A NEW IMAGE QUALITY METRIC BASED ON MIX-SCALE TRANSFORM

Min Liu, Guangtao Zhai, Ke Gu, and Xiaokang Yang

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

ABSTRACT

In the study of image quality assessment (IQA), multi-scale methods are often considered since they can supply more flexibility than single-scale algorithms by incorporating the variations of viewing conditions, such as famous MS-SSIM and VIF. However, it has been demonstrated in some recent researches that scale-transform based single-scale methods also achieved fairly well performances in terms of the correlation between the quality predictions and the subjective scores. Both of the single-scale based IQA metrics conform to a two-step model: image transform using proper scale coefficient and the SSIM metric, but the influence of different components in SSIM (luminance, contrast and structural similarity) on the chosen scale coefficient is not taken into account. In our research, it was found that the scale transform coefficients for different components should be different. This paper accordingly proposes a new MIx-Scale (MIS) based IQA algorithm. Experimental results on six publicly available databases (LIVE, TID2008, CSIQ, Toyama, IVC and LIVE Multiply Distortion) are provided to confirm that the proposed IQA metric can constantly lead to higher prediction accuracy for image quality assessment than traditional scale-based methods.

Index Terms**—** Image quality assessment (IQA), full-reference (FR), mix-scale (MIS)

1. INTRODUCTION

Image quality assessment (IQA) is a classical research direction, and it can play an important role in many areas of digital image processing, such as the development and optimization of image compression, storage, transmission and etc. Generally, image quality assessment includes two categories: subjective assessment and objective assessment. According to the instruction on subjective assessment in the ITU-R BT.500 [1] that is made by the International Telecommunication Union (ITU), the former designs experiments and obtains subjective mean opinion scores (MOSs) of observers. Due to some remarkable shortages of subjective assessment (e.g. inconvenience, time-consuming and expensiveness), objective IQA methods are extremely needed to automatically predict image quality. Among all the existing objective metrics, a pair of the most famous IQA metrics are mean-squared error (MSE) and peak signal-to-noise ratio (PSNR), for they are quite convenient and have the definite physical meaning.

During the latest decade, it has been widely recognized by more and more relative researchers that MSE/PSNR is not well correlated with human judgment of image quality/MOS [2]. Consequently, several advanced metrics like structural similarity index (SSIM) provide an alternative and complementary approach to solve the problem of IQA. These methods introduced in [3]-[4] belong to single-scale type of approaches. Despite of fairly well performance of abovementioned IQA metrics, it is not difficult to find that one significant impact factor "scale" was not taken into consideration. In order to fill this blank study, an increasing number of IQA algorithms have been developed based on the multi-scale strategy, for example, multi-scale SSIM (MS-SSIM) [5], information fidelity criterion (IFC) [6], visual information fidelity (VIF) [7], and information content weighting (IW) PSNR/SSIM [8]. The encouraging prediction accuracies of these IQA algorithms suggest the effectiveness of the multi-scale strategy.

Recently, some valuable researches [9]-[10] introduced another solution to the scale problem, and pointed out that the suitable single scale-transform based IQA method can also attain inspiring performance. Especially, the SAST model based PSNR/SSIM method [10] is executed on the resized reference and distorted images depending on the optimal scale parameter estimated from the image size and viewing distance. This approach is effective because as the viewing distance increases, the viewing angel shrinks and less image details can be noticed. Inspired by this, we have tested the influence of different components of SSIM (luminance, contrast and structural similarity) on the different selected scale coefficient on the prediction accuracy of IQA. Experimental results supported our supposition that different components should be set as various scale-transform coefficients. So, this paper proposes a new MIx-Scale based SSIM (MIS-SSIM) method.

The rest of this paper is organized as follows. In Section 2, we review and analyze the definition of SSIM, MS-SSIM and etc. Section 3 proposes our MIS-SSIM metric by combining different scale-transform coefficients into different components of SSIM. In Section 4, experimental results using LIVE database [11], TID2008 subsets [12], CSIQ subsets [13], IVC subsets [14], Toyama database [15] and a recently released LIVE Multiply Distortion database [16] are reported and analyzed. Finally, Section 5 concludes this paper.

