An improved full-reference image quality metric based on structure compensation

Ke Gu, *Guangtao Zhai*, *Xiaokang Yang*, and *Wenjun Zhang*

Institute of Image Communication and Information Processing, Shanghai Jiao Tong University, Shanghai, China
Shanghai Key Laboratory of Digital Media Processing and Transmissions

***Abstract*—During the last two decades, image quality assessment has been a major research area, which considerably helps to promote the development of image processing. Following the tremendous success of Structural SIMilarity (SSIM) index in terms of the correlation between the quality predictions and the subjective scores, many improved algorithms have been further exploited, such as Multi-Scale SSIM (MS-SSIM) and Information content Weighted SSIM (IW-SSIM). However, a growing number of researchers have been devoted to the study of the effects of uneven responses to different image distortion categories on prediction accuracy of the quality metrics. Inspired by this, we propose an improved full-reference image quality assessment paradigm based on structure compensation. Experimental results on Laboratory for Image and Video Engineering (LIVE) database and Tampere Image Database 2008 (TID2008) are provided to confirm our introduced approach has superior prediction performance as compared to mainstream image quality metrics. Besides, it is worth emphasizing that our algorithm not introduces other operators but only applies the SSIM function to compensate itself, and furthermore, it also has an effective capability of image distortion classification.**

I. Introduction

The need of perceptual image quality assessment (IQA) is amplified by the rapid growth of multimedia applications, such as the development and optimization of image/video compression, storage and transmission algorithms. Existing IQA algorithms fall into two categories: subjective assessment and objective assessment. The subjective assessment method should be the ultimate quality gauge for digital images, but it is usually time-consuming, expensive and not practical for real-time image processing systems. The Mean-Squared Error (MSE) and its equivalent the Peak Signal-to-Noise Ratio (PSNR) are still the most widely used objective quality metrics as the benchmark in practice, because of both their convenience and their clear physical meaning as distortion/fidelity measures. However, it has been extensively recognized that MSE and PSNR are not well correlated with human judgment of quality, i.e. the Mean Opinion Score (MOS).

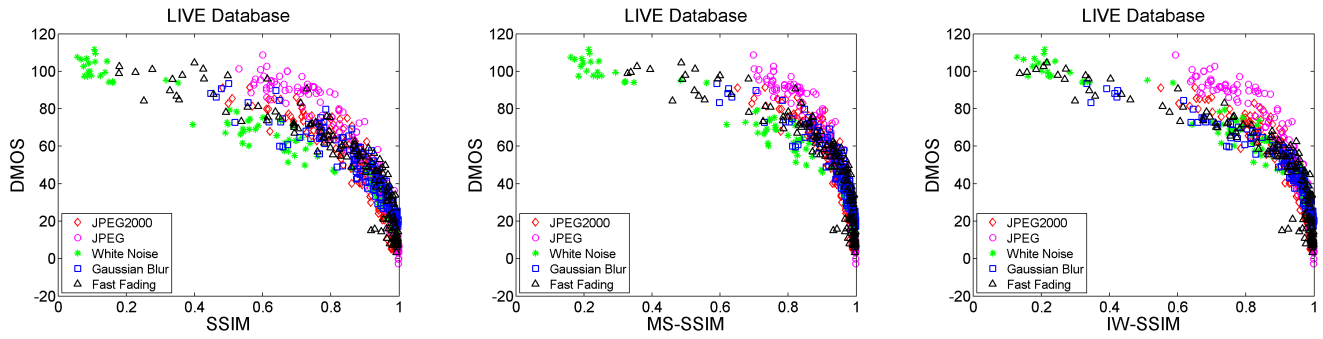
Inspired by a classical cognitivist paradigm of psychology [1], a large set of the so called cognitivist methods have been proposed thorough the years. The most popular cognitivist method Structural SIMilarity (SSIM) index [2], focusing on structural information substantially, made a breakthrough in the study of image quality metrics. And then, an increasing number of improved approaches have been further developed, such as Multi-Scale SSIM (MS-SSIM) [3] and Information content Weighted SSIM (IW-SSIM) [4]. In addition, various other cognitivist algorithms, including Visual Information Fidelity (VIF) [5] and a pixel based version of VIF (VIFP) [5], also have been proposed for better quality prediction. Very recently, researchers in the area of IQA realized the importance of distortion classification, which the SSIM metric hardly takes into account. For example, the Virtual Cognitive Model (VICOM) [6] tries to differentiate images with respect to their impairment type first to overcome the obstacle of uneven responses to common impairment sources.

