STRUCTURAL SIMILARITY WEIGHTING FOR IMAGE QUALITY ASSESSMENT

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

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

ABSTRACT

Recently, there has been a trend of investigating weighting/pooling strategies in the research of image quality assessment (IQA). The saliency maps, information content maps and other weighting strategies were reportedly to be able to amend performance of IQA metrics to a sizable margin. In this work, we will show that local structural similarity is itself an effective yet simple weighting scheme leading to substantial performance improvement of IQA. More specifically, we propose a Structural similarity Weighted SSIM (SW-SSIM) metric by locally weighting the SSIM map with local structural similarities computed using SSIM itself. Experimental results on LIVE database confirm the performance of SW-SSIM as compared to some major weighting/pooling type of IQA methods, such as MS-SSIM, WSSIM and IW-SSIM. We would like to emphasize that our SW-SSIM is merely a straightforward realization of a more general framework of locally weighting IQA metric using itself as similarity measures.

Index Terms— Image quality assessment (IQA), saliency, structural similarity, pooling

1. INTRODUCTION

Image quality assessment (IQA) plays a significant role in image processing. Perceptual IQA metrics can assist the development and optimization of image acquisition, communication, compression, storage and display systems. Generally, image quality assessment can be divided into two kinds according to the application scenarios. The first kind of methods are the so-called subjective approaches that are generally recognized as the ultimate image quality gauge. The other kind is objective assessment aiming at automatic prediction of human response to image quality. Due to the fact that the subjective assessment is accompanied with some remarkable drawbacks, such as being time-consuming, expensive and laborious, there has been an increasing interest in developing objective IQA metrics. Depending on the availability of reference images, objective IQA algorithms can be further classified into three categories: namely the well-known full-reference (FR), reduced-reference (RR) and no-reference (NR) algorithms. In this paper, we focus on FR image quality metric.

The full-reference IQA approaches generally adopt a two-stage method: local distortion/fidelity measurement and pooling. Following the great success of SSIM [1] and VIF [2], local distortion/fidelity type of methods have attracted a great deal of attention from researchers. Meanwhile, a number of weighting/pooling strategies [3]-[10] also have been proposed for the second stage. Considering the fact that the perceived quality of image highly relies upon the scale at which the image is observed, MS-SSIM [3] was exploited by incorporating various viewing conditions. Through incorporating properties of the human visual system (HVS) into image quality metrics, WSSIM [8] was developed as a saliency map weighting based approach. To sum up, the basic idea of [3]-[8] is a common hypothesis that the pooling strategy should take into consideration the human visual fixation or visual region-of-interest detection.

However, as can be seen from Fig. 2-3, the saliency maps computed by visual attention model (VAM), or even those recorded directly by eye-tracker [8], do not always capture the regions with the most "apparent" distortions. Clearly, the reference images alone cannot generate faithful estimates of the saliency features. Therefore, a newly proposed approach S_NW-SSIM [9] appropriately integrated saliency features from both the original and distorted images based on the method of [11] to form the final saliency map. Unfortunately, this new saliency map based weighting strategy only brought limited improvement of prediction accuracy for IQA.

Recently, statistical information theory inspired IW-SSIM [10] has achieved very good performance and is currently the de facto benchmark for pooling-type of IQA methods. It is both interesting and intriguing to realize that the root of IW-SSIM, namely, the information content weighting (IW) map is originated from the IQA metric of VIF [2]. It is a natural speculation that this IQA based pooling strategy is indeed very effective in improving the performance of the IQA metrics themselves. In our research, it was found that the weighting map based on the notion of local similarity estimated directly using IQA metrics can be very useful for improving the prediction accuracy of IQA. With this observation, we design a low complex and high performance Structural similarity based Weighting (SW) strategy, and name the corresponding IQA method the Structural similarity Weighted SSIM (SW-SSIM).

The remainder of this paper is organized as follows. Section 2 first provides the detailed description of the proposed SW-SSIM paradigm, and then justifies the effectiveness of our SW map by comparing different major saliency maps. In Section 3, experimental results using the LIVE database [12] are reported and analyzed. Finally, some concluding remarks are given in Section 4.