2. BACKGROUND

According to the research of IQA in [4], the human visual system (HVS) is highly adapted for luminance, contrast and structural information between a distorted image and its reference image. In general, the luminance of the surface of an object being observed is defined as the product of the illumination and the reflection, while the contrast is measured as the difference in luminance that makes an object (or its representation in an image) distinguishable. And the structural information in an image represents the structure of objects in the scene. Here we emphasize an important thing that the above-mentioned three components are relatively independent of others, which has been proved in [4].

Depending on luminance, contrast and structural comparisons, a basic full-reference SSIM algorithm [4] is proposed to predict the image quality. The luminance, contrast and structural similarities between two local image patches extracted from the reference and distorted images are defined as

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

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

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

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

$$\mu_x = \frac{1}{N}\sum_{i=1}^{N}\omega_i x(i) \quad (4)$$

$$\sigma_x = \frac{1}{N-1}(\sum_{i=1}^{N}\omega_i(x(i) - \mu_x))^{1/2} \quad (5)$$

$$\sigma_{xy} = \frac{1}{N-1}\sum_{i=1}^{N}\omega_i(x(i) - \mu_x)(y(i) - \mu_y) \quad (6)$$

The SSIM index evaluating the overall image quality is defined by

$$SSIM(x,y) = \frac{1}{M}\sum_{i=1}^{M} SSIM_M AP(x(i), y(i)) \quad (7)$$

$$= \frac{1}{M}\sum_{i=1}^{M} l(x(i), y(i))c(x(i), y(i))s(x(i), y(i)) \quad (8)$$

where $x(i)$ and $y(i)$ are the image contents at the ith local window. x and y are the reference and distorted images. M is the number of local windows in the image.

However, the successful SSIM metric was not considering the influence of scale, so the MS-SSIM incorporating different viewing conditions was exploited to solve this problem. As suggested in [5], a system takes the reference and distorted image signals as the input, then iteratively applies a low-pass filter and downsamples the filtered image by a factor by 2. We indicate the original image as Scale 1, and the highest scale as Scale N, which is achieved after $N-1$ iterations. The luminance comparison is computed only at Scale N. At the jth scale, the contrast comparison and the structure comparison are individually calculated and denoted as $c_j(x,y)$ and $s_j(x,y)$. We combine three components at different scales to achieve the overall SSIM evaluation

$$SSIM(x,y) = [l_N(x,y)] \cdot \Pi_{j=1}^{N}[c_j(x,y)][s_j(x,y)] \quad (9)$$

It is particularly noteworthy that single-scale and multi-scale are both considered in this formula, in order to make different parameter settings (including all single-scale and multi-scale settings) comparable. Finally, [5] found out that multi-scale SSIM model outperforms the best single-scale model.

Recently, some valuable researches [9]-[10] have developed another solution for the above-mentioned problem, and have demonstrated that the suitable single scale-transform based SSIM method can also attain inspiring performance. A simple and empirical method mentioned in [9] has been exploited for SSIM to determine the downsampling scale S for evaluating images viewed from a typical distance:

$$S = max(1, round(H/256)) \quad (10)$$

with H being the image height. So, for the reference image x and distorted image y, the improved SSIMz metric processed by scale transfer can be defined as

$$\begin{aligned} SSIMz &= SSIM(x,y) \\ &= SSIM(R(L(x), S), R(L(y), S)) \quad (11) \end{aligned}$$

where $R(\cdot)$ and $L(\cdot)$ indicate image resizing and low-pass filtering function, respectively. In addition, the recently proposed SAST model also achieved encouraging results through using the concept of human visual angle and angle of gaze to estimate the optimal scale coefficients, as illustrated in [10].

