NONLINEAR ADDITIVE MODEL BASED SALIENCY MAP WEIGHTING STRATEGY FOR IMAGE QUALITY ASSESSMENT

Ke Gu, Guangtao Zhai, Xiaokang Yang, Li Chen 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

Most state-of-the-art image quality metrics are based on the two-step approach: local distortion/fidelity measurement and pooling. During the pooling stage, many weighting strategies have been proposed incorporating properties of the distortion itself, various masking effects and visual attention. Recently, researchers have devoted great enthusiasm and effort to the improvement of image quality assessment using visual saliency models. In this research, it is noticed that visual saliency features of both the original image and the distorted one have impacts on the process of image quality assessment. To reduce the overlapping effects, a nonlinear additive model is proposed to integrate saliency features from the original and distorted images towards improved error weighting results. Our extensive experimental studies on four publicly available image databases (LIVE, TID2008, CSIQ and A57) indicate that the proposed improved nonlinear additive model based saliency map weighting strategy constantly leads to higher prediction accuracy for image quality assessment than traditional methods.

Index Terms—Image quality assessment (IQA), visual attention (VA), saliency map, nonlinear additive model

1. INTRODUCTION

Perceptual image quality assessment (IQA) plays an important part in many areas of digital image processing, such as the development and optimization of image compression, storage and transmission algorithms. Existing IQA approaches fall into two categories: subjective assessment and objective assessment. Although the subjective assessment approach should be the ultimate quality gauge for digital images, it is usually time-consuming, expensive and not practical for real-time image processing systems. The Mean-Squared Error (MSE) and its relative the Peak Signal-to-Noise Ratio (PSNR) are still the most widely used objective quality metrics, both due to their convenience and due to their clear physical meaning as distortion/fidelity measures. However, it has been widely recognized that MSE and PSNR are not well correlated with human judgment of quality, i.e. the Mean Opinion Score (MOS). Objective image quality metrics based on the human visual system (HVS) are potentially more reliable for accurate quality prediction [1], for example, the well-known quality metrics of SSIM [2] and VIF [3]. It has been demonstrated that visual attention model based on combining some low-level aspects of the HVS (e.g. frequency sensitivity, luminance masking and texture masking) can improve the performance of IQA methods and this topic has been a very active research area [4]-[7].

Intuitively, the distortion in prominent area (a saliency region) attracts more of the viewer's attention, and is more annoying than those in any other areas. This idea has been exploited in [4]-[7], where the performances of the image quality metrics are improved by weighting the measured local distortions with the saliency map. Understandably, severe distortion usually hinders accurate saliency feature detection so most saliency map weighted IQA methods (e.g. [5]-[7]) choose to work only with saliency map computed on the distortion-free reference image. However, those methods are clearly based on the assumption that the saliency features in the original and distorted images are similar to each other. Admittedly, when the distorted image is of high quality, the above assumption is almost valid. On the other hand, as the quality of the distorted image drops, the noticeable artifacts may dominate the saliency detection process and results in a discrepancy between the saliency features of the reference and the distorted images. Therefore, an accurate weighting strategy should be based on saliency features not only from the reference image but also the distorted one. Moreover, inspired by former works on masking effect estimation, a nonlinear additive model [9]-[10] is used in this research to account for the possible overlapping effects between the saliency maps generated by the bottom-up attention model [8] on the original and distorted images.

The remainder of this paper is organized as follows. Section 2 first reviews a classic bottom-up saliency model and points out the shortage of applying the model on the original image alone in error pooling for image quality assessment. Then we outline the saliency model based weighting strategies using linear/nonlinear additive models. Section 3 introduces the linear/nonlinear additive model based IQA algorithms. In Section 4, experimental results using the LIVE database [11], TID2008 [12], CSIQ database [13] and A57 database [14] are reported and the results are analyzed. Finally, conclusion is drawn and future work direction is pointed out in Section 5.

This work is supported by National Nature Science Foundation under Grant 60932006 and 61102098.