2. THE PROPOSED IMAGE QUALITY METRIC

While the most prevailing weighting/pooling scheme for IQA methods are those based on the visual attention mechanisms and saliency features, in this research, we investigate the possibility of weighting with local similarity measures quantified by the IQA method itself. More specifically, we proposed to compute the local structural similarities through applying SSIM to local image blocks and further weight the SSIM map with the local similarity.

2.1. The Proposed SW-SSIM Metric

The first part of the proposed SW-SSIM metric is a structural similarity weighting map that is evaluated by a modified SSIM used on local blocks. The weighting coefficients for a block is estimated by calculating the structural similarity between it and its eight neighboring blocks.

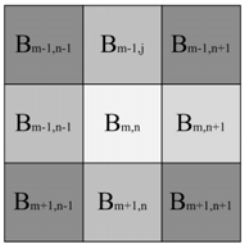

Fig. 1. Illustration of structural similarity weighting strategy for a block $B_{m,n}$ with the size of $M \times M$. Its surrounding eight blocks are used to compute the structural similarity as the weighting value of the current block $B_{m,n}$.

For a block $B_{m,n}$ with the size of $M \times M$ is located in an image. Its weighting coefficients $\nu(m,n,N)$ are computed as follows:

$$\nu(m,n,N) = \frac{\sum_{k=1}^{8} \nu_s(k) \cdot m_s(k,N)}{\sum_{k=1}^{8} \nu_s(k)} \tag{1}$$

where N is an additional variable to be defined later. $\nu_s(k)$ are eight constant model parameters. $k = \{1...4\}$ indicate top, right, bottom and left light gray blocks, and $k = \{5...8\}$ indicate top-right, bottom-right, bottom-left and top-left dark gray blocks, as shown in Fig. 1.

The measures of similarity degree $m_s(k,N)$ between $B_{m,n}$ and one of its eight neighboring blocks can be evaluated by

$$\begin{aligned}
m_s(1,N) &= SSIM(B_{m,n}, B_{m-1,n}, N) \\
m_s(2,N) &= SSIM(B_{m,n}, B_{m,n+1}, N) \\
m_s(3,N) &= SSIM(B_{m,n}, B_{m+1,n}, N) \\
m_s(4,N) &= SSIM(B_{m,n}, B_{m,n-1}, N) \\
m_s(5,N) &= SSIM(B_{m,n}, B_{m-1,n+1}, N) \\
m_s(6,N) &= SSIM(B_{m,n}, B_{m+1,n+1}, N) \\
m_s(7,N) &= SSIM(B_{m,n}, B_{m+1,n-1}, N) \\
m_s(8,N) &= SSIM(B_{m,n}, B_{m-1,n-1}, N)
\end{aligned} \tag{2}$$

where SSIM is a revised version of [1] to be defined as follows. Consider X and Y to be two matrices with the same size. Let μ_X, μ_Y, σ_X^2, σ_Y^2 and σ_{XY} be the means, variances and covariance between X and Y. The luminance, contrast and structural similarities are estimated as

$$l(X,Y,N) = \frac{2\mu_X\mu_Y + C_1}{\mu_X^2 + \mu_Y^2 + C_1} \tag{3}$$

$$c(X,Y,N) = \frac{2\sigma_X\sigma_Y + C_2}{\sigma_X^2 + \sigma_Y^2 + C_2} \tag{4}$$

$$s(X,Y,N) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \tag{5}$$

where C_1, C_2 and C_3 are small constants used to avoid instability when the denominators are very close to zero. Here we use a $N \times N$ circular-symmetric Gaussian weighting function $\mathbf{w} = \{w_{ij} | i,j = 1,\ldots,N\}$, with standard deviation of 1.5 samples, normalized to unit sum ($\sum_{i=1}^{N}\sum_{j=1}^{N} w_{ij} = 1$). The statistics μ_X, σ_X^2 and σ_{XY} can be computed by