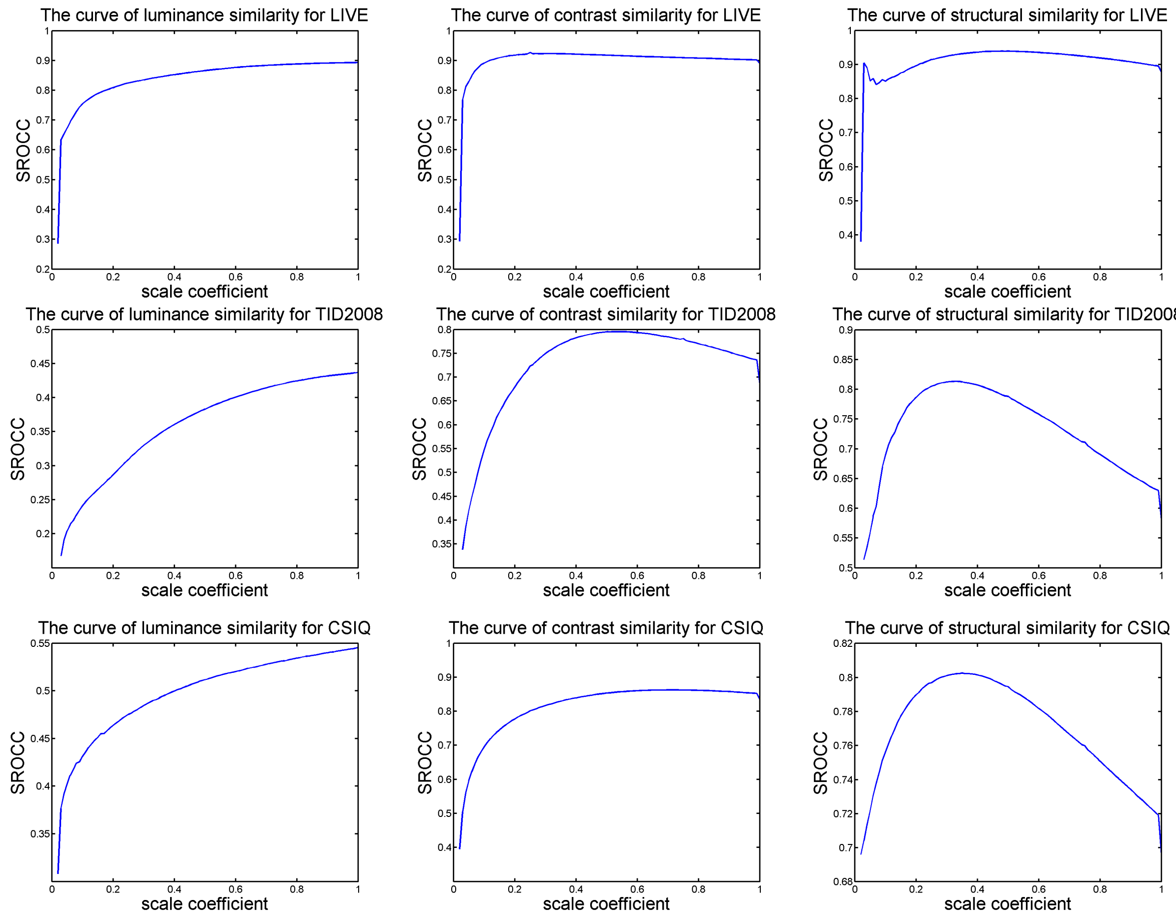

Fig. 1. The changing of curve of luminance, contrast, structural similarities under different scale coefficients for (a)-(c): LIVE database; (d)-(f): TID2008 database; (g)-(i): CSIQ database.

3. THE PREDICTION MODEL

3.1. The Analysis of the influence of Scale on Different Components in SSIM Metric

Despite of fairly well performance of SSIMz and SAST methods, in our research, we noticed that the optimal scale coefficients for different components in SSIM should be different. Here, we adopt six databases (LIVE, TID2008, CSIQ, IVC, Toyama and LIVE Multiply Distortion databases) as testing beds. Furthermore, we consider five common distortion types (e.g. JPEG2K, JPEG, White Noise, Blur, and Fast Fading) as suggested by LIVE database. Fig. 1 illustrates the changing of curves of luminance, contrast, and structural similarities in three large databases, namely LIVE, TID2008 and CSIQ. It is not difficult to find that, for the luminance similarity in Fig. 1 (a), (d), and (g), the corresponding curves exert less influence on the prediction accuracy as the scale coefficient becomes smaller. In other words, the luminance similarity had better be computed at the most proper scale of original size. Moreover, to discard the luminance similarity probably brings higher performance. However, the trend for curves of contrast and structural similarities are quite different. Differing from luminance (see the Fig. 1 (b)-(c), (e)-(f), and (h)-(i)), contrast and structural similarities achieve the maximum correlation values when the scale is small, around 0.5, which can explain why the SSIMz processed by scale transform (with coefficient = 0.5) can obtain higher performance.