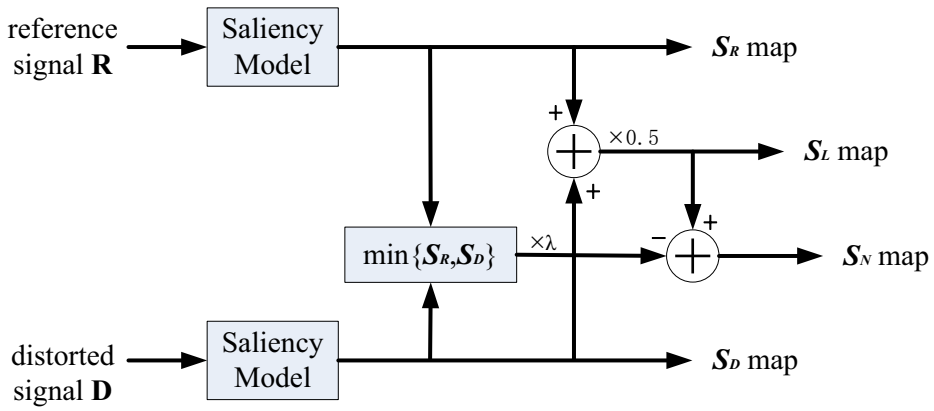


Figure 1. Diagram for computing saliency maps (SR map: saliency map for the reference image, SD map: saliency map for the distorted image, SL map: saliency map based on SR and SD maps with linear additive model, SN map: saliency map based on SR and SD maps with nonlinear additive model)

2. LINEAR/NONLINEAR ADDITIVE MODEL BASED SALIENCY WEIGHTING

2.1. Bottom-up Saliency Model

Visual attention model is an important basic research area for image quality assessment. Visual attention models generally fall into two categories: top-down and bottom-up approaches. The top-down method is usually driven by a certain task when viewing a scene, such as searching for a specific target. So, top-down attention models are usually based on visual feature detection that is correlated with such tasks. In the bottom-up methods, a computational model for identifying visual attention regions is developed using low-level features of visual signals.

In this study, we employed a classic bottom-up visual attention model [8]. Inspired by the behavior and the neuronal architectures of the early primate visual system, this saliency model is to construct a single topographical saliency map first by combining multi-scale image features, such as colors, intensity, orientations and other visual information. Then, a winner-take-all network that implements a neutrally distributed maximum detector is performed to detect the most salient locations step by step. Finally, a saliency map of an image can be computed, which can depict the saliency distribution over different locations. Fig. 2 (c)-(d) illustrate saliency maps of reference and distorted images for JPEG distortion, and Fig. 3 (c)-(d) are results for Fast-fading distortion.

2.2. Linear/Nonlinear Additive Model

Most of saliency weighted IQA methods just computed the saliency map of reference images. As we discussed, when the reference and distorted images are of comparable quality, a weighting strategy based on only the saliency map from the original reference image is very valid. However, this approach tends to fail as the quality level of the distorted image drops. When the quality of the distorted image is quite low, as illustrated in Fig. 2-3 (c)-(d), saliency maps of reference image and distorted image can be very different. Considering the fact that it is the distorted images that are delivered to the viewers in typical visual communication systems, for an accurate saliency map weighting strategy, saliency features from both the original and the distorted image should be considered. According to the analysis above and for the sake of comparison, a linear additive model shown in Fig. 1 is used to incorporate both saliency maps. The resulting overall saliency map based on linear additive model (S_L) is simply given by:

$$S_L = \frac{S(x)+S(y)}{2} = \frac{S_R + S_D}{2} \quad (1)$$

where x and y indicate reference and distorted images, $S(x)$ represents the saliency map of x. S_R and S_D are saliency maps of reference and distorted images respectively.

Although the S_L map contains salient regions from both saliency maps S_R and S_D, we can find a serious problem introduced by the straightforward linear additive model: saliency regions shared by both the reference and the distorted images are excessively emphasized while regions belonging to single image are overly lessened. For example, the individual saliency regions in red circle of Fig. 2-3 (c)-(d) are lessened in Fig. 2-3 (e), while the same saliency regions in green circle of Fig. 2-3 (c)-(d) are enhanced in Fig. 2-3 (e). According to the working mechanism of saliency formation in V1 area, the retina and LGN, orientation, motion and luminance (the former two are clearly coded by V1 cells and the last one is controlled by activated cells in the retina and LGN [9]) all play significant roles in forming the final saliency map. By carrying out three experiments on the combination of orientation and motion contrast, combination of orientation or motion and luminance contrast and combinations of orientation or motion and color contrast in [9], Northdurft gave two conclusions: First, the underlying neural processes are far more independent of orientation or motion and luminance than the case for the