In our research, it is observed that the application of Gaussian weighting window to compute the local statistics mean (defined by Eq. (4) later) in SSIM introduces different degrees of ambiguity for various types of image distortion. The ambiguity of mean can be estimated by the computing structural similarity between an image and its mean map. Then our proposed IQA paradigm can be evaluated by an effective nonlinear combination between SSIM method and structure compensation. In addition, our defined structure compensation can also be used to discriminate image distortion categories. Still, it is worth mentioning that the structure compensation is computed by using the SSIM index. In other words, it will be demonstrated in this paper that the shortcoming of SSIM, namely the ignorance of uneven responses to different distortion tpyes, can be overcome by SSIM itself.

The remainder of this paper is organized as follows. Section II first reviews SSIM as well as two effective improved approaches, MS-SSIM and IW-SSIM, and then points out their common shortage. In Section III, the significance and definition of structure compensation are described in detail. Section IV explicitly proposes our new image quality metric Structure Compensation based SSIM (SC-SSIM), and also presents the application of structure compensation to classify various image distortion types and its practical value. In Section V, experimental results using the Laboratory for Image and Video Engineering (LIVE) database [7] and Tampere Image Database 2008 (TID2008) [8] are reported and analyzed. Finally, conclusion is drawn in Section VI.

II. Background

The basic spatial domain SSIM algorithm [2] makes use of separated comparisons of local luminance, contrast and structure between a distorted image and its reference image. The luminance, contrast and structural similarities between



(a) SSIM (b) MS-SSIM (c) IW-SSIM

Fig. 1. Scatter plots of DMOS vs. (a): SSIM; (b): MS-SSIM; (c): IW-SSIM on LIVE database.

two local image patches extracted from the reference and distorted images are evaluated as

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

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

$$s(x,y) = \frac{2\sigma_x\sigma_y + C_3}{\sigma_x^2 + \sigma_y^2 + C_3} \tag{3}$$

where $C_1 = (K_1L)^2$, $C_2 = (K_2L)^2$ and $C_3 = C_2/2$. Using spatial patch with Gaussian weighting window $\omega = \{\omega_i | i = 1, 2, \ldots, N\}$, with standard deviation of 1.5 samples as well as normalized to unit sum ($\omega_i = 1$), the estimation of local statistics mean μ_x, standard deviation σ_x and cross-correlation σ_{xy} are given by

$$\mu_x = \Sigma_{i=1}^{N} \omega_i x_i \tag{4}$$

$$\sigma_x = (\Sigma_{i=1}^{N} \omega_i (x_i - \mu_x))^{\frac{1}{2}} \tag{5}$$

$$\sigma_{xy} = \Sigma_{i=1}^{N} \omega_i (x_i - \mu_x)(y_i - \mu_y). \tag{6}$$

Finally, the SSIM index evaluating the overall image quality is defined by

$$\begin{aligned} SSIM(X,Y) &= \frac{1}{M}\Sigma_{i=1}^{M} SSIM\_MAP(x_i, y_i) \\ &= \frac{1}{M}\Sigma_{i=1}^{M} l(x_i, y_i)c(x_i, y_i)s(x_i, y_i) \end{aligned} \tag{7}$$

where x_i and y_i are the image contents at the ith local window. X and Y are the reference and distorted images, respectively. M is the number of local windows in the image. Fig. 1 (a) displays the scatter plot of DMOS vs. SSIM.

Following the massive success of SSIM in terms of the correlation between the quality predictions and the subjective score, many improved metrics have been proposed later. Since the perceived quality of an image heavily depends upon the scale at which the image is analyzed, MS-SSIM [1] is exploited considering the effects of varied viewing distances. Besides, due to the fact that the pooling stage following local distortion/fidelity measurement lacks theoretical supports and reliable computational models, IW-SSIM [4] is enlightened by several recent successful image quality metrics [5], [9]. Scatter plots of DMOS vs. MS-SSIM and IW-SSIM are also shown in Fig. 1 (b)-(c), as compared with SSIM.