$$\mu_X = \frac{1}{N^2} \sum_{i,j=1}^{N} w_{ij} x_{ij} \tag{6}$$

$$\sigma_X^2 = \frac{1}{N^2 - 1} \sum_{i,j=1}^{N} w_{ij} (x_{ij} - \mu_X)^2 \tag{7}$$

$$\sigma_{XY} = \frac{1}{N^2 - 1} \sum_{i,j=1}^{N} w_{ij} (x_{ij} - \mu_X)(y_{ij} - \mu_Y) \tag{8}$$

and μ_Y and σ_Y^2 have the similar definitions as Eq. (6)-(7).

Then, the SSIM_MAP is defined as the product of the luminance, contrast and structural similarities:

$$\begin{aligned}
&SSIM_MAP(X,Y,N) \\
&\qquad = l(X,Y,N) \cdot c(X,Y,N) \cdot s(X,Y,N).
\end{aligned} \tag{9}$$

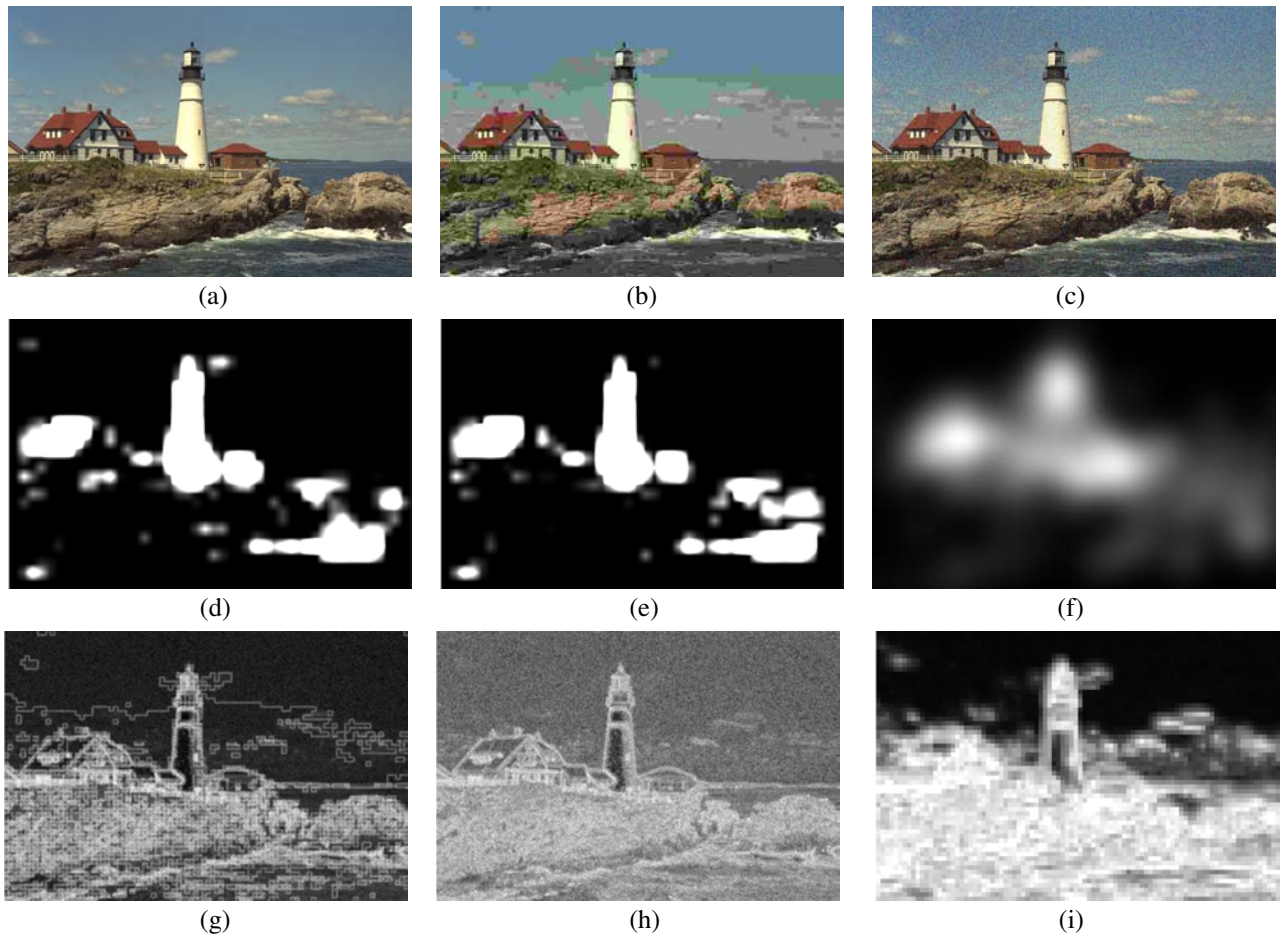

Fig. 2. The first example of comparison of different saliency maps (the whiter the regions are, the larger the saliency is): (a) Reference image; (b) JPEG compressed image; (c) White noise image; (d) $S_N W$ map [9] of (b); (e) $S_N W$ map of (c); (f) Eye-tracking based visual attention map [8] of (b)-(c); (g) IW map [10] of (b); (h) IW map of (c); (i) The proposed SW map of (b)-(c). Notice that eye-tracking based visual attention map and our SW map have the same results for (b) and (c) due to their common reference image.

And the corresponding SSIM index evaluating the overall image quality is defined by

