Self-adaptive scale transform for IQA metric

Ke Gu, *Student Member, IEEE*, Guangtao Zhai, *Member, IEEE*,
Xiaokang Yang, *Senior Member, IEEE*, and Wenjun Zhang, *Fellow, IEEE*

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

***Abstract*—Recently, an increasing number of image quality assessment (IQA) algorithms have been developed based on multi-scale methods, such as MS-SSIM, IFC, VIF and IW-PSNR/SSIM. Inspired by the achievement of multi-scale type of IQA algorithms, this paper proposes a self-adaptive scale transform based IQA approach. Using image size and viewing distance as input variables, we construct a self-adaptive scale transform function to estimate the suitable scale transform coefficient for the following image quality metrics. Two of the most well-known full-reference IQA methods (PSNR and SSIM), and three publicly-available subjectrated image databases (LIVE, IVC and Toyama-MICT) with clear image size and viewing distance values are used as testing beds in this paper. Experimental results and comparative studies on different combinations of IQA methods and image databases suggest the effectiveness and the robustness of the proposed approach.**

I. Introduction

Image quality assessment (IQA) is a significant research area in the image processing. Generally, image quality metric can be classified into two categories: subjective assessment and objective assessment. The subjective IQA method should be the ultimate quality gauge for images, but it is usually time-consuming, expensive and impractical for real-time image processing systems. For objective image quality metrics, the Mean-Squared Error (MSE) and its equivalent the Peak Signal-to-Noise Ratio (PSNR) are the most widely used approaches as the benchmark in practice, due to both the convenience and the definite physical meaning.

During the last decade, the research of image quality assessment has made great progress. On one hand, plenty of image databases have been built according to the instruction on subjective assessment in the ITU-R BT.500 [1], which is made by the International Telecommunication Union (ITU); on the other hand, a great many objective IQA algorithms have been developed, and achieved fairly well performances in terms of the correlation between the quality predictions and the subjective scores. Among them, the most well-known Structural SIMilarity (SSIM) index [2] is substantially focusing on structural information. Then, an increasing number of image quality metrics by using multiple scales have been further exploited, such as Multi-Scale SSIM (MS-SSIM) [3], Information Fidelity Criterion (IFC) [4], Visual Information Fidelity (VIF) [5], and Information content Weighting (IW) PSNR/SSIM [6]. Their better results suggest that the reliable scale chosen is probably a key factor for perceptual image quality assessment.

So far, there have been very few studies on the influence of scale transform on the prediction accuracy of image quality metrics [7]. A simple, empirical approach has been later proposed for SSIM, making it achieve more accurate prediction scores. Besides, note that the detailed demands on image size and viewing distance in the laboratory environment are not parts of the recommendations in ITU-R BT.500 [1]. As a result, existing image databases adopt different kinds of image size and viewing distance values during subjective experiments.

Inspired by this, it has been further testified in our research that the valid scale transform can be able to improve the performances of quality metrics. We believe it can be explained by the fact that human perception to details in images largely depends the resolution of Human Visual System (HVS). As the viewing distance becomes farther and farther, the resolution of human eyes reduces, and the distinguishment of the tiny artifacts in an image is gradually more difficult, as shown in Fig 1. So, this paper designs a valid scale transform function, mainly consisting of image size and viewing distance, to approximately simulate the real mechanism of HVS, and on this base, develops a Self-Adaptive Scale Transform (SAST) model based IQA method.

The remainder of this paper is organized as follows. Section II proposes the SAST model, taking into account the concept of human visual angle and angle of gaze. In Section III, experimental results are reported and analyzed, by using PSNR and SSIM as testing IQA algorithms, and the LIVE database [8], IVC database [9], and Toyama-MICT database [10] as testing beds, because of the clear viewing distance and image size values during their subjective experiments [7]. Finally, a conclusion and future work are given in Section IV.