It is not difficult to find from Fig. 1 that uneven responses to different types of distortions indeed have important influences on the whole prediction accuracy, although this problem has been lessened a certain degree by MS-SSIM and IW-SSIM, as illustrated in Fig. 1 (b)-(c). So, an effective compensation method on this inconsistency is extremely needed, which is the topic of the next section.

III. Structure Compensation

It is explicitly explained in [2] that the application of spatial patch extracted is under the assumption that luminance and contrast can vary across scene, and the Gaussian weighting window is employed to avoid undesirable "blocking" artifacts, as defined in Eq. (4)-(6). However, due to the fact that the Gaussian weighting window of spatial patch in pixel domain presents a kind of low-pass filtering effect in spatial frequency domain, mean, standard deviation and cross-correlation maps are blurred prior to the evaluation of SSIM.

Intuitively, it is believed that the ambiguity, stemming from the application of low-pass filtering stated above, is of different degrees for the original image and the distorted image. Here we first define the above-mentioned ambiguity as

$$AMB(X) = S(X, \overline{X}) = S(X, B(X)) \tag{8}$$

where X is an image, $S(\cdot)$ and $B(\cdot)$ represents a similarity measuring and a blurred function, respectively. More specifically, the ambiguity of a reference image is constant while that of its distorted image varies with different distortion categories. For example, the ambiguity of a white noise image is larger than that of its reference image, but for JPEG2000 the result is completely on the contrary. As illustrated in Fig. 2-3, the ambiguity map of X ($AMB(X)$, defined by Eq. (9) later) has more black regions (i.e. more dissimilarity between X and μ_X) than the ambiguity map of Y ($AMB(Y)$) in Gaussian blur distortion; however, the situation is exactly

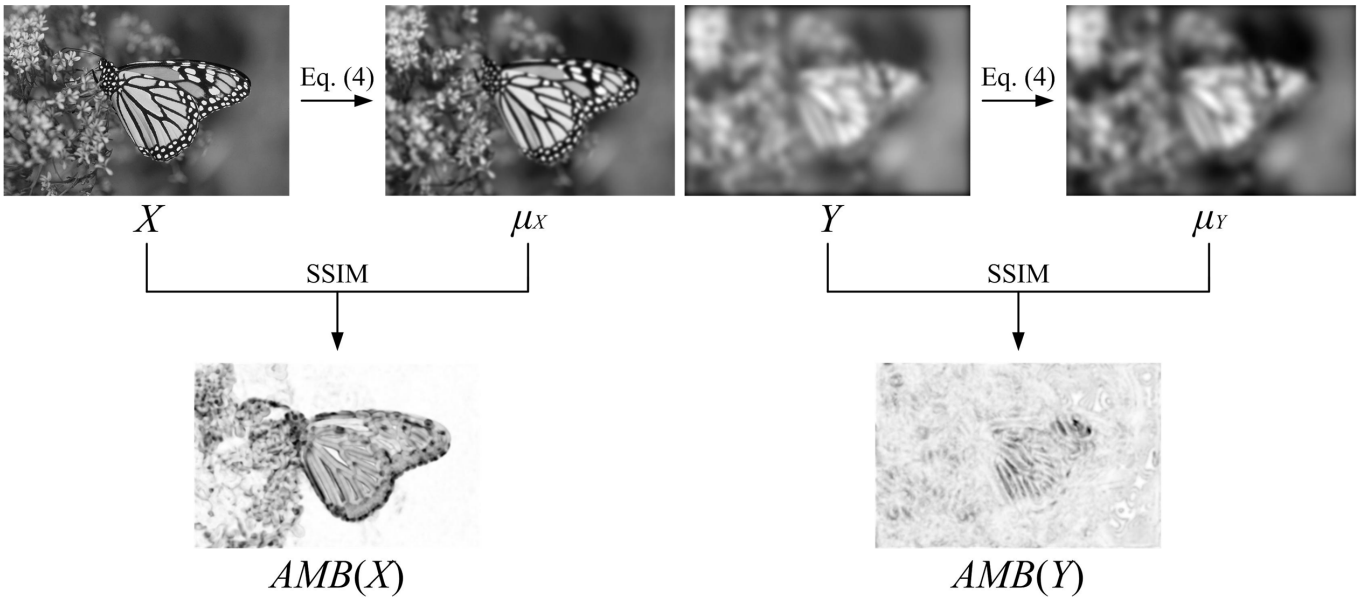


Fig. 2. X, μ_X and $AMB(X)$ are the reference image of Y, the mean map of X defined by Eq. (4) and the corresponding ambiguity map (more black regions indicating more dissimilarity between X and μ_X), respectively. Y, μ_Y and $AMB(Y)$ have similar definitions but for a Gaussian blurred image.

