Image Quality Assessment based on Improved Structural SIMilarity *

Jinjian Wu[1], Fei Qi[2], and Guangming Shi[3]

School of Electronic Engineering, Xidian University,
Xi'an, Shaanxi, 710071, P.R. China
[1]jinjian.wu@mail.xidian.edu.cn
[2]fred.qi@ieee.org
[3]gmshi@xidian.edu.cn

Abstract. In this paper, we propose a novel image quality assessment (IQA) based on an Improved Structural SIMilarity (ISSIM) which considers the spatial distributions of image structures. The existing structural similarity (SSIM) metric, which measures structure loss based on statistical moments, i.e., the mean and variance, represents mainly the luminance change of pixels rather than describing the spatial distribution. However, the human visual system (HVS) is highly adapted to extract structures with regular spatial distributions. In this paper, we employ a self-similarity based procedure to describe the spatial distribution of image structures. Then, combining with the statistical characters, we improve the structural similarity based quality metric. Furthermore, considering the viewing condition, we extend the ISSIM metric to the multi-scale space. Experimental results demonstrate the proposed IQA metric is more consistent with the human perception than the SSIM metric.

Keywords: Image Quality Assessment, Structural Similarity, Statistical Character, Spatial distribution, Self-Similarity

1 Introduction

As a mathematical technology of the human behaviors in image quality evaluation, objective image quality assessment (IQA) metric has been widely used in various image processing application, e.g., compression, transmission and restoration [4]. The simplest and most common quality metrics are the mean square error (MSE) and the peak signal-to-noise ratio (PSNR), which directly compute the differences between the reference and distorted images. But both metrics do NOT accord with the human visual perception well, since the signal error is not equivalent to the degradation of visual quality in the human visual system (HVS).

* This work is supported by Natural Science Foundation of China under Grant NO. 60805012, 61033004, 61070138, 61227004, and 61003148.

Considering the perceptual characteristic of the HVS, Wang et al. introduced a structural similarity (SSIM) based quality metric [8]. The SSIM metric is under the assumption that the HVS is highly adapted to extract structural information from an input scene. In the SSIM metric, the image structure is represented by statistical characters, e.g., the mean and variance, and image quality is measured based on the similarity between these statistical characters. This metric imitates the human perception on image structure and returns a better assessment result (be more consistent with the HVS) than MSE and PSNR. Furthermore, Wang et al. improved the SSIM metric by taking the variations of the viewing conditions into account, and introduced a multi-scale structural similarity (MS-SSIM) based quality metric [10]. As an extension of the single-scale SSIM metric, the MS-SSIM metric further promotes the performance on image quality assessment. In [3], Li and Bovik segmented the image into three types of region, i.e., plain, edge, and texture, and gave different weights to the quality results (evaluated by the SSIM metric) of these regions. In addition, the edge structure represents the major information for visual perception and plays a crucial role in the recognition for image content [1][5]. And therefore, Liu et al. [5] improved the SSIM metric by considering the edge similarity.

Though the SSIM and MS-SSIM metrics achieve great success in subjective quality assessment, their statistic based structural descriptors, i.e., mean, variance, and covariance, are too rough to represent the complex image structure [3]. For example, the statistical variance mainly represents the luminance difference but gives little information about the spatial distribution of image structure. The HVS is not only sensitive to the luminance difference but also to the spatial distribution [11] of image structure. Therefore, a more precise structural descriptor, which can effectively represent the spatial structure, is required for much accurate quality assessment.

In this paper, we introduce a novel structure descriptor to improve the structural similarity based quality metric. Since the image structure is determined by the arrangements of and relations among pixels [11], a self-similarity based procedure, which is a valid representation of the relationship among pixels, is adopted to describe the spatial distribution of image structures. At the meanwhile, the statistical characters, which mainly represent the luminance change, is adopted to describe the luminance difference of image structures. Employing both statistical character and self-similarity to represent the luminance change and spatial distribution of the structure, respectively, a much precise structure descriptor is constructed. Then, with the novel structural descriptor, we improve the quality assessment between the reference and distorted images, which we call improved structural similarity (ISSIM) based quality metric. Furthermore, considering the variations of the viewing conditions, we extend the ISSIM metric into multi-scale space and introduce a multi-scale ISSIM (MS-ISSIM) metric. Experimental results demonstrate that the ISSIM/MS-ISSIM metric outperforms the SSIM/MS-SSIM metric.