$$SSIM(X,Y,N) = \frac{1}{P}\sum_{p=1}^{P} SSIM\_MAP(x_p, y_p, N) \quad (10)$$

with P being the number of local windows in the image.

Besides, by first downsampling reference image R and distorted image D with scale transform coefficient Z followed by SSIM metric, a simple and empirical method [13] is also took into account:

$$SSIM_Z = SSIM(F_Z(R), F_Z(D), N') \quad (11)$$

where $F_Z(\cdot)$ is a downsampling function given by [13], and its scale Z can be evaluated by

$$Z = max(1, round(min(H,W)/256)) \quad (12)$$

with H and W being the height and width of the reference image, respectively.

Eventually, the proposed SW-SSIM metric is given by applying SW map weighting SSIM_MAP:

$$\begin{aligned} &SW\text{-}SSIM(F_Z(R), F_Z(D)) \\ &\quad = SW\text{-}SSIM(R', D') \\ &\quad = \frac{\sum_{i,j} SW_{i,j} \cdot SSIM\_MAP(r'_{i,j}, d'_{i,j}, N')}{\sum_{i,j} SW_{i,j}} \end{aligned} \quad (13)$$

where

$$SW_{i,j} = 1 - \nu(\lceil i/M \rceil, \lceil j/M \rceil, N) \quad (14)$$

satisfying that $m = \lceil i/M \rceil$ and $n = \lceil j/M \rceil$.

All the parameters used in our paradigm are provided as follows: $\nu_s(k = 1...4) = 1$, $\nu_s(k = 5...8) = 0.25$,$N = 7$, $N' = 11$, and $M = 4$.