In the recently proposed SAST model [10], the authors took the viewing conditions (display resolution and viewing distance) into consideration, and made it a point that suitable scale chosen is a key factor for the research of IQA. Although the exact viewing conditions of CSIQ, TID2008 and LIVE Multiply Distortion databases cannot be found, we believe their viewing conditions should be within the appropriate range, similar to that of LIVE, IVC, and Toyama databases. Therefore, a group of three scale coefficients for luminance, contrast, and structural similarities will have better results than SSIM, SSIMz and MS-SSIM metrics, which will be illustrated in the next subsection.

3.2. The Proposed MIS-SSIM Metric

Encouraged by the progress achieved by the SSIMz metric, we redefine the three components of MIS-SSIM model:

$$l(x_l, y_l) = \frac{2\mu_{x_l}\mu_{y_l} + C_1}{\mu_{x_l}^2 + \mu_{y_l}^2 + C_1} \quad (12)$$

$$c(x_c, y_c) = \frac{2\sigma_{x_c}\sigma_{y_c} + C_2}{\sigma_{x_c}^2 + \sigma_{y_c}^2 + C_2} \quad (13)$$

$$s(x_s, y_s) = \frac{\sigma_{x_s y_s} + C_3}{\sigma_{x_s}\sigma_{y_s} + C_3} \quad (14)$$

where $x_j = R(L(x), S_j)$, and $y_j = R(L(y), S_j)$. $S = \{S_l, S_c, S_s\}$ indicate one group of mix-scale coefficients for luminance, contrast and structural similarities. Their exact values used in this paper will be provided in the next section.

In hence, the MIS-SSIM index evaluating the overall image quality is defined by

$$MIS - SSIM(R(L(x), S), R(L(y), S)) \quad (15)$$

$$= MIS - SSIM(x_j, y_j | j = l, c, s) \quad (16)$$

$$= \prod_{j=l,c,s} \frac{1}{M_j} \sum_{i=1}^{M_j} l(x_k(i), y_k(i)) \quad (17)$$

where M_l, M_c and M_s are the number of local windows in the image for luminance, contrast and structural similarities.

It is demonstrated that luminance, contrast, and structural similarity play different effects on the IQA. To illustrate in detail, we come to three conclusions. Firstly, contrast has more effect on IQA than luminance because the human visual system is more sensitive to contrast than absolute luminance. Secondly, structural similarity has the largest effect on IQA. Thirdly, compared with images after down-sampling, luminance, contrast and structural similarity have different maps respectively.

4. EXPERIMENTAL RESULTS AND ANALYSIS

In this paper, the scale coefficients of contrast and structural similarities for all six databases are set as (S_c=) 0.40 and (S_s=) 0.22, while the term of luminance similarity is not applied here. Then, as suggested by VQEG [17], we use nonlinear regression with a four parameter logistic function for mappings of the scores of four metrics (SSIM, SSIMz, MS-SSIM and the proposed MIS-SSIM) to subjective MOS/DMOS values:

$$q(x) = \frac{\beta_1 - \beta_2}{1 + exp{-}(x - \beta_3)/\beta_4} + \beta_2 \quad (18)$$

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

Three commonly used performance metrics, Pearson Linear Correlation (PLCC), Spearman Rank-Order Correlation Coefficient (SROCC) and Root Mean Square Error (RMSE) are employed to further evaluate SSIM, SSIMz, MS-SSIM and the proposed MIS-SSIM metrics on LIVE, TID2008, C-SIQ, IVC, Toyama and LIVE Multiply Distortion databases. It is worth emphasizing that this experiment only adopts five distortion types JPEG2K, JPEG, White Noise, Blur, and Fast Fading, or their version of multiply distortions. All the PLC-C, SROCC and RMSE results are illustrated in Table I-III in detail. The scatter plots of MOS/DMOS versus model predictions of SSIM, SSIMz, MS-SSIM and MIS-SSIM are shown in Fig. 2 - Fig. 7. It can be seen that the proposed MIS-SSIM paradigm has achieved better results than the other three IQA metrics used in our test.

Note that the proposed MIS-SSIM belongs to a general model, meaning that it can be easily used to modify other powerful IQA metrics, such as some improved versions of S-SIM [xx]-[xx]. Besides, it is also needed to emphasize that we are capable of expanding the proposed algorithm, which only makes some small changes on SSIM, to video quality assessment (VQA), since its simplicity and SSIM has been inserted into most existing video coding systems and softwares as one of benchmark image quality metrics. Accordingly, we have a reason to believe that the MIS-SSIM based VQA approach will have exiciting performance.