The rest part of this paper is organized as follows. In section 2, we propose a novel IQA metric following the introduction of a precise structural descriptor based on the self-similarity of image content. Performance of the proposed ISSIM metric is evaluated with several experiments on the TID2008 database [6] in Section 3. Conclusions and discussions are given in Section 4.

2 Structural Descriptor and Quality Metric

In this section, we firstly analyze the spatial distribution of structure based on self-similarity procedure. Then combining with statistical character, a much precise structural descriptor is proposed. Finally, with the novel structural descriptor, an improved structural similarity based quality metric is introduced.

2.1 Self-Similarity and Structural Descriptor

The structural information represents the primary visual contents of the input scene, and the HVS is highly sensitive to it [8]. As Fig. 1 shows, the two original images, (a) and (c), are composed of two types of pixels. Pixels in the two images are regularly and irregularly arranged, respectively. Since the spatial distributions of pixels in the two images are different, they present different structures. When contaminated by the same Gaussian white noise, as shown in Fig. 1(b) and (d), the two images have different quality degradations.

The statistical character cannot represent precisely the spatial distributions of the structure though it is effective to describe the rough features of the structure. As the two original images, shown by Fig. 1(a) and (c), are composed with the same amount of black and white pixels, they have the same statistical values. According to the SSIM metric, the two original images are with similar structural descriptors [8]. As a result, when the two original images are contaminated by the same noise, they will acquire the same quality values based on the SSIM metric. Obviously, this result does not accord with the human perception, since the HVS is much more sensitive to noise in Fig. 1(b) which has regular structures than in Fig. 1(d) which has irregular structures. Therefore, a more precise structural descriptor is required for accurate image quality assessment.

The structure appears as the relations among pixels [11]. As shown in Fig. 1(a), pixels in this regular image are strongly correlated with their surroundings and present self-repeating structures. The HVS is highly adapted to extract the homogeneous structures, and easily find out the distortion according to the comparison among them. However, the spatial distribution of pixels in an irregular image, as shown in Fig. 1(c), is disordered, the HVS is unable to accurately predict the structure and becomes insensitive to the distortion in it. Therefore, the regularity of the structure directly determines the sensitiveness of noise, and we need to consider the regularity for structural description.

Self-similarity, which describes the intrinsic relations among pixels, is an effective description of structural regularity [2]. An image with regular structure appears strong self-similarity, e.g., Fig. 1 (a), while an irregular image presents dissimilar structures, e.g., Fig. 1 (c). In this paper, we adopt self-similarity to represent the spatial distribution of the structure.

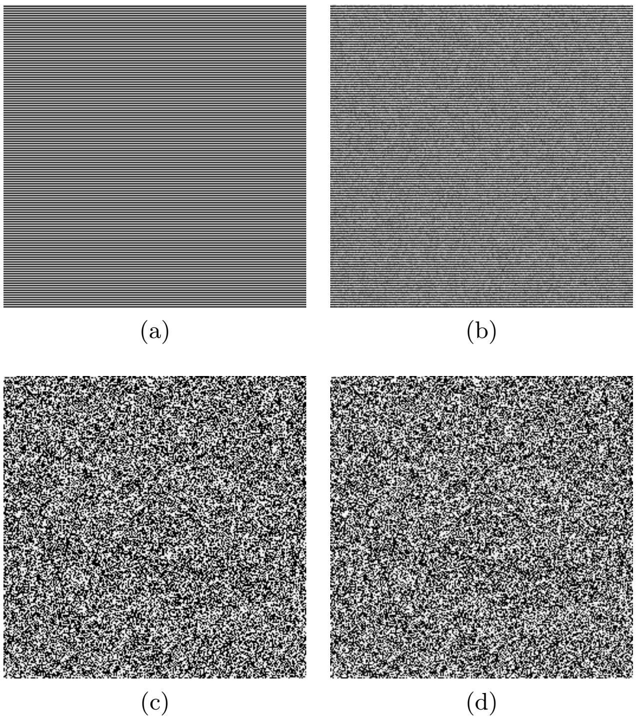


Fig. 1: Image structures have significant effects to visual quality assessment. (a) and (c) are the original images, which are composed of two types of pixels (black and white). (b) and (d) are distorted images contaminated by a same noise. Though (a) and (c) are with the same mean and variance, their structural characters are very different. The HVS is much more sensitive to the noise in (b) than that in (d).