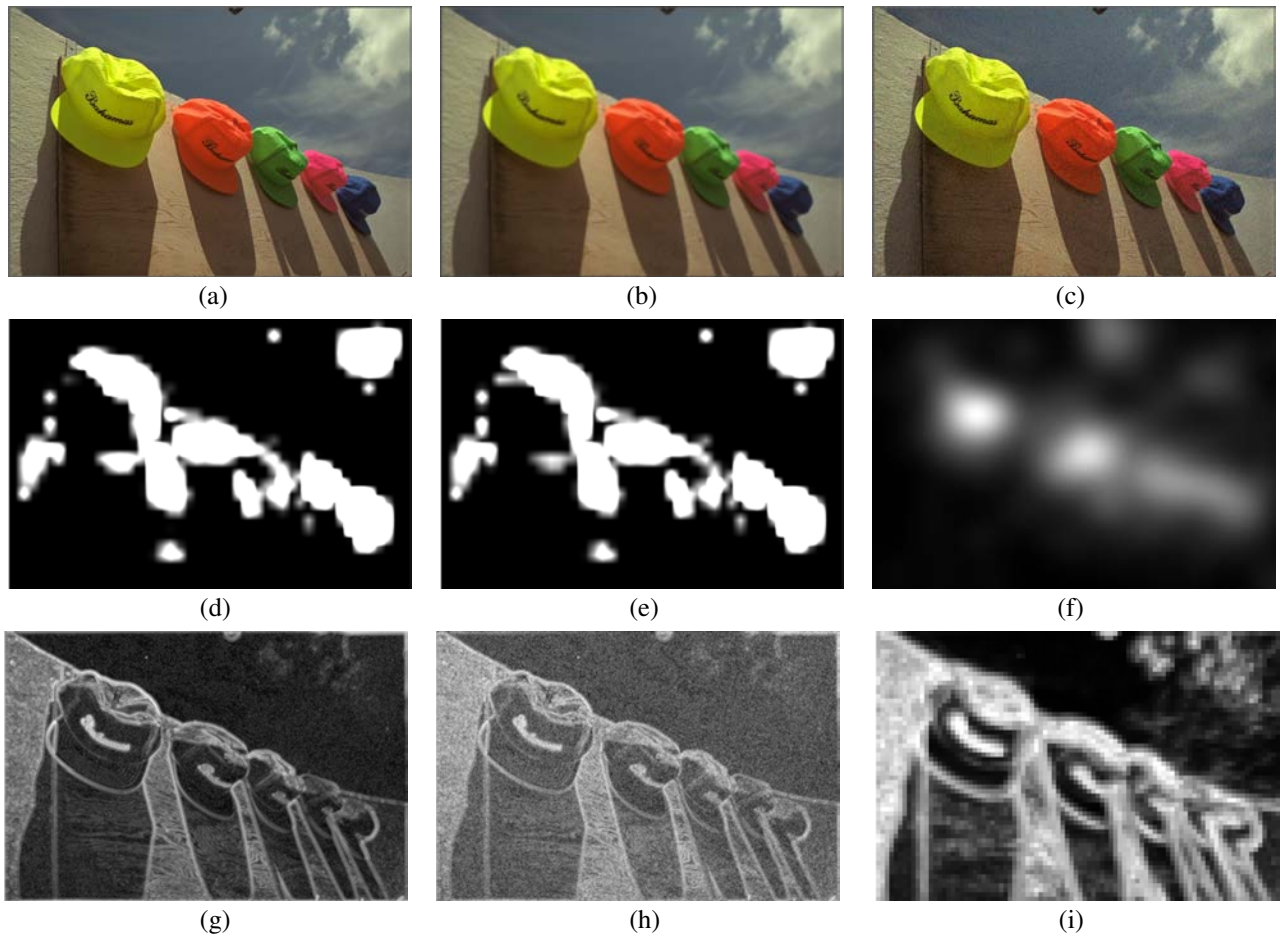

Fig. 3. The second example of comparison of different saliency maps (the whiter the regions are, the larger the saliency is): (a) Reference image; (b) Gaussian blur image; (c) White noise image; (d) S_NW map [9] of (b); (e) S_NW map of (c); (f) Eye-tracking based visual attention map [8] of (b)-(c); (g) IW map [10] of (b); (h) IW map of (c); (i) The proposed SW map of (b)-(c). Notice that eye-tracking based visual attention map and our SW map have the same results for (b) and (c) due to their common reference image.