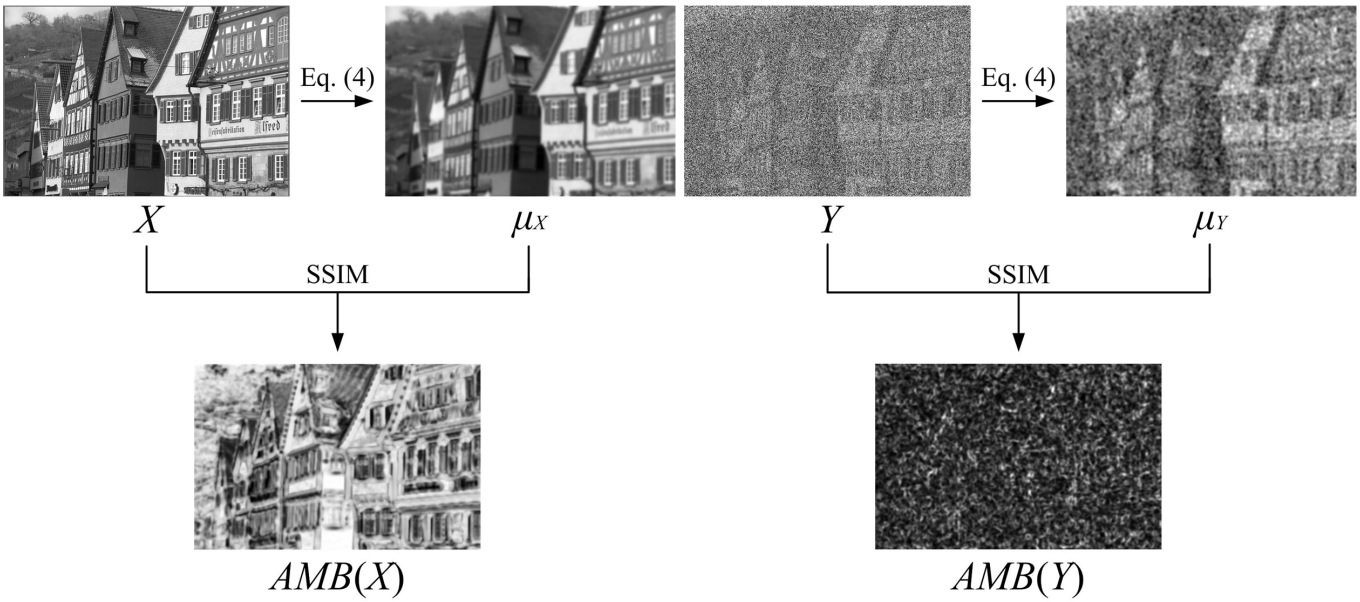


Fig. 3. X, μ_X and $AMB(X)$ are the reference image of Y, the mean map of X defined by Eq. (4) and the corresponding ambiguity map (more black regions indicating more dissimilarity between X and μ_X), respectively. Y, μ_Y and $AMB(Y)$ have similar definitions but for a white noise image.

opposite for white noise distortion. The phenomena can be explained by the fact that white noise images have more high-frequency components than JPEG2000, Gaussian blurred and fast fading images. In addition, the absolute difference of ambiguity ($|AMB(X) - AMB(Y)|$) between the reference and distorted image becomes larger with the lower quality of the distorted image (Y). Consequently, it is based on a belief in this paper that different degrees of ambiguity can be as an important and meaningful compensation on SSIM algorithm.

Firstly, using SSIM operator to estimate the structural similarity between an image and its mean map, just as illustrated in Fig. 2-3, the ambiguities for reference and distorted images is evaluated by

$$AMB(X) = SSIM(X, \mu_X) \tag{9}$$

and

$$AMB(Y) = SSIM(Y, \mu_Y) \tag{10}$$

where μ_X and μ_Y indicate the mean maps of reference and distorted images respectively, which can be computed by Eq.