Considering both the statistical character and self-similarity, which respectively represent the luminance difference and the spatial distribution, we introduce a precise structural descriptor. Let X be the reference image, and $\Omega(x)$ be a local region of a pixel $x \in X$. The self-similarity of a pixel x is measured by the similarity coefficients between pixel x and its surrounding, denoted by $\{d_1(x), \cdots, d_N(x)\}$. The similarity coefficient between the central pixel x and its ith neighbor y_i, with $i = 1, \cdots, N$, is computed as [11],

$$d_i(x) = \exp\left(-\frac{\left\|F(x) - F(y_i)\right\|_2^2}{2h_x^2}\right), \tag{1}$$

where $F(x)$ and $F(y_i)$ denote the vectors formed by concatenating all columns in $\Omega(x)$ and $\Omega(y_i)$, and h_x is defined as [11],

$$h_x = \begin{cases} \sigma_0 & \text{if } \sigma_x \le \sigma_0 \\ \sigma_0(\frac{\sigma_0}{\sigma_x})^{0.5} & \text{else} \end{cases}, \tag{2}$$

where σ_x is the variance of the local region $\Omega(x)$, and σ_0 is the mean variance value of the image.

2.2 Improved Structural Similarity Based Quality Metric

With the precise structure descriptor, we propose an improved structural similarity based quality metric. The SSIM metric compute the similarity on three statistical components, which are luminance similarity, contrast similarity, and structural similarity [8]. Let X' be the distorted image of X, for any two pixels $x \in X$ and $x' \in X'$, the SSIM metric is as follows [8],

$$l(x, x') = \frac{2\mu_x\mu_{x'} + C_1}{\mu_x^2 + \mu_{x'}^2 + C_1}, \tag{3}$$

$$c(x, x') = \frac{2\sigma_x\sigma_{x'} + C_2}{\sigma_x^2 + \sigma_{x'}^2 + C_2}, \tag{4}$$

$$s(x, x') = \frac{\sigma_{xx'} + C_3}{\sigma_x\sigma_{x'} + C_3}, \tag{5}$$

$$\mathrm{SSIM}(x, x') = l(x, x')\, c(x, x')\, s(x, x'), \tag{6}$$

where μ_x and $\mu_{x'}$ are the means of the local patches, which are with a size of 11×11, centered at x and x', respectively, σ_x and $\sigma_{x'}$ are the standard variance, $\sigma_{xx'}$ is the covariance of the two patches, C_1, C_2 and C_3 are small constants to make sure the denominator not being zero (Please refer to [8] for more details about the SSIM metric).

As it can be seen that (6) is based on the statistical characters, and cannot effectively represent the changes in the spatial distribution of image structures. Here, we compute the similarity of the spatial distribution based on the similar coefficients provided in (1). The similarity of the spatial distribution between the reference image X and the distorted image X' is computed as,

$$\mathrm{SD}(x, x') = \frac{1}{N} \sum_{i=1}^{N} \frac{2 d_i(x) d_i(x') + C_4}{d_i^2(x) + d_i^2(x') + C_4}, \tag{7}$$

where $C_4 = (K_1 L)^2$, L is the gray level of the image and $K_1 = 0.01$ (similar as in [8]).

Combining (6) and (7), the ISSIM metric is acquired,

$$\mathrm{ISSIM}(x, x') = \alpha\, \mathrm{SSIM}(x, x') + \beta\, \mathrm{SD}(x, x'), \tag{8}$$

where α and β are the relative importance of the two parts, and in this paper, we simply set $\alpha = \beta = 0.5$.