2.2. Comparison of Different Weighting Strategy

We have illustrated in Fig. 2-3 the results of different weighting strategy for two original images each subject to two different types of distortion (JPEG compression and white noise injection in Fig. 2 and Gaussian blur and white noise distortion in Fig. 3). It can be observed that the S_NW saliency map (Fig. 2-3 (d)-(e)) using a classical bottom-up saliency model [11] cannot effectively capture the most distortion regions. Still, it is noticed that the visual attention map (Fig. 2-3 (f)) recorded from eye-tracking experiments [8] still cannot catch the most distorted areas. On the other hand, the VIF metric inspired IW map (Fig. 2-3 (g)-(h)) in [10] can capture part of the most distorted regions. And our proposed SW map (Fig. 2-3 (i)) indeed apprehends most of the apparent distorted regions and is therefore expected to improve the IQA more effectively. For instance, see one obvious example in the right-middle part in Fig. 2 (b). Only IW-SSIM and our proposed metric are capable of detecting the distortions, as shown in Fig. 2 (g), (i).

Table 1. Comparison of different scores for Fig. 2-3 (b)-(c) using subjective and objective pooling-type of methods, including the results of DMOS, S_NW-SSIM, WSSIM, IW-SSIM, SW-SSIM, and the most two popular PSNR and SSIM methods.

Comparison of Different Scores				
Algorithm	Fig. 2 (b)	Fig. 2 (c)	Fig. 3 (b)	Fig. 3 (c)
DMOS	**78.168**	**55.536**	**57.781**	**51.670**
PSNR	24.906	21.977	29.726	27.619
SSIM	0.8351	0.6585	0.9192	0.8019
S_NW-SSIM	0.8675	0.8190	0.9061	0.8594
WSSIM	0.7036	0.4586	0.8595	0.5284
IW-SSIM	0.8520	0.8294	0.9328	0.9014
SW-SSIM	**0.7936**	**0.8191**	**0.8620**	**0.8742**

Since the HVS is not working in a pixel by pixel manner in accessing image qualities, the proposed block-based SW-SSIM metric outperforms the point-wise metric of IW-SSIM.

Table 1 also tabulates some quality scores for Fig. 2-3 (b)-(c), including the above-mentioned pooling-type of algorithms, differential mean opinion scores (DMOS), and the popular PSNR and SSIM methods. Among the IQA methods in the test, only our proposed metric gives Fig. 2 (c) a score of higher than Fig. 2 (b), which corresponds to the fact that Fig. 2 (b) has larger DMOS than Fig. 2 (c) (i.e. the image quality of Fig. 2 (c) is higher than that of Fig. 2 (b)). Similar example can be found in Fig. 3 (b) and (c). Note that we only consider the comparison between Fig. 2 (b) and (c) (or Fig. 3 (b) and (c)), because the distorted images in each pair are from the same reference image, which can overcome the influence of different image contents on the accuracy of IQA metrics. So, we have a reason to believe that our proposed method is further more correlated with the DMOS than all other IQA metrics used in our test.

3. EXPERIMENTAL RESULTS

Mappings of the scores of eight metrics PSNR, SSIM [1], SSIM' [13], MS-SSIM, S_NW-SSIM, WSSIM, IW-SSIM and the proposed SW-SSIM methods to subjective scores are obtained using nonlinear regression with a four-parameter logistic function as suggested by VQEG [14]:

$$q(\varepsilon) = \frac{\gamma_1 - \gamma_2}{1 + exp(-\frac{(\varepsilon - \gamma_3)}{\gamma_4})} + \gamma_2 \qquad (15)$$

with ε and $q(\varepsilon)$ being the input score and the mapped score, respectively. The free parameters γ_1 to γ_4 are to be determined during the curve fitting process.

This paper applies three commonly used performance metrics, Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-Order Correlation Coefficient (SROCC) and Root Mean-Squared Error (RMSE), to further compare the competitive SW-SSIM metric and the other seven methods (i.e., PSNR, SSIM, SSIM', MS-SSIM, S_NW-SSIM, WSSIM, and IW-SSIM) on the LIVE database. Table 2-4 illustrate their performance results, and the scatter plots of SW-SSIM on five different distortion categories of data sets and the whole LIVE database are shown in Fig. 4. Clearly, our algorithm attains better result than most of the existing pooling-type of IQA methods used in our research.

