikh, M.F. Sabir, A.C. Bovik, A statistical evaluation of recent full reference image quality assessment algorithms, IEEE Transactions on Image Processing 15 (11) (2006) 3441–3451.
[31] X. Ran, N. Farvardin, A perceptually-motivated three-component image model—Part I: description of the model, IEEE Transactions on Image Processing 4 (4) (1995) 401–415.
[32] J.L. Li, G. Chen, Z.R. Chi, et al., Image coding quality assessment using fuzzy integrals with a three-component image model, IEEE Transactions on Fuzzy Systems 12 (1) (2004) 99–106.
[33] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, LIVE Image Quality Assessment Database, (online) http://live.ece.utexas.edu/research/quality.
[34] Z. Wang, A.C. Bovik, Mean squared error: love it or leave it? A new look at signal fidelity measures, IEEE Signal Processing Magazine 26 (1) (2009) 98–117.