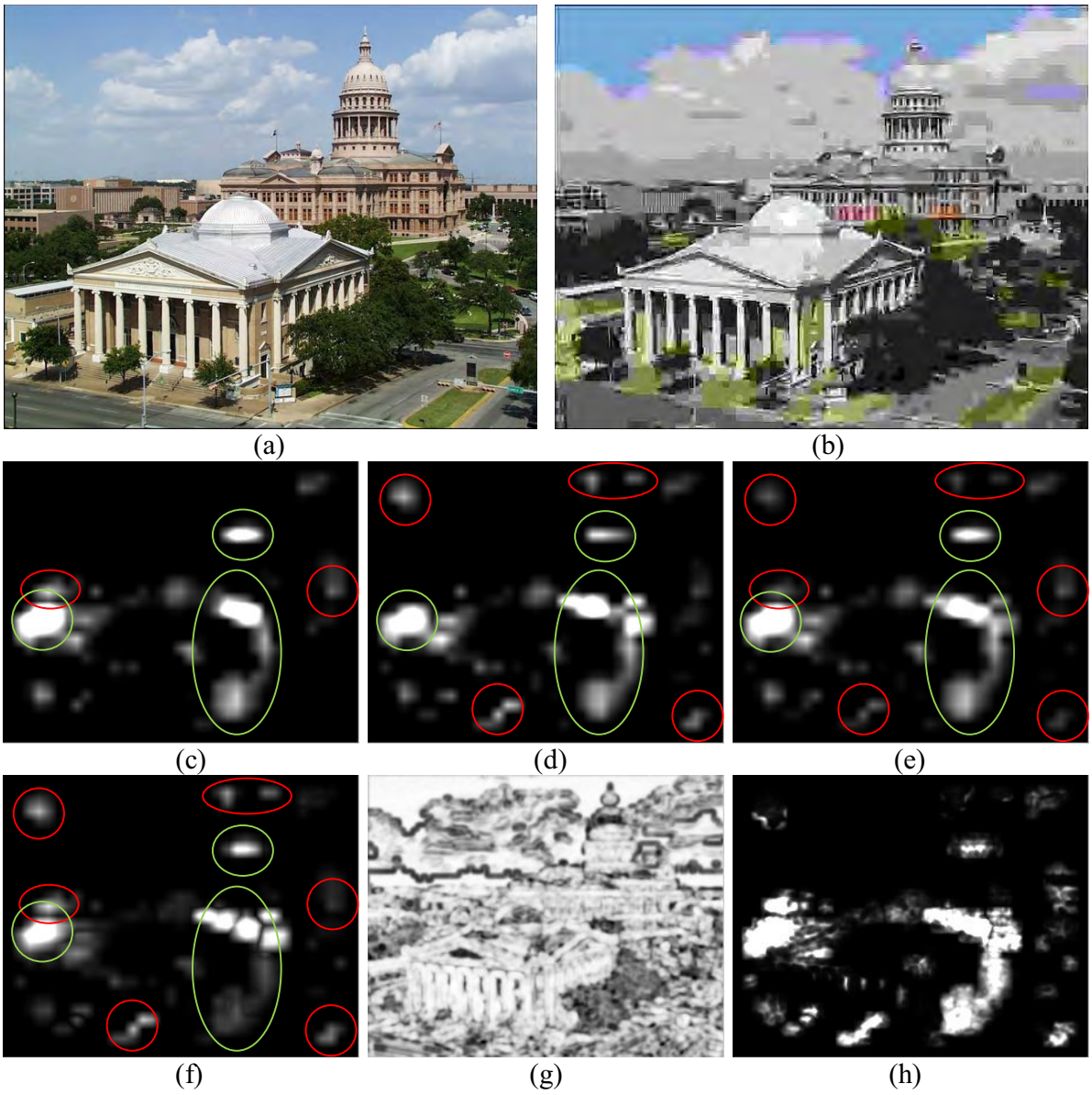


Figure 2. Illustration of saliency maps (more whiter more saliency) and ssim_map (more blacker more distorted): (a) reference image; (b) JPEG distorted image with its dmos value 97.2468; (c) saliency map of reference image; (d) saliency map of distorted image; (e) saliency map based on linear additive model; (f) saliency map based on nonlinear additive model; (g) ssim_map; (h) saliency map (f) weighted ssim_map