Table 1. PLCC VALUES (AFTER NONLINEAR REGRESSION) OF S-SIM, SSIMZ, MS-SSIM AND MIS-SSIM ON LIVE, TID2008, CSIQ, IVC, TOYAMA AND LIVE MULTIPLY DISTORTION DATABASES.

Person Linear Correlation Coefficient (PLCC)				
Databases	SSIM	SSIMz	MS-SSIM	MIS-SSIM
LIVE	0.9014	**0.9369**	0.9338	0.9134
TID2008	0.7362	0.8653	0.8679	**0.8699**
CSIQ	0.8513	0.9295	0.9390	**0.9439**
IVC	0.8583	0.9272	0.9154	**0.9288**
Toyama	0.7978	0.8899	0.8926	**0.9099**
LIVE Multiply Distortion	0.7333	0.8827	0.8749	**0.8924**
All	0.8130	0.9053	0.9058	**0.9114**

Table 2. SROCC VALUES (AFTER NONLINEAR REGRESSION) OF SSIM, SSIMZ, MS-SSIM AND MIS-SSIM ON LIVE, TID2008, CSIQ, IVC, TOYAMA AND LIVE MULTIPLY DISTORTION DATABASES.

Spearman Rank Orderd Correlation Coefficient (SROCC)				
Databases	SSIM	SSIMz	MS-SSIM	MIS-SSIM
LIVE	0.9104	**0.9490**	0.9448	0.9369
TID2008	0.7641	0.8947	0.8867	**0.8964**
CSIQ	0.8767	0.9324	0.9416	**0.9557**
IVC	0.8424	0.9153	0.9154	**0.9288**
Toyama	0.7870	0.8830	0.8870	**0.9083**
LIVE Multiply Distortion	0.6459	0.8455	0.8392	**0.8609**
All	0.8044	0.9033	0.9024	**0.9145**

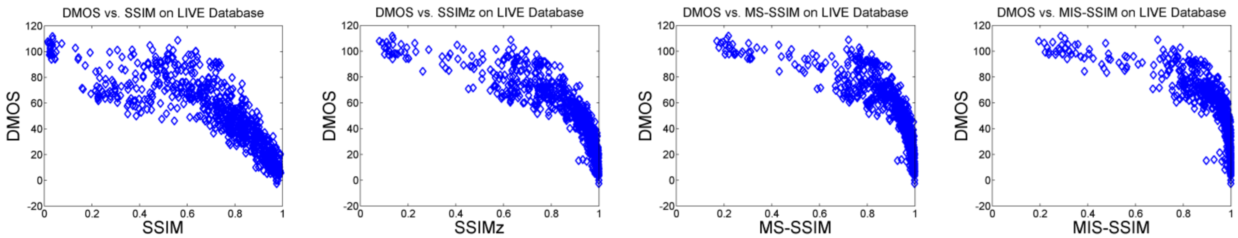

Fig. 2. Scatter plots of DMOS vs. SSIM, SSIMzMS-SSIM and the proposed MIS-SSIM metrics on the LIVE database.

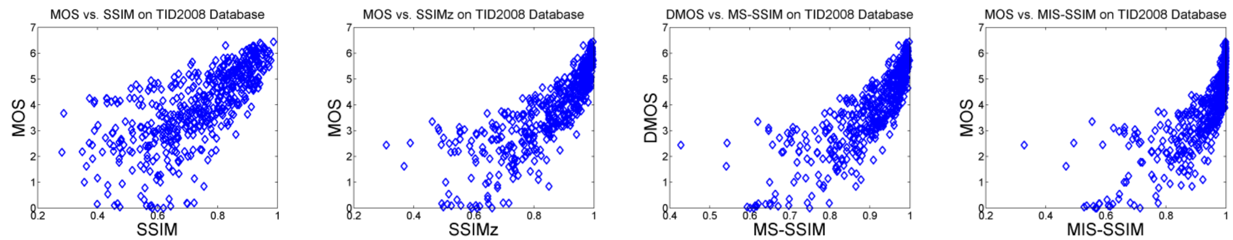

Fig. 3. Scatter plots of MOS vs. SSIM, SSIMzMS-SSIM and the proposed MIS-SSIM metrics on the TID2008 subsets, including JPEG, JPEG2K, White Noise, Blur and Fast Fading distortions.