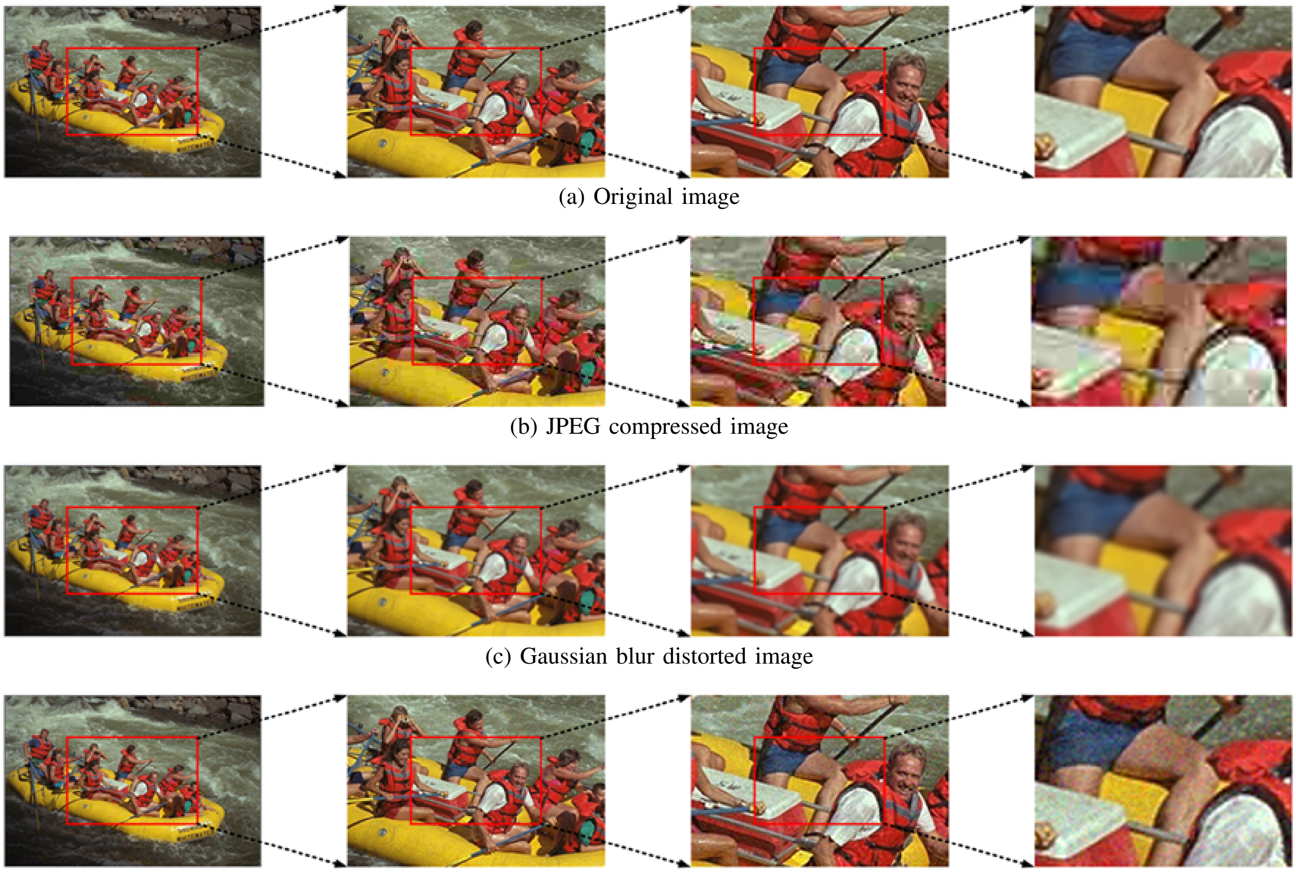
II. The SAST Model

It is demonstrated in [7] that the external factors, including image size and viewing distance, have considerable influences on the prediction accuracy of image quality metrics. A simple, empirical method has been exploited for SSIM to determine the downsampling scale Z for evaluating images viewed from a typical distance:

$$Z_\alpha = max(1, round(H_I/256)) \tag{1}$$

where H_I indicates the image height. For an image X, its version after scale transform X' can be computed by

978-1-4673-5762-3/13/$31.00 ©2013 IEEE



(a) Original image

(b) JPEG compressed image

(c) Gaussian blur distorted image

(d) White noise image

Fig. 1. Changes of resolution of human eyes with the closer viewing distance.