combinations of orientation and motion; Second, the additive effects of combination of color and motion are pronounced but clearly different from linear summation. Those conclusions suggest that overlapping effects should exist in the saliency maps of the reference and the distorted images. To diminish the effects of the common saliency regions in the reference and the distorted images, we can penalize the regions accordingly by subtracting a local minimum between the saliency estimation. Here, as illustrated in Fig. 1, we adopts the so called nonlinear additive model based saliency map (S_N) defined as follows

$$S_N = \frac{S(x)+S(y)}{2} - \lambda \bullet min\{S(x), S(y)\}$$

$$= \frac{S_R + S_D}{2} - \lambda \bullet min\{S_R, S_D\} \quad (2)$$

where λ is an parameter adjusting the penalization strength. Clearly, value for λ depends on the level of saliency feature shared between the reference and the distorted images. In

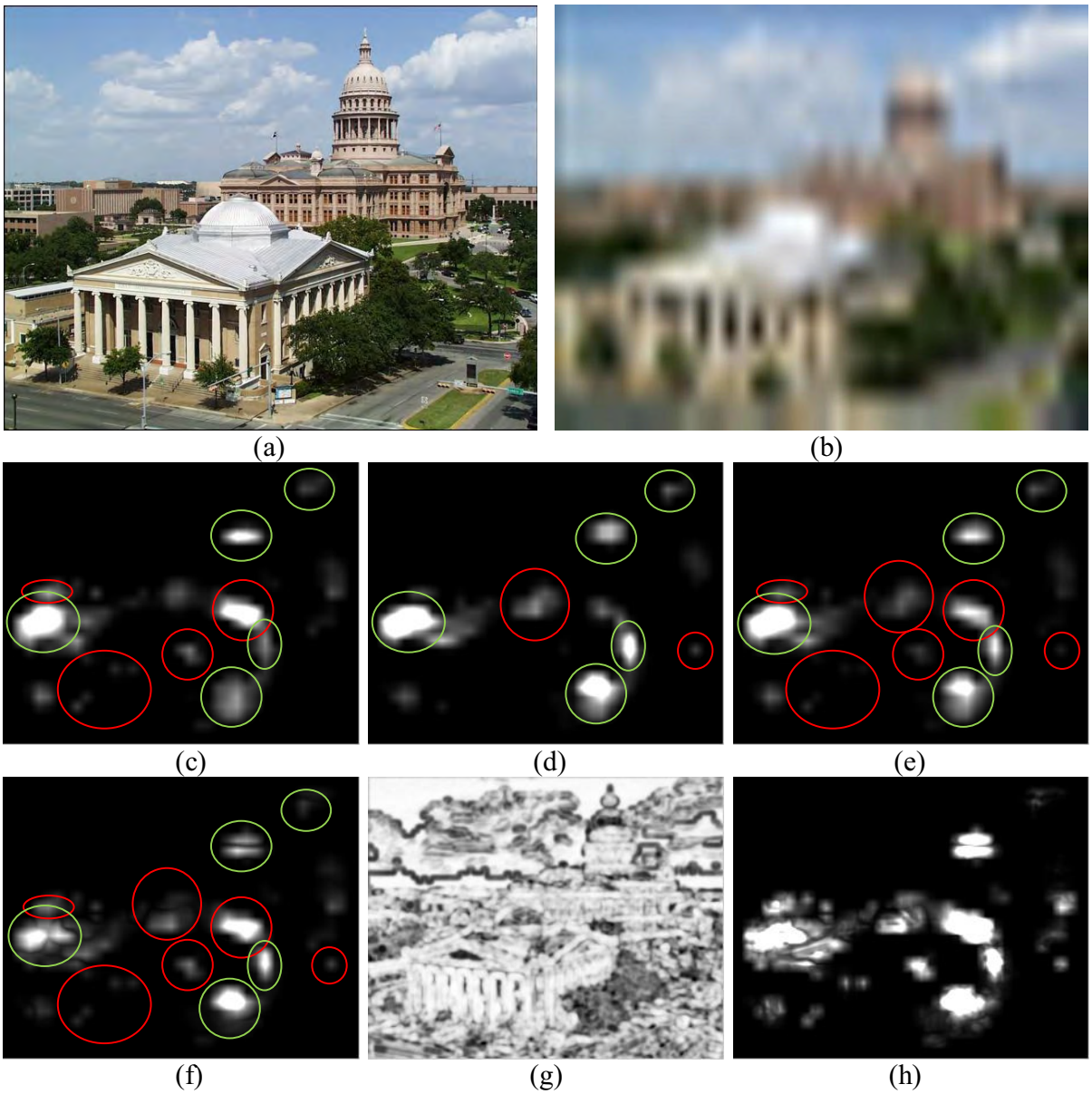


Figure 3. Illustration of saliency maps (more whiter more saliency) and ssim_map (more blacker more distorted): (a) reference image; (b) Fast-fading distorted image with its dmos value 101.3192; (c) saliency map of reference image; (d) saliency map of distorted image; (e) saliency map based on linear additive model; (f) saliency map based on nonlinear additive model; (g) ssim_map; (h) saliency map (f) weighted ssim_map

our research, it is noticed that the optimal value of λ is related to the type of image distortion. Table I lists the optimal values for λ, obtained by our extensive tests for distortions of JP2K, JPEG, WN, Gblur and Fast-fading in LIVE database in terms of the correlation between IQA scores and subjective scores. To simplify our algorithms, in the later experiments, we set λ=0.45 for all type of distortions in this research. The nonlinear additive model based saliency detection results shown in Fig. 2-3 (f) are found to be more reliable and more correlated to our experience than the results of S_R, S_D and S_L.