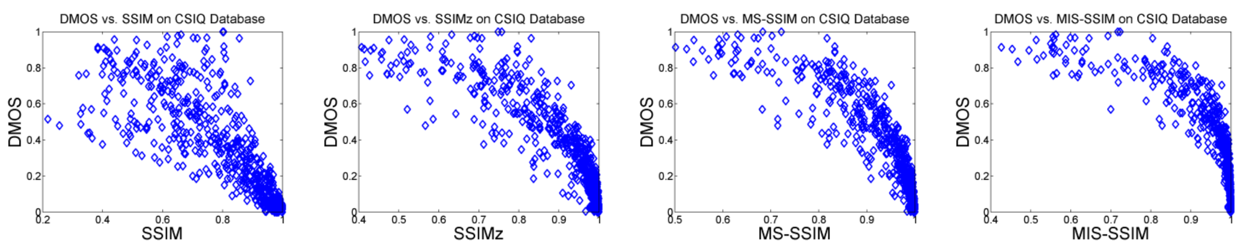

Fig. 4. Scatter plots of DMOS vs. the SSIM, SSIMz, MS-SSIM and the proposed MIS-SSIM metrics on the CSIQ subsets, including JPEG, JPEG2K, White Noise and Blur distortions.

Table 3. RMSE VALUES (AFTER NONLINEAR REGRESSION) OF S-SIM, SSIMZ, MS-SSIM AND MIS-SSIM ON LIVE, TID2008, CSIQ, IVC, TOYAMA AND LIVE MULTIPLY DISTORTION DATABASES.

Root Mean Square Error (RMSE)				
Databases	SSIM	SSIMz	MS-SSIM	MIS-SSIM
LIVE	11.8323	**9.5503**	9.7788	11.1215
TID2008	0.9954	0.7371	0.7306	**0.7254**
CSIQ	0.1483	0.1042	0.0972	**0.0933**
IVC	0.6373	0.4651	0.4665	**0.4263**
Toyama	0.7545	0.5709	0.5641	**0.5192**
LIVE Multiply Distortion	12.8586	8.8867	9.1596	**8.5346**
All	4.5377	**3.3857**	3.4662	3.5701

5. CONCLUSIONS

In this paper, based on the changeable effects of luminance, contrast and structural similarities on the IQA during the s-cale variations, we could get better performance by combin-ing these three components under different scales together. We have tested the proposed improved SSIM method for the IQA algorithms on LIVE, TID2008, CSIQ, IVC, Toyama and LIVE Multiply Distortion databases. Results of experiments verified that our proposed MIS-SSIM method has superior prediction accuracy for IQA.

This paper proposes a general model which can be used to other improved SSIM or VIF model. Besides, it can be expanded to video quality assessment which is superior than VIF, because VIF is based on wavelet,

6. REFERENCES

[1] ITU, “Methodology for the subjective assessment of the quality of television pictures,” Recommendation, International Telecommunication Union/ITU Radio communication Sector, 2009.

[2] Z. Wang and A. C. Bovik, “Mean squared error: Love it or leave it? A new look at signal fidelity measures,” *IEEE Signal Processing Mag.*, vol. 26, pp.98-117, January 2009.

[3] Z. Wang and A. C. Bovik, “A universal image quality index,” *IEEE Signal Processing Letters*, vol. 9, pp. 81-84, March 2002.

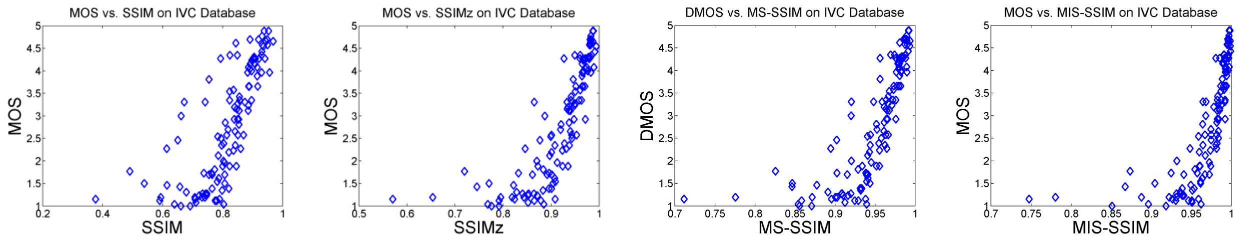

Fig. 5. Scatter plots of MOS vs. SSIM, SSIMzMS-SSIM and the proposed MIS-SSIM metrics on the IVC database subsets, including Jpeg, Jpeg2K and Blur distortions.