In addition, considering the viewing conditions, we extend the ISSIM metric into multi-scale and introduce a multi-scale ISSIM (MS-ISSIM) based quality metric. We downsample the original images into multi levels and operate the ISSIM metric on each one,

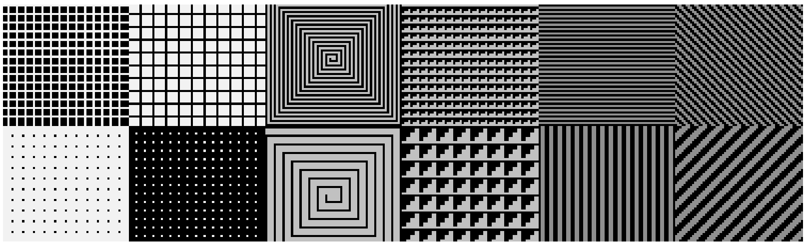
$$\text{MS-ISSIM}(x, x') = \prod_{l=1}^{M} \mathrm{ISSIM}_l^{\gamma(l)}(x, x') \tag{9}$$

where M is the highest level. In this paper, we set $M = 5$ and γ to be 0.0448, 0.2856, 0.3001, 0.2363, and 0.1333 from $l = 1$ to $l = M$, respectively, according to [10].

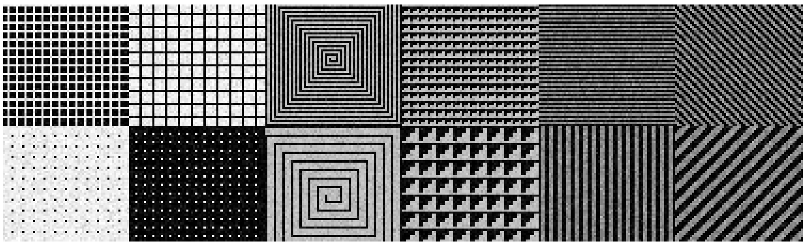
3 Experimental Results and Discussion

In this section, we firstly analyze the effectiveness of the proposed metric on images with representative structures. Then we verify the proposed metric by comparing with the SSIM/MS-SSIM metric on the TID2008 [6]. The TID2008 database contains 1700 distorted images, which are generated from 25 reference images with 17 type of distortions at 4 different noise levels. Its corresponding Mean Opinion Score (MOS) is obtained through a complicated result, which is achieved by more than 800 experiments with a large number of observers from three countries (Finland, Italy, and Ukraine).

The proposed metric is based on a more precise structural descriptor than that in the SSIM metric. With the precise structural descriptor, the character of the image can be further analyzed. For a clear view, a part of one reference image from the TID2008 database, which is composed of 12 types of structures, is chosen, as shown in Fig. 2 (a). And Fig. 2 (b) is its corresponding white noise contaminated image. Though under the same noise, these patches with different structures present different visual quality degradations. Intuitively, the HVS is highly sensitive to the distortions on these patches with regular structures. And therefore, the more regular the patch is, the more seriously its quality degrades.

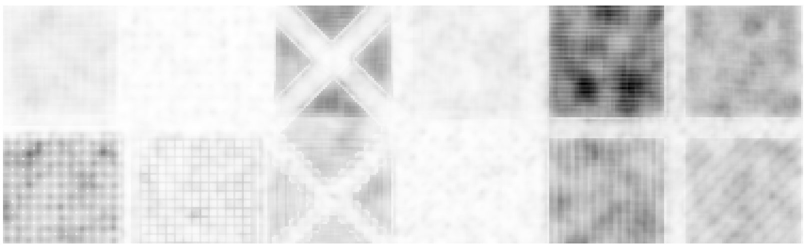


(a)



(b)

(c)



(d)

Fig. 2: IQA results comparison on a concept image. (a) Reference image. (b) distorted image. (c) SSIM based IQA result. (d) ISSIM based IQA result.

For example, the right four patches are with highly self-similar structures and the noise in them is easy to be sensed, while the two patches in the fourth column are less self-similar than their nearby patches and are much robust to noise.

The evaluation results from the SSIM metric (The SSIM code is downloaded from Wang's homepage) and the ISSIM metric on contaminated image are shown in Fig. 2 (c) and (d), respectively. In Fig. 2(c), it can be seen that the results on the right 6 patches are almost the same, which is against the perception of the HVS. The output of the ISSIM metric presents different assessment result on these 6 patches. As shown in Fig. 2 (d), patches with regular structures (i.e., the right 4 patches) have more quality loss than these irregular ones (the two patches in the fourth column). Therefore, the output of the ISSIM metric is more consistent with the human visual perception than the SSIM metric.