TABLE I. λ VALUES AND THEIR CORRESPONDING PLCC AND SROCC OF JP2K, JPEG, WN, GBLUR AND FAST-FADING DISTORTED IMAGES

DISTORTION (image numbers)	λ	PLCC	SROCC
JP2K (169)	0.50	0.9650	0.9613
JPEG (175)	0.39	0.9790	0.9760
WN (101)	0.44	0.9801	0.9971
GBLUR (101)	0.50	0.9517	0.9499
FAST-FADING (101)	0.27	0.9393	0.9516

3. IMPROVING SALIENCY MAP WEIGHTING IMAGE QUALITY METRIC

With our linear/nonlinear additive model based saliency maps, we can immediately weight the local distortion map, and then the final objective image quality scores can be computed by

$$Q = \frac{\sum_i^N \omega(i) q(i)}{\sum_i^N \omega(i)} \tag{3}$$

where $\omega(i)$ is the value of saliency map, $q(i)$ indicates the value of local distortion map, and N represents the size (number of pixels) of the local distortion map.

Here we use SSIM [2] as a tool for measuring local distortion, which is defined by:

$$SSIM = \frac{1}{M}\sum_{i=1}^{M} ssim\_map(x, y) \tag{4}$$

where x and y are the reference and distorted images, and M is the size of the image. The saliency map $\{\omega(i)\}$, created by computing saliency map at each image location, can be used as a weighting function for SSIM pooling. Thus, we define S_R weighted SSIM (S_RW-SSIM) as follows

$$S_R W\text{-}SSIM = \frac{\sum_i^N \omega^{S_R}(x_i, y_i) ssim\_msp(x_i, y_i)}{\sum_i^N \omega^{S_R}(x_i, y_i)} \tag{5}$$

and similarly, S_D weighted SSIM (S_DW-SSIM), S_L weighted SSIM (S_LW-SSIM), and S_N weighted SSIM (S_NW-SSIM) are given by:

$$S_D W\text{-}SSIM = \frac{\sum_i^N \omega^{S_D}(x_i, y_i) ssim\_msp(x_i, y_i)}{\sum_i^N \omega^{S_D}(x_i, y_i)} \tag{6}$$

$$S_L W\text{-}SSIM = \frac{\sum_i^N \omega^{S_L}(x_i, y_i) ssim\_msp(x_i, y_i)}{\sum_i^N \omega^{S_L}(x_i, y_i)} \tag{7}$$

$$S_N W\text{-}SSIM = \frac{\sum_i^N \omega^{S_N}(x_i, y_i) ssim\_msp(x_i, y_i)}{\sum_i^N \omega^{S_N}(x_i, y_i)} \tag{8}$$

where ω^{S_R}, ω^{S_D}, ω^{S_L} and ω^{S_N} indicate the values of S_R, S_D, S_L and S_N map.