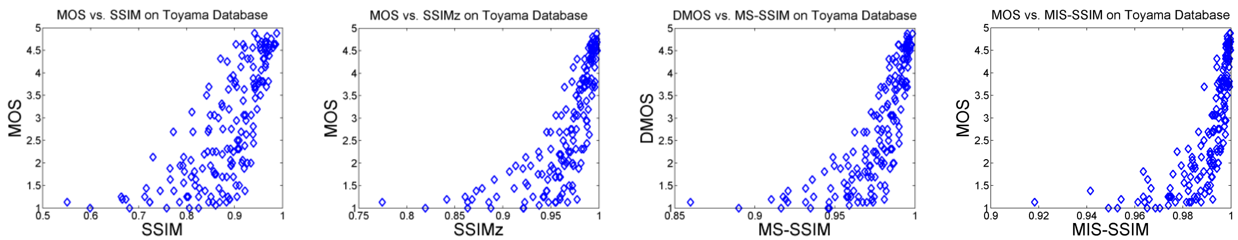

Fig. 6. Scatter plots of MOS vs. SSIM, SSIMzMS-SSIM and the proposed MIS-SSIM metrics on the Toyama database.

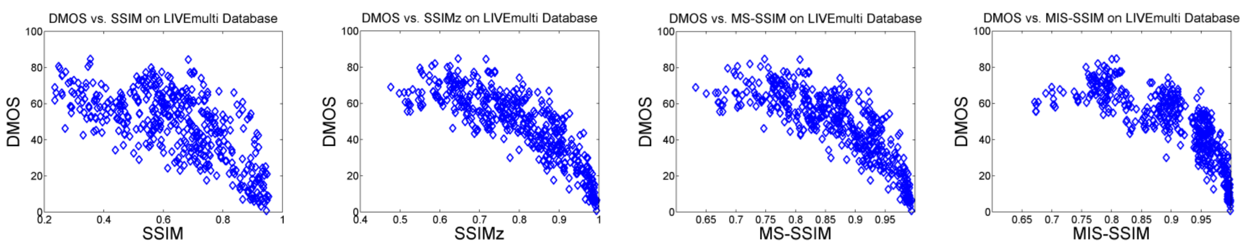

Fig. 7. Scatter plots of DMOS vs. SSIM, SSIMz, MS-SSIM and the proposed MIS-SSIM metrics on the LIVE multiply distortion database.

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

[5] 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.

[6] 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.

[7] 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.

[8] Z. Wang and Q. Li, "Information content weighting for perceptual image quality assessment," *IEEE Trans. Image Process.*, vol. 20, no. 5, pp. 1185-1198, 2011.

[9] W. Lin and C.-C. Jay Kuo, "Perceptual visual quality metrics: A survey," *J. Vis. Commun. Image R.*, vol. 22, no. 4, pp. 297-312, May 2011.

[10] K. Gu, G. Zhai, X. Yang, and W. Zhang, "Self-adaptive scale transform for IQA metric," *Proc. IEEE International Symposium on Circuits and Systems*, 2013.

[11] H. R. Sheikh, Z. Wang, L. Cormack, and A. C. Bovik, "LIVE image quality assessment Database Release 2," [Online]. Available: http://live.ece.utexas.edu/research/quality.

[12] 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 Radio electronics*, vol.10, pp. 30-45, 2009.

[13] E. C. Larson, and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, 2010.

[14] A. Ninassi, P. Le Callet, and F. Autrusseau, "Subjective quality assessment-IVC database," [Online]. Available: http://www2.irccyn.ec-nantes.fr/ivcdb.

[15] Y. Horita, K. Shibata, Y. Kawayoke, and Z. M. P. Sazzad, "MICT image quality evaluation database," [Online]. Available: http://mict.eng.utoyama.ac.jp/mict/index2.html.

[16] Dinesh Jayarraman, Anish Mittal, Anush K.Moorthy, and Alan C.Bovik, "Objective quality assessment of multiply distorted images," *Proceedings of Asilomar Conference on Signals, Systems and Computers*, 2012.

[17] 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/.