$$X' = R(L(X), Z) \tag{2}$$

where $R(\cdot)$ and $L(\cdot)$ indicate image resize and low-pass filtering function. Then, the improved PSNR and SSIM using downsampling scale Z_α are given by

$$\begin{aligned} PSNR_\alpha &= PSNR(X', Y') \\ &= PSNR(R(L(X), Z_\alpha), R(L(Y), Z_\alpha)) \end{aligned} \tag{3}$$

and

$$\begin{aligned} SSIM_\alpha &= SSIM(X', Y') \\ &= SSIM(R(L(X), Z_\alpha), R(L(Y), Z_\alpha)). \end{aligned} \tag{4}$$

Table II-III present that the above-mentioned methods have gained better results. Since there are not explicit parametric choices for viewing condition variations in the major existing IQA algorithms, it naturally suggests designing a valid self-adaptive scale transform model to help to promote the prediction performance of IQA approaches.

First of all, it is believed that the changes of resolution should be a gradual process, and the jumping scale transform of Z_α is quite unreasonable. For example, at the same certain viewing distance, $Z_\alpha = 3$ for an image with 1024×650 and $Z_\alpha = 2$ for an image with 1024×630 shows that the larger image looks smaller, and vice versa. Moreover, it is stated in [11] that the amount of detail resolvable by the human eyes is primarily limited by the density of the light-sensitive rods and cones on the eyes' retina. Therefore, for a viewing distance D, we define the visual scope S of human eyes as

$$S = H \cdot W \tag{5}$$

with H and W being the visual height and width. As illustrated in Fig. 2, they can be estimated by

$$H = 2\,tan(\frac{\theta_H}{2}) \cdot D \tag{6}$$

and

$$W = 2\,tan(\frac{\theta_W}{2}) \cdot D. \tag{7}$$

Generally, θ_H and θ_W are around 120^o and 150^o [12]. When a viewer pays attention to the details of an image so as to score it, the real view angle (i.e. angle of gaze) will become narrower, about one third of common angle of visual. So, we choose $\theta_H = 40^o$ and $\theta_W = 50^o$ in this paper. Just as shown in Fig. 1, it becomes increasingly difficult to distinguish the tiny distortions in an image, with the viewing

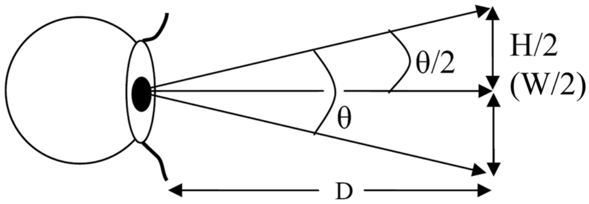


Fig. 2. The human visual angle.

TABLE I
DESCRIPTION OF LIVE, IVC AND TOYAMA-MICT DATABASES.

Database name	Image size ($W_I \times H_I$)	D / H_I	Image number
LIVE	768×512, 480×720, 640×512, 632×505, 634×505, 618×453, 610×488, 627×482, 634×438	3∼3.75	779
IVC	512×512	4	185
Toyama-MICT	768×512	6	168

distance being farther and farther. Accordingly, the reliable scale transform of an image should be carefully chosen before objective IQA methods to be used. To approximately stimulate the real mechanism of HVS, this paper define the scale Z of SAST model as the root square of the ratio of image size to visual scope as follows

$$\begin{aligned} Z_S &= \sqrt{\frac{H_I \cdot W_I}{H \cdot W}} \\ &= \sqrt{\frac{1}{4\,tan(\frac{\theta_H}{2}) \cdot tan(\frac{\theta_W}{2})} \cdot (\frac{H_I}{D})^2 \cdot \frac{W_I}{H_I}}, \end{aligned} \tag{8}$$

where W_I represents the image width. Consequently, the proposed SAST model based PSNR and SSIM can be evaluated by

$$\begin{aligned} PSNR_S &= PSNR(X', Y') \\ &= PSNR(R(L(X), Z_S), R(L(Y), Z_S)) \end{aligned} \tag{9}$$

and

$$\begin{aligned} SSIM_S &= SSIM(X', Y') \\ &= SSIM(R(L(X), Z_S), R(L(Y), Z_S)). \end{aligned} \tag{10}$$

III. Experimental Results

This paper adopts LIVE, IVC and Toyama-MICT databases as testing beds, for there are definite viewing distance and image size values during the subjective experiments, as tabulated in Table I.