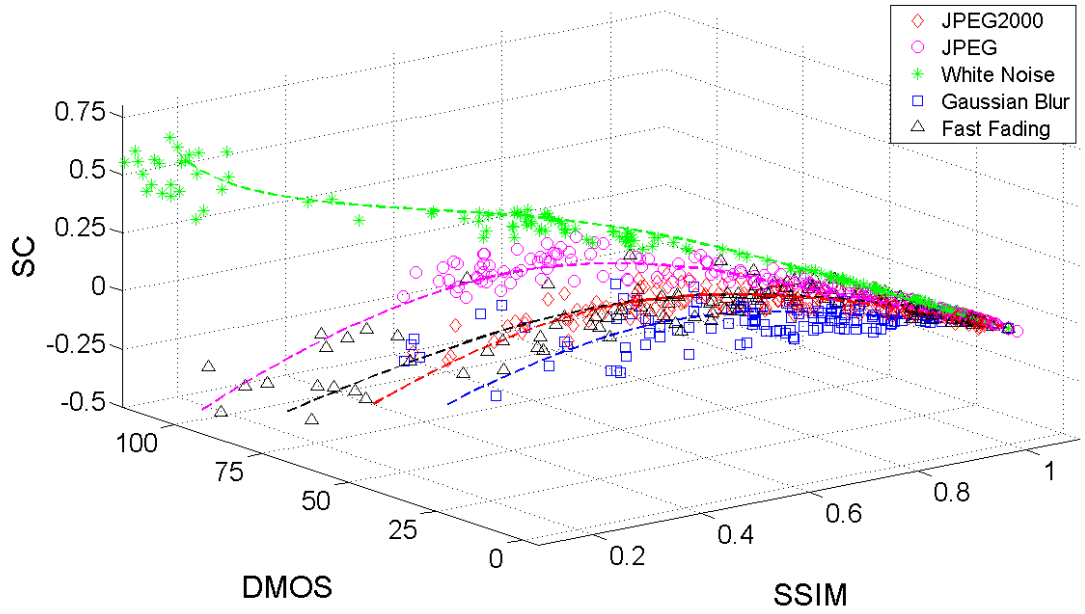


Fig. 4. Scatter plot of SC vs. SSIM and DMOS on LIVE database for five categories of image distortions, and their corresponding nonlinear fitting curves (dash lines).

(4). Secondly, the difference of ambiguities can be estimated as structure compensation (SC):

$$SC(X,Y) = AMB(X) - AMB(Y). \tag{11}$$

Fig. 4 illustrates the scatter plot of SC vs. SSIM and DMOS and provides a good sign to improve SSIM.

IV. The Proposed Method

For the application of structure compensation, we are still facing a major difficulty: how to combine SSIM and SC. To solve this problem, we introduced an empirical nonlinear additive model. Thus, the SC-SSIM is finally given by

$$\begin{aligned} SC\text{-}SSIM &= F_{SC}(SSIM, SC) \\ &= SSIM + \alpha(SC)^{\gamma_1} + \beta \begin{cases} (SC)^{\gamma_2} & \text{if } SC \geq 0 \\ (-SC)^{\gamma_3} & \text{otherwise} \end{cases} \end{aligned} \tag{12}$$

where α, β, γ_1, γ_2 and γ_3 are model parameters. To quantify these coefficients, all the images in LIVE database are randomly divided into two groups (training group with 644 images and testing group with 338 images) with respect to different reference images. Then, all the parameters in Eq. (12) can be obtained by a series of training to find the most reliable values.

To sum up, our SC-SSIM mainly has three steps, as displayed in Fig. 5: First, predict the SSIM between reference and distorted images; Second, evaluate their SC value; Third, compute the SC-SSIM result based on the nonlinear combination between SSIM and SC.

Furthermore, it is observed in the projection of Fig. 4 on SSIM-SC plane, as shown in Fig. 6, that the structure compensation also can be regarded as an effective categorical indicator. By finding the minimum distance value among a point vector $(SC, SSIM)$ and five nonlinear fitting curves in Fig. 6, the distortion type of the image represented by this point vector can be determined.