Table 2. Pearson Linear Correlation Coefficient (PLCC) values (after nonlinear regression) of PSNR, SSIM, SSIM', MS-SSIM, S_NW-SSIM, WSSIM, IW-SSIM and the proposed SW-SSIM methods on whole LIVE database (779 images) and five data sets of different distortion categories.

Pearson Linear Correlation Coefficient (PLCC)						
Algorithm	JP2K	JPEG	WN	Gblur	FF	All
PSNR	0.899	0.887	**0.985**	0.795	0.889	0.870
SSIM [1]	0.934	0.947	0.964	0.907	0.942	0.901
SSIM' [13]	0.966	0.978	0.969	0.946	0.949	0.938
MS-SSIM	0.969	0.981	0.972	0.953	0.920	0.933
S_NW-SSIM	0.962	0.978	0.973	0.948	0.938	0.941
WSSIM	0.959	0.967	0.974	0.926	0.954	0.921
IW-SSIM	0.971	0.981	0.968	0.962	0.931	0.942
SW-SSIM	**0.974**	**0.981**	0.972	**0.970**	**0.955**	**0.950**

Table 3. Spearman Rank-Order Correlation Coefficient (SROCC) values (after nonlinear regression) of PSNR, SSIM, SSIM', MS-SSIM, S_NW-SSIM, WSSIM, IW-SSIM and the proposed SW-SSIM methods on whole LIVE database (779 images) and five data sets of different distortion categories.

Spearman Rank-Order Correlation Coefficient (SROCC)						
Algorithm	JP2K	JPEG	WN	Gblur	FF	All
PSNR	0.895	0.880	**0.985**	0.783	0.890	0.875
SSIM [1]	0.935	0.944	0.962	0.894	0.941	0.910
SSIM' [13]	0.961	0.976	0.969	0.952	0.954	0.947
MS-SSIM	0.965	0.979	0.972	0.958	0.931	0.944
S_NW-SSIM	0.958	0.975	0.980	0.948	0.950	0.952
WSSIM	0.953	0.963	0.968	0.935	0.956	0.929
IW-SSIM	0.964	0.980	0.966	0.972	0.944	0.956
SW-SSIM	**0.967**	**0.982**	0.980	**0.972**	**0.963**	**0.961**

Table 4. Root Mean-Squared Error (RMSE) values (after nonlinear regression) of PSNR, SSIM, SSIM', MS-SSIM, S_NW-SSIM, WSSIM, IW-SSIM and the proposed SW-SSIM methods on whole LIVE database (779 images) and five data sets of different distortion categories.

Root Mean-Squared Error (RMSE)						
Algorithm	JP2K	JPEG	WN	Gblur	FF	All
PSNR	11.01	14.65	**4.718**	11.44	13.03	13.46
SSIM [1]	8.534	9.907	6.845	8.964	9.496	11.83
SSIM' [13]	6.473	6.502	6.811	6.115	8.964	9.445
MS-SSIM	6.363	5.999	6.569	5.855	9.703	9.306
S_NW-SSIM	6.889	6.552	6.460	5.876	9.828	9.219
WSSIM	7.079	8.029	6.342	6.961	8.538	10.59
IW-SSIM	6.034	6.111	6.931	5.148	10.41	9.131
SW-SSIM	**5.758**	6.117	6.596	**4.651**	**8.468**	**8.557**

Besides, we want to emphasize that the proposed SW-SSIM method is more valid for the current hot topic of Management Information Systems (MIS), especially for video coding. With the recent emergence of high-definition/ultra-high-definition TV programs, human beings have succeed in pushing the multimedia entertainment to a new level. However, in the meantime, more powerful techniques for video coding, quality assessment and etc are highly required. Note, SW-SSIM is performing in a blockwise manner, which makes it suitable for parallel processing to achieve high executive speed. And moreover, our metric has very high portability as it is solely based on SSIM that has been inserted into many existing video processing systems. By taking the remarkable superior performance of SW-SSIM into consideration, we therefore believe that our approach has great potentials to be directly or indirectly used in the video quality assessment of high-definition/ultra-high-definition TV programs.