Mappings of the scores of six metrics PSNR, PSNR_α, SSIM, SSIM_α, and our PSNR_S and SSIM_S to subjective Mean Opinion Scores (MOSs) or Differential MOSs (DMOSs) are achieved by applying nonlinear regression with a four-parameter logistic function as suggested by VQEG [13]:

$$q(x) = \frac{\beta_1 - \beta_2}{1 + exp(-(x - \beta_3)/\beta_4)} + \beta_2 \tag{11}$$

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

Three usually used performance metrics, Pearson Linear Correlation Coefficient (PLCC), Spearman Rank-order Correlation Coefficient (SRCC), and Root Mean-Squared Error (RMSE) as suggested by VQEG [13], are employed to further evaluate the proposed SAST model based PSNR/SSIM metric and the other four IQA methods on LIVE, IVC and Toyama-MICT databases. Their performance values and directly average results are illustrated in Table II-III, and all the corresponding scatter plots are displayed in Fig. 3. As we expected, by having image processed with a reliable scale transform, PSNR_S and SSIM_S have obtained inspiring results, especially the performance of PSNR_S, which is the best of all the PSNR type of methods.

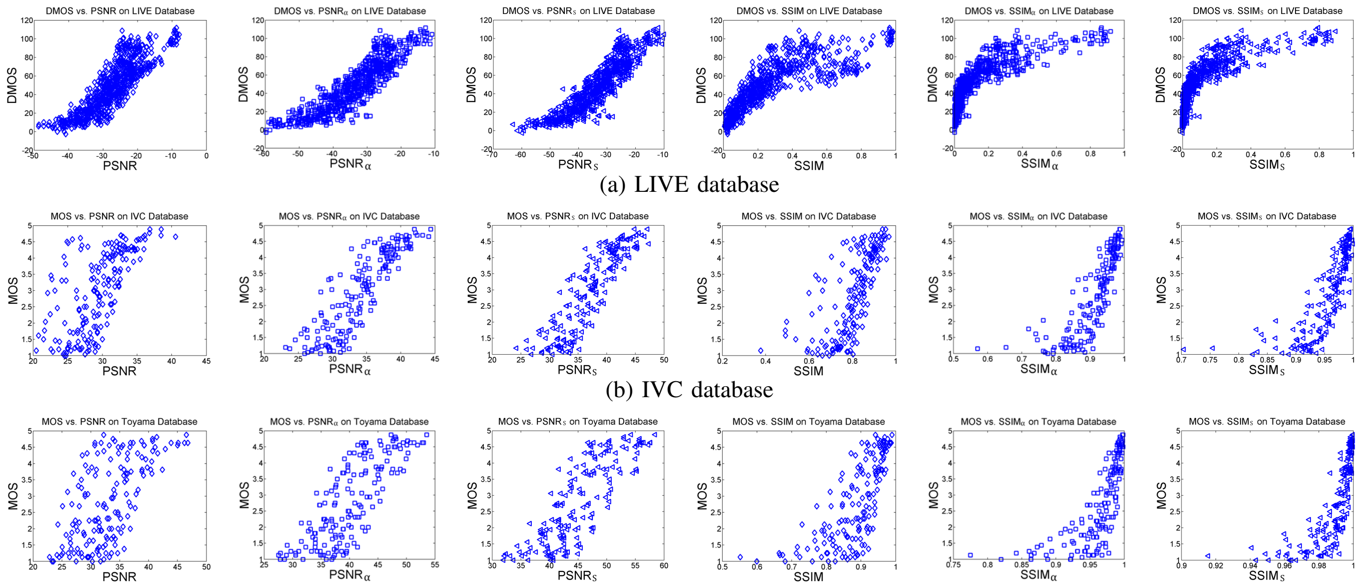
Besides, there are two points worth to be mentioned. First, the proposed SAST model is simple, effective, and more close to the real mechanism of HVS. Second, notice that our SSIM_S has worse accuracy than SSIM_α on LIVE and IVC databases. We believe it is probably because image content, including complexity and distortion types, has fairly important effect on performance of SSIM, which is just not considered by the SAST model yet. Third, we believed that it is probably more accurate to measure downsampling scale by the square root of the ratio of Gaussian/Gabor energies between image size and visual scope, which has been proved with quite good performance in the research of IQA [14]-[15].