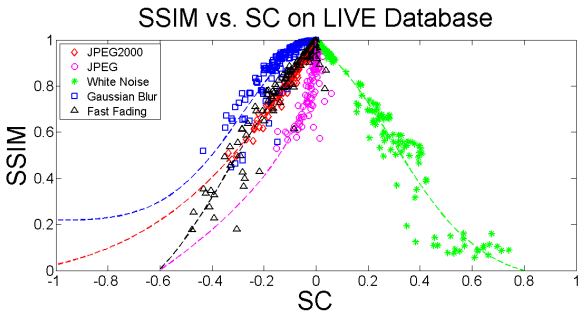


Fig. 6. Projection of Fig. 4 on SSIM-SC plane for five categories of image distortions, and their corresponding nonlinear fitting curves (dash lines).

Finally, it is worth emphasizing that our defined SC-SSIM can be rewritten by taking Eq. (9)-(11) into Eq. (12):

$$\begin{aligned} &SC\text{-}SSIM(X,Y) = F_{SC}(SSIM(X,Y) - SC(X,Y)) \\ &= F_{SC}(SSIM(X,Y), (SSIM(X,\mu_X) - SSIM(Y,\mu_Y))) \end{aligned} \tag{13}$$

and this indicates our algorithm not introduces other operators but only employs the SSIM function. In other words, we successfully use SSIM metric to remedy its shortage of ignoring inconsistent responses to different categories of image distortion.

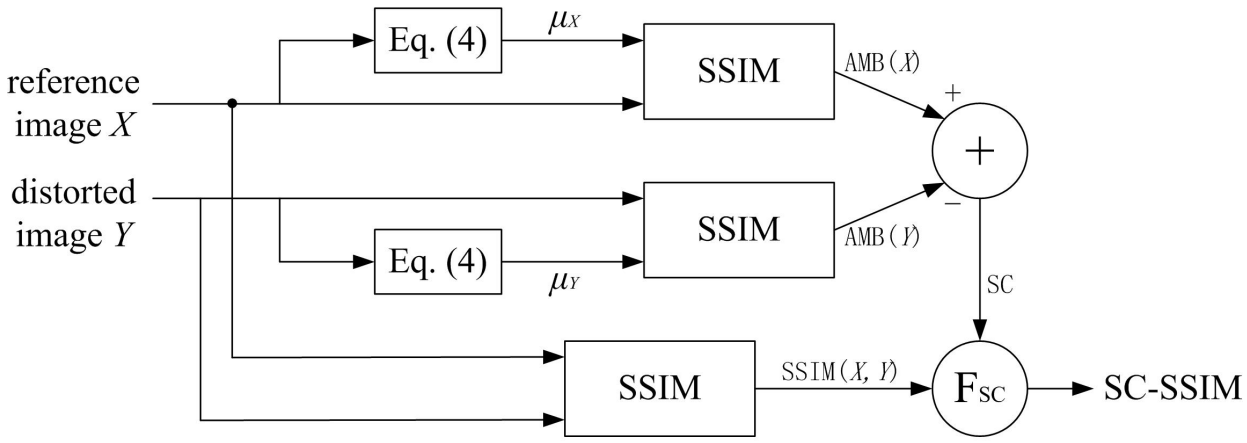


Fig. 5. Illustration of three steps of the SC-SSIM metric.

V. Experimental results

Mappings of the scores of six metrics SSIM, MS-SSIM, IW-SSIM, VIF, VIFP and the proposed SC-SSIM to subjective scores are obtained using nonlinear regression with a four-parameter logistic function as suggested by VQEG [10]:

$$q(x) = \frac{\beta_1 - \beta 2}{1 + exp(-(x - \beta_3)/\beta_4)} + \beta_2 \qquad (14)$$

with x being the input score and $q(x)$ the mapped score and β_1 to β_4 are free parameters to be determined during the curve fitting process.

Five commonly used performance metrics, Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-order Correlation Coefficient (SRCC), Kendall's Rank-order Correlation Coefficient (KRCC), Average Absolute prediction Error (AAE) and Root Mean-Squared (RMS) error as suggested by VQEG [10], are employed to further evaluate the competitive SC-SSIM metric and the six mainstream methods, namely SSIM, MS-SSIM, IW-SSIM, VIF, VIFP and DIP [11] on LIVE database and TID2008 subsets (Gaussian noise, Gaussian blur, JPEG, JPEG2000 and JPEG2000 transmission errors). Their values are illustrated in Table I-II and the scatter plots of D-MOS/MOS vs. SC-SSIM on LIVE database/TID2008 subsets are displayed in Fig. 7-8. It can be seen that our proposed SC-SSIM paradigm has achieved much better results than the six mainstream IQA metrics on the whole database. Besides, Table III presents SRCC results of VICOM and SC-SSIM, and as expected, our SC-SSIM also has higher prediction accuracy.