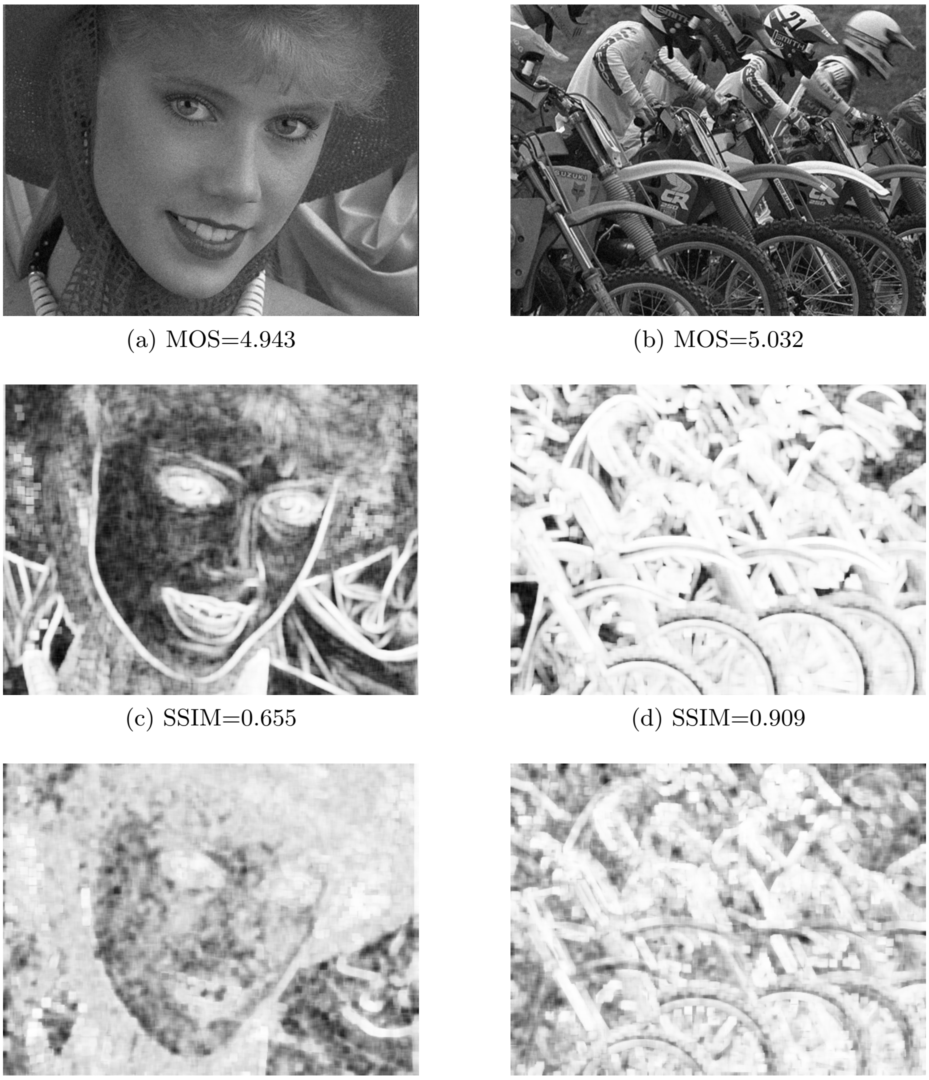
For further analyzing the performance of the proposed algorithm against the SSIM metric, two natural images are chosen for comparison. The experimental results are shown in Fig. 3, which indicates that the structural degradations of the two contaminated images are limited, and their subjective qualities, where Fig. 3(a) is MOS=4.943 and Fig. 3(b) is MOS=5.032, are very near. Since the SSIM metric only adopts the statistical character for structural analysis, as shown in Fig. 3(c), the quality degradation on the smooth region is overestimated. As a result, according to the SSIM metric, the quality of Fig. 3(a) with SSIM=0.655 is far worse than that of Fig. 3(b) with SSIM=0.909, because the former possesses a large smooth region while the latter has a smaller smooth area. Though the luminance difference of the smooth region has been changed under the distortion, the self-similarity is almost unchanged and the structure degradation is small according to the HVS. With the ISSIM metric, the structural similarity is computed on both the luminance change and spatial distribution. The outputs of the ISSIM metric on the two images, Fig. 3(a) with ISSIM=0.992 and Fig. 3(b) with ISSIM=0.997, are much similar. Therefore, the ISSIM metric performs more consistently with the HVS than the SSIM metric does.