TABLE II
PLCC, SRCC AND RMSE VALUES (AFTER NONLINEAR REGRESSION) OF PSNR, PSNR_α, PSNR_S, SSIM, SSIM_α AND SSIM_S ON LIVE (779 IMAGES), IVC (185 IMAGES), TOYAMA-MICT (168 IMAGES) DATABASES.

LIVE database [8]			
	PLCC	SRCC	RMSE
PSNR	0.8701	0.8756	13.469
PSNR_α	0.9031	0.9056	11.735
PSNR_S	**0.9137**	**0.9164**	**11.104**
SSIM	0.9014	0.9104	11.832
SSIM_α	0.9251	0.9355	10.376
SSIM_S	**0.9306**	**0.9446**	**10.002**

IVC database [9]			
	PLCC	SRCC	RMSE
PSNR	0.7195	0.6887	0.8462
PSNR_α	0.8912	0.8828	0.5527
PSNR_S	**0.8956**	**0.8893**	**0.5419**
SSIM	0.7923	0.7785	0.7433
SSIM_α	**0.9122**	**0.9030**	**0.4993**
SSIM_S	0.9046	0.8912	0.5195

Toyama-MICT database [10]			
	PLCC	SRCC	RMSE
PSNR	0.6352	0.6130	0.9665
PSNR_α	0.8003	0.7942	0.7504
PSNR_S	**0.8355**	**0.8276**	**0.6876**
SSIM	0.7962	0.7865	0.7571
SSIM_α	0.8917	0.8844	0.5664
SSIM_S	**0.9079**	**0.9042**	**0.5247**



(a) LIVE database

(b) IVC database

(c) Toyama-MICT database

Fig. 3. Scatter plots of DMOS/MOS vs. PSNR, PSNR_α, PSNR_S, SSIM, SSIM_α and SSIM_S on LIVE, IVC and Toyama-MICT databases.

IV. Conclusion

In this paper, we propose a self-adaptive scale transform model based IQA paradigm, mainly relying on the concept of human visual angle and angle of gaze to simulate the real mechanism of HVS. Experimental results on LIVE, IVC and Toyama-MICT databases are provided to confirm that PSNR_S and SSIM_S have obtained inspiring results. In addition, our SAST model is quite simple, effective, and has achieved the most superior performance of the three of PSNR type of methods. Finally, it is worth emphasizing that this research suggests a new direction of IQA approach, by taking the external factors, including image size and viewing distance, into consideration.

However, it has been observed that the prediction accuracy of SSIM_S is not the best of all the SSIM type of methods, which is possibly due to the fact that SSIM is a image content based IQA approach. So, future work will be devoted to improve the SAST model, by adding more image content features, such as complexity and distortion category.

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 III
Directly average results of performance values in Table II.

	PLCC	SRCC	RMSE
PSNR	0.7416	0.7257	5.0937
PSNR_α	0.8648	0.8609	4.3460
PSNR_S	**0.8816**	**0.8778**	**4.1111**
SSIM	0.8300	0.8251	4.4443
SSIM_α	0.9097	0.9076	3.8139
SSIM_S	**0.9143**	**0.9133**	**3.6820**

References

[1] ITU, "Methodology for the subjective assessment of the quality of television pictures," *Recommendation, International Telecommunication Union/ITU Radiocommunication Sector*, 2009.

[2] 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, Apr. 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*, Nov. 2003.

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

[5] 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.

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

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

[8] 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/

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

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

[11] H. Knoche, and M. Sasse, "The sweet spot: how people trade on size and definition on mobile devices," *In Proc. of the 16th ACM Int. Conf. on Multimedia*, 2008.

[12] E. Bruce Goldstein, Sensation and Perception. 1979.

[13] 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/.

[14] G. Zhai, W. Zhang, X. Yang, S. Yao, and Y. Xu, "GES: a new image quality assessment metric based on energy features in Gabor transform domain," *Proc. IEEE Int. Symposium on Circuits and Systems*, pp. 1715-1718, 2006.

[15] G. Zhai, W. Zhang, Y. Xu, and W. Lin, "LGPS: Phase based image quality assessment metric," *IEEE Workshop on Signal Processing Systems*, pp. 605-609, 2007.