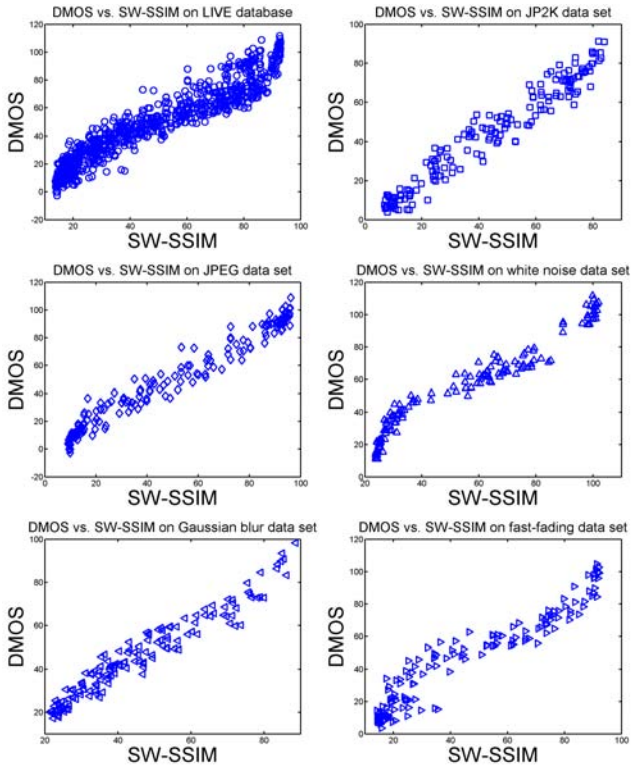

Fig. 4. Scatter plots of DMOS vs. SW-SSIM on all the LIVE database and five various distortion types of data sets.

4. CONCLUSION

In this paper, we propose a new weighting/pooling strategy for IQA based on local structural similarity measure computed using the IQA itself. We further develop a Structural similarity Weighting SSIM (SW-SSIM) image quality metric with prediction accuracy outperforming most of existing weighting/pooling type of IQA algorithms on the LIVE database. We want to point out that our method is quite fitted to be applied in multimedia entertainment, one major direction of management information systems, due to its high prediction accuracy, parallel operation of the blockwise manner, and easy portability in video processing system. In addition, note that the work in this paper is merely an easy realization of a more general framework, namely locally weighing IQA metrics using IQA metrics themselves. Other more delicate combinations are expected to bring even higher performance improvement. The research on this direction is largely warranted not only by the simplicity and elegancy of the proposed framework, but also by the encouraging performance gain obtained in the tests reported in this paper.

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).

5. REFERENCES

[1] 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.

[2] 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.

[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] G. H. Chen, C. L. Yang, and S. L. Xie, "Gradient-based structural similarity for image quality assessment," *IEEE ICASSP*, pp. 2929-2932, 2006.

[5] D. Rao and L. Reddy, "Image quality assessment based on perceptual structural similarity," *Pattern Recognition and Machine Intelligence*, pp. 87-94, 2007.

[6] A. K. Moorthy and A. C. Bovik, "Visual Importance Pooling for Image Quality Assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 2, pp. 193-201, April 2009.

[7] C. Li and A. C. Bovik, "Content-partitioned structural similarity index for image quality assessment," *Signal Processing: Image Communication*, vol. 25, no. 7, pp. 517-526, August 2010.

[8] H. Liu and I. Heynderickx, "Visual attention in objective image quality assessment: Based on eye-tracking data," *IEEE Transaction on Circuits and System for Video Technology*, vol. 21, no. 7, pp. 971-982, April 2011.

[9] K. Gu, G. Zhai, X. Yang, L. Chen, and W. Zhang, "Nonlinear additive model based saliency map weighting strategy for image quality assessment", *IEEE International Workshop on Multimedia Signal Processing*, 2012.

[10] 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.

[11] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 20, pp. 1254-1259, November 1998.

[12] 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.

[13] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, The ssim index for image quality assessment. <http://ece.uwaterloo.ca/∼z70wang/research/ssim/>.

[14] 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/.