For a comprehensive analysis, we make a comparison between the SSIM metric and the ISSIM metric over the whole TID2008 database. To evaluate the performance of the two metrics on a common space during our experiment, we firstly employed a five-parameter mapping function [7] to nonlinearly regress the computational quality scores S_0,

$$S_r = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2 (S_0 - \beta_3))} \right) + \beta_4 \, S_0 + \beta_5, \tag{10}$$

where β_j, $j = 1, \cdots, 5$, are the parameters to be fitted. The scatter plots of the computational score from the four metrics, i.e., SSIM, ISSIM, MS-SSIM, and MS-ISSIM, versus the mean opinion score (MOS) on the whole TID2008 database are shown in Fig. 4.

Then, five criteria are adopted [9] for result evaluation. The criteria are Spearman Rank-order Correlation Coefficient (SRCC), Kendall Rank-order Correlation Coefficient (KRCC), Pearson Linear Correlation Coefficient (PLCC), Mean Absolute Error (MAE), and Root Mean-Squared Error (RMSE). A better IQA



(a) MOS=4.943 (b) MOS=5.032

(c) SSIM=0.655 (d) SSIM=0.909

(e) ISSIM=0.992 (f) ISSIM=0.997

Fig. 3: IQA results comparison on natural images. (a) and (b) White noise contaminated images. (c) and (d) Outputs of the SSIM metric. (e) and (f) Outputs of the ISSIM metric. The subjective qualities (represented by MOS values) are very similar and the ISSIM metric coincides with this, while the SSIM metric shows that (a) is far worse than (b).

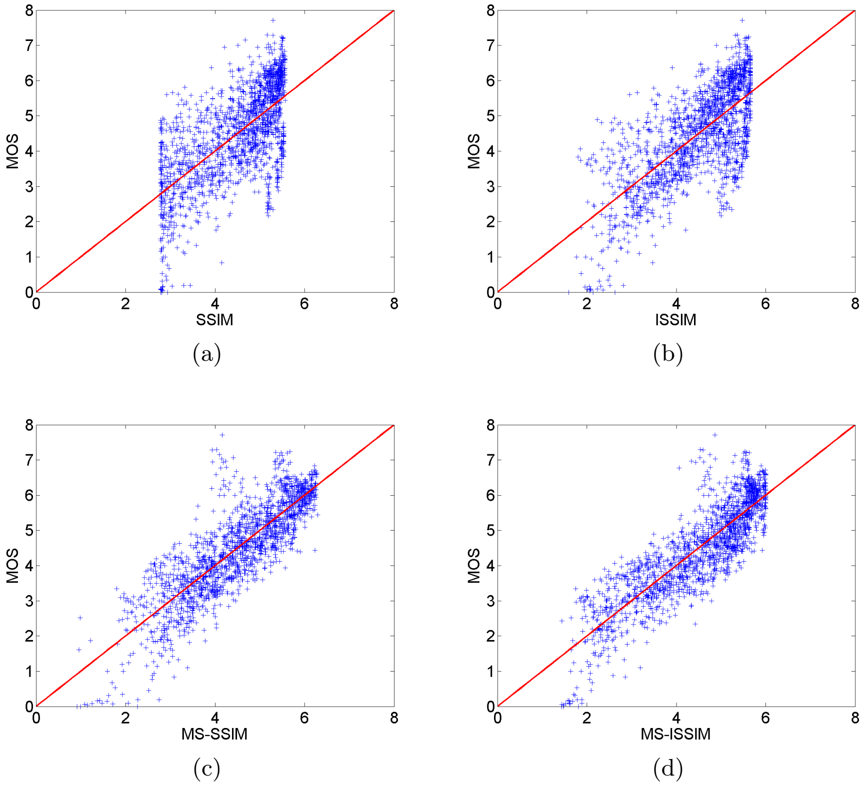


Fig. 4: Scatter plots of subject scores vs. the computational scores (mapping scores of the SSIM, ISSIM, MS-SSIM, and MS-ISSIM) for the TID2008 database.