Except superior performance, it is important to clarify that our paradigm has two paramount merits: First, low computational complexity and high execute speed because of parallel processing of SSIM operation; Second, strong portability due to the fact that the structure compensation can be computed by the SSIM algorithm and some basic operations, such as addition and multiplication.

VI. Conclusion

In this paper, we propose a new structure compensation based SSIM paradigm. This research is devoted to three valuable findings. Above all, through the compensation of uneven responses to different categories of distortions, the proposed SC-SSIM paradigm has higher prediction accuracy on LIVE database and TID2008 subsets. Second, the structure compensation also can be as a categorical indicator to fast and effectively discriminate different image distortion types. Third, the SC-SSIM has low computational complexity, high execute speed and strong portability for it can be computed by the application of SSIM function only. Experimental results on LIVE database and TID2008 subsets verify that the performances of the proposed methods are clearly better than SSIM, MS-SSIM, IW-SSIM, VIF, VIFP and VICOM algorithms.

Acknowledgment

This work was supported in part by NSERC, NSFC (61025005, 60932006, 61001145), SRFDP (20090073110022), postdoctoral foundation of China 20100480603,201104276, postdoctoral foundation of Shanghai 11R21414200, the 111 Project (B07022) and STCSM (12DZ2272600).

TABLE I
PLCC, SRCC, KRSS, AAE AND RMS VALUES (AFTER NONLINEAR REGRESSION) OF SSIM, MS-SSIM, IW-SSIM, VIF, VIFP, AND SC-SSIM ON LIVE DATABASE (779 IMGES).

	PLCC	SRCC	KRCC	AAE	RMS
SSIM [2]	0.9383	0.9478	0.7961	7.5251	9.4508
MS-SSIM [3]	0.9402	0.9512	0.8043	7.4382	9.3121
IW-SSIM [4]	0.9425	0.9566	0.8174	7.4416	9.1344
VIF [5]	0.9594	0.9633	0.8273	6.2323	7.7102
VIFP [5]	0.9594	0.9618	0.8249	6.1186	7.7143
DIP [11]	0.9601	0.9642	0.8292	6.2206	7.6472
SC-SSIM	0.9620	0.9655	0.8364	6.0727	7.4610

TABLE II
PLCC, SRCC, KRSS, AAE AND RMS VALUES (AFTER NONLINEAR REGRESSION) OF SSIM, MS-SSIM, IW-SSIM, VIF, VIFP, AND SC-SSIM ON TID2008 SUBSETS (500 IMAGES), INCLUDING GAUSSIAN NOISE, GAUSSIAN BLUR, JPEG, JPEG2000 AND JPEG2000 TRANSMISSION ERRORS.

	PLCC	SRCC	KRCC	AAE	RMS
SSIM [2]	0.8576	0.8868	0.6939	0.5943	0.7563
MS-SSIM [3]	0.8669	0.8858	0.6930	0.5959	0.7332
IW-SSIM [4]	0.8977	0.9140	0.7345	0.5218	0.6481
VIF [5]	0.9178	0.9082	0.7373	0.4614	0.5838
VIFP [5]	0.9044	0.8872	0.7092	0.4995	0.6276
DIP [11]	0.9212	0.9051	0.7286	0.4792	0.6113
SC-SSIM	0.9211	0.9369	0.7776	0.4409	0.5727

TABLE III
SRCC VALUES (AFTER NONLINEAR REGRESSION) OF VICOM AND SC-SSIM ON LIVE DATABASE (982 IMAGES) AND TID2008 SUBSETS (500 IMAGES), INCLUDING GAUSSIAN NOISE, GAUSSIAN BLUR, JPEG, JPEG2000 AND JPEG2000 TRANSMISSION ERRORS.