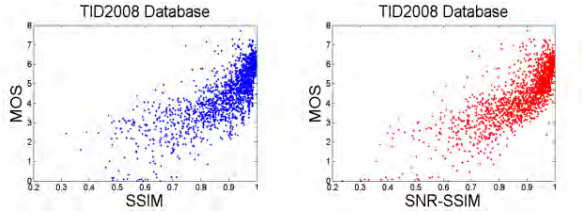
4. EXPERIMENTAL RESULTS

Mappings of the scores of these five metrics SSIM, S_DW-SSIM, S_RW-SSIM, S_LW-SSIM and S_NW-SSIM to subjective scores are obtained using nonlinear regression with a 4-parameter logistic function as suggested by VQEG [15]:

$$q(x) = \frac{\beta_1 - \beta_2}{1 + exp\left(-(x - \beta_3)/\beta_4\right)} + \beta_2 \tag{9}$$

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.

Two commonly used performance metrics, Pearson Linear Correlation Coefficient (PLCC) and Spearman Rank-order Correlation Coefficient (SRCC) as suggested by VQEG [15], are employed to further evaluate SSIM and competitive S_RW-SSIM, S_DW-SSIM, S_LW-SSIM and S_NW-SSIM metrics on LIVE, TID2008, CSIQ and A57. Fig. 4 shows the scatter plots of MOS vs. SSIM/S_NW-SSIM on TID2008 database and all the PLCC and SRCC values are illustrated in Fig. 5, 6, 7 and 8 respectively.



(a) (b)

Figure 4. (a): Scatter plots of MOS vs. SSIM on TID2008 database; (b): Scatter plots of MOS vs. S_NW-SSIM on TID2008 database.

We have some valuable observations on the performances of the IQA algorithms: First, S_DW-SSIM has higher prediction accuracy than S_RW-SSIM, and this coincides with the fact that subjective scores mainly depend on the injuring degree of distorted images (i.e. saliency maps of distorted images play more important roles). Second, S_LW-SSIM and S_NW-SSIM generally have much better performance because of using both saliency maps. Third, S_NW-SSIM performs the best of all, due to the fact that $\boldsymbol{S_N}$ map is based on combining saliency maps of reference and distorted images as well as reducing the overlapping effects in both saliency maps.

5. CONCLUSION

In this paper, we first point out the limits of the present saliency map weighted IQA methods in considering only saliency features from the reference image. Since image observation process is usually conducted on the distorted image in the real world, saliency features from the distorted image should also be explored. And to combat the possible overlapping effects between the common saliency features between the original and the distorted images, a nonlinear additive model approach is used in combining the saliency maps from the original and the distorted images. We have tested the proposed improved saliency map weighting strategy for the IQA algorithm of SSIM on LIVE, TID2008, CSIQ and A57 databases. The experimental results verify that the linear/nonlinear additive model based weighting strategies outperform methods using only the original or the distorted image while the nonlinear additive model based weighting method has the best result.

As analyzed, optimal weighting parameter $\boldsymbol{\lambda}$ in the nonlinear additive model is closely related to different kind of distortion, so our future work will involve the distortion

classification towards even higher performance improvement for weighting IQA algorithms using saliency maps.

6. REFERENCES

[1] S. Winkler, "Vision models and quality metrics for image processing applications," Ph.D. dissertation, Dept. Elect., EPFL, Lausanne, 2002.

[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] H. R. Sheikh and A. C. Bovik, "Image information and visual quality," IEEE Transactions on Image Processing, vol.15, no.2, pp. 430–444, February 2006.

[4] R. Barland and A. Saadane, "Blind quality metric using a perceptual importance map for JPEG-2000 compressed images," in Proc. IEEE International Conference of Image Processing, pp. 2941–2944, October 2006.

[5] A. K. Moorthy and A. C. Bovik, "Perceptually significant spatial spooling techniques for image quality assessment," in Proc. Electronic Imaging, 2009.

[6] Junyong You, Perkis A. and Gabbouj, M., "Improving image quality assessment with modeling visual attention," IEEE European Workshop on Visual Information Processing, July 2010.

[7] H. Liu and Ingrid 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.

[8] 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.

[9] H. C. Nothdurft, "Salience from feature contrast: additivity across dimensions," Vision Research, vol. 40, pp. 1183–1201, 2000.