metric should have higher SRCC, KRCC, and PLCC, while lower MAE and RMSE values. As listed in Table 1, the ISSIM metric has higher SRCC, KRCC and PLCC values, and lower MAE and RMSE values than the SSIM metric. Therefore, the ISSIM metric has an obvious improvement to the SSIM metric. Meanwhile, the MS-ISSIM metric also performs better on all the five evaluation criteria than the MS-SSIM metric, which further confirms that the proposed structural descriptor is more accurate and effective than that employed by the SSIM metric.

Table 1: Performance comparisons of image quality assessment algorithms on TID2008 database.

Criteria	SSIM	ISSIM	MS-SSIM	MS-ISSIM
SRCC	0.641	**0.677**	0.850	**0.867**
KRCC	0.467	**0.499**	0.657	**0.673**
PLCC	0.643	**0.705**	0.782	**0.806**
MAE	0.831	**0.771**	0.669	**0.630**
RMSE	1.027	**0.951**	0.836	**0.794**

4 Conclusion

In this paper, an improved structural similarity based image quality assessment is proposed. Existing statistical characters based structural descriptor mainly represents the luminance change of the structure while cannot effectively represent the spatial distribution. Since the HVS is highly sensitive to the spatial distribution of image structures, we adopted self-similarity to describe structural character in detail. And then, combining both the luminance change and spatial distribution of the structure, an accurate structural descriptor is introduced. According to the novel structural descriptor, we improve the structural similarity based image quality assessment and introduce an ISSIM metric. Moreover, we extend the ISSIM metric into multi-scale space and deduce a MS-ISSIM based quality metric. Experimental results demonstrate the ISSIM/MS-ISSIM metric outperforms the SSIM/MS-SSIM metric.

References

1. Chen, G., Yang, C., Po, L., Xie, S.: Edge-Based structural similarity for image quality assessment. In: Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on. vol. 2, p. II (May 2006)
2. Katkovnik, V., Foi, A., Egiazarian, K., Astola, J.: From local kernel to nonlocal Multiple-Model image denoising. Int. J. Computer Vision 86, 1–32 (Jul 2009)
3. Li, C., Bovik, A.C.: Three-component weighted structural similarity index. pp. 72420Q–72420Q–9. SPIE (2009)

4. Lin, W., Kuo, C.J.: Perceptual visual quality metrics: A survey. J. Visual Communication and Image Representation 22(4), 297–312 (2011)
5. Liu, A., Lin, W., Narwaria, M.: Image quality assessment based on gradient similarity. IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society 21(4), 1500–1512 (Apr 2012)
6. Ponomarenko, N., Lukin, V., Zelensky, A., Egiazarian, K., Carli, M., Battisti, F.: Tid2008 - a database for evaluation of full-reference visual quality assessment metrics. Advances of Modern Radioelectronics 10, 30–45 (2009)
7. (VQEG), V.Q.E.G.: Final report from the video quality experts group on the validation of objective models of video quality assessment ii (2003), http://www.vqeg.org/
8. Wang, Z., Bovik, A., Sheikh, H., Simoncelli, E.: Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing 13(4), 600–612 (2004)
9. Wang, Z., Li, Q.: Information content weighting for perceptual image quality assessment. Image Processing, IEEE Transactions on 20(5), 1185–1198 (2011)
10. Wang, Z., Simoncelli, E., Bovik, A.: Multiscale structural similarity for image quality assessment. In: Signals, Systems and Computers, 2003. Conference Record of the Thirty-Seventh Asilomar Conference on. vol. 2, pp. 1398–1402 Vol.2 (2003)
11. Wu, J., Qi, F., Shi, G.: Self-similarity based structural regularity for just noticeable difference estimation. Journal of Visual Communication and Image Representation 23(6), 845–852 (2012)