	5 parameters VICOM [6]	6 parameters VICOM [6]	SC-SSIM
SRCC (LIVE)	0.9740	0.9750	0.9816
SRCC (TID2008)	0.9290	0.9300	0.9369

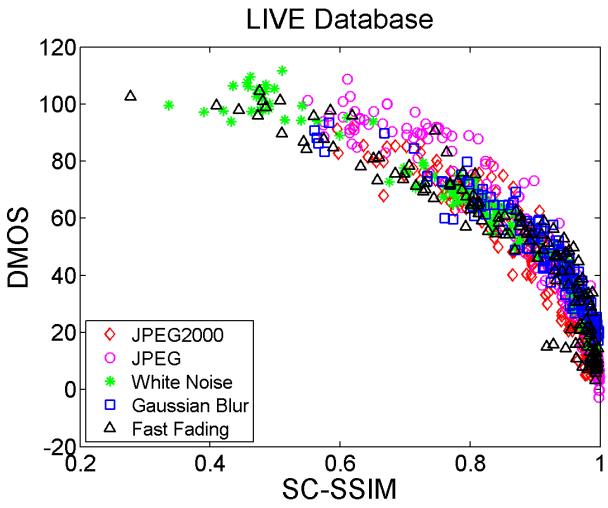


Fig. 7. Scatter plot of DMOS vs. SC-SSIM on LIVE databse.

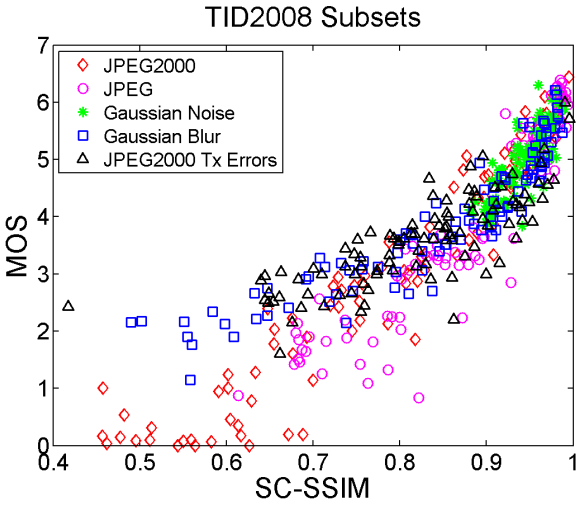


Fig. 8. Scatter plot of MOS vs. SC-SSIM on TID2008 subsets.

REFERENCES

[1] G. A. Miller, "The cognitive revolution: a historical perspective," *Trends in Cognitive Sciences*, vol. 7, no. 3, pp. 141-144, March 2003.

[2] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transaction on Image Processing*, vol. 13, no. 4, pp. 600-612, April 2004.

[3] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multi-scale structural similarity for image quality assessment," *IEEE Asilomar Conference Signals, Systems and Computers*, November 2003.

[4] Z. Wang and Qiang Li, "Information content weighting for perceptual image quality assessment," *IEEE Transaction on Image Processing*, vol. 20, no. 5, pp. 1185-1198, 2011.

[5] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," *IEEE Trans. Image Process.*, vol. 15, no. 2, pp. 430-444, February 2006.

[6] Licia Capodiferro, Giovanni Jacovitti and Elio D. Di Claudio, "Two-dimensional qpproach to full reference image quality Assessment Based on Positional Structural Information," *IEEE Trans. Image Process.*, 2012.

[7] H. R. Sheikh, K. Seshadrinathan, A. K. Moorthy, Z. Wang, A. C. Bovik and L. K. Cormack, "Image and video quality assessment research at LIVE," [Online]. Available: http://live.ece.utexas.edu/research/quality/

[8] N. Ponomarenko, V. Lukin, A. Zelensky, K. Egiazarian, M. Carli and F. Battisti, "TID2008-A database for evaluation of full-reference visual quality assessment metrics," *Advances of Modern Radioelectronics*, vol. 10, pp. 30-45, 2009.

[9] H. R. Sheikh, A. C. Bovik, and G. de Veciana, "An information fidelity criterion for image quality assessment using natural scene statistics," *IEEE Trans. Image Process.*, vol. 14, no. 12, pp. 2117-2128, 2005.

[10] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," March 2000, http://www.vqeg.org/.

[11] Ke Gu, G. Zhai, X. Yang and W. Zhang, "A new psychovisual paradigm for image quality assessment: from differentiating distortion types to discriminating quality conditions", *Signal, Image and Video Processing*, 2012.