[10] Yang X. K., Lin W. S., Lu Z. K., Ong E. P. and Yao S. S., "Just-noticeable-distortion profile with nonlinear additivity model for perceptual masking in color images," IEEE International Conference on ICASSP, April 2003.

[11] 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/

[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 Radioelectronics, Vol. 10, pp. 30–45, 2009.

[13] E. C. Larson and D. M. Chandler, "Categorical image quality (CSIQ) database," [Online]. Available: http://vision.okstate.edu/csiq

[14] D. M. Chandler and S. S. Hemami, "VSNR: A wavelet-based visual signal-to-noise ratio for natural images," IEEE Transaction on Image Processing. [Online]. Available: http://foulard.ece.cornell.edu/dmc27/vsnr/vsnr.html

[15] 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/

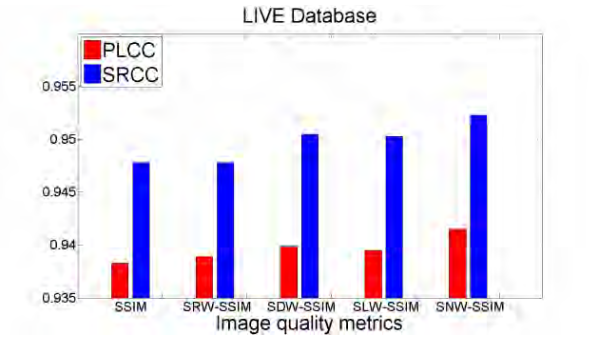


Figure 5. Correlation coefficients of five metrics SSIM, S_RW-SSIM, S_DW-SSIM, S_LW-SSIM and S_NW-SSIM for LIVE database

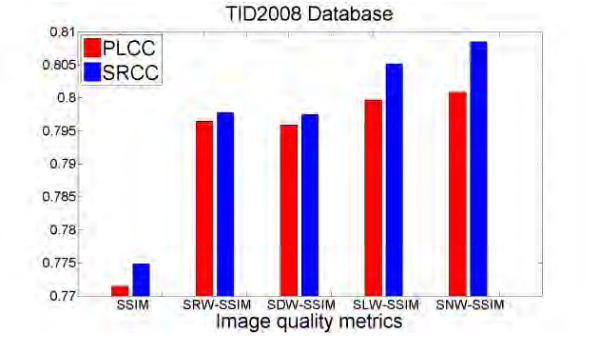


Figure 6. Correlation coefficients of five metrics SSIM, S_RW-SSIM, S_DW-SSIM, S_LW-SSIM and S_NW-SSIM for TID2008 database

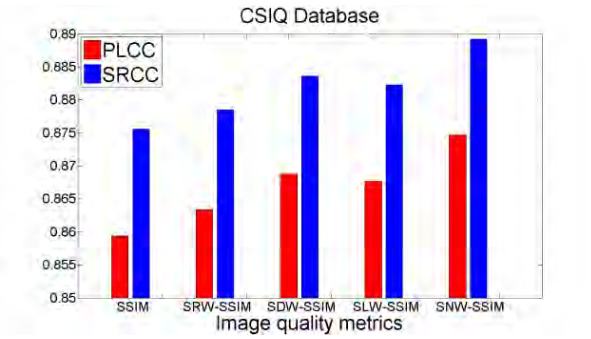


Figure 7. Correlation coefficients of five metrics SSIM, S_RW-SSIM, S_DW-SSIM, S_LW-SSIM and S_NW-SSIM for CSIQ database

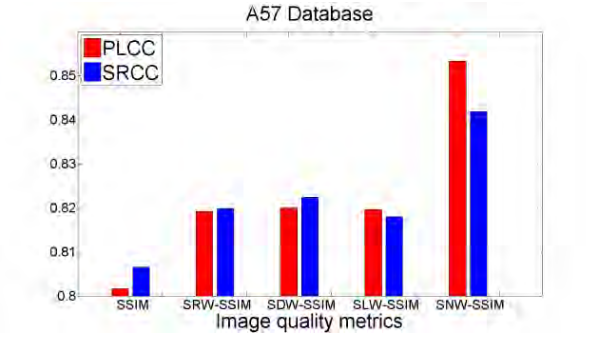


Figure 8. Correlation coefficients of five metrics SSIM, S_RW-SSIM, S_DW-SSIM, S_LW-SSIM and S_NW-SSIM for A57 database