Full-Reference Image Quality Assessment via Region-Based Analysis

Ke Gu, Wenjun Zhang, Ci Wang and Guangtao Zhai
Institute of Image Communication and Information processing
Shanghai Jiaotong University
Shanghai, China

Abstract **— Image quality assessment (IQA) plays an important role in many image processing systems. Based on the fact that different image components have different visual impact, we propose a new region-based method using a three-region image model and Back Propagation (BP) neural network. Experimental results show that our algorithm is obviously better than the peer image quality assessment method, using the Laboratory for Image and Video Engineering (LIVE) database as a test-bed.**

Keywords**-Image quality assessment (IQA); back propagation (BP) neural network; region-based analysis.**

I. INTRODUCTION

Image quality assessment (IQA) is important for many applications. Existing approaches fall into two kinds: subjective assessment and objective assessment. While the subjective assessment is the terminal gauge of image quality (IQ), it is time-consuming, expensive and cannot be used in real-time image processing system. The mean-squared error (MSE) and its derived peak signal-to-noise ratio (PSNR) are still the most widely used objective metrics, due to their convenience and clear physical meaning. However, it has been proved that they are not well fitted with human judgment of quality [1]. Subsequently, Wang et al [2] presented that the human visual system (HVS) is highly adapted to extract structural information from visual scenes. Thus, the Structural Similarity (SSIM) [3] Index is proposed, and experimental results demonstrate that it is more consistent with HVS than MSE/PSNR.

During the last three decades, many studies have investigated the utility of HVS for IQA, and a great deal of effort has been made to develop new IQA methods [4]-[6]. Hantao Liu et al maintained that a human VA saliency map is generally originated from the spatial pattern of fixations in eye-tracking experimental data [7], and put human beings' saliency map into PSNR and SSIM metrics by locally weighting the corresponding distorting map, like the combination strategy in [8]. Thus, the method combining PNSR/SSIM and human beings' saliency map, which is called WPSNR/WSSIM [9], is proposed.

However, there is one fatal shortage of VA model, based on eye-tracking experiments with unimpaired images under natural viewing conditions, that the data of eye-tracking experiment are just for those source images of the LIVE database [10], not for all kinds of natural images. So this IQA method is difficult to be used in image processing system, which includes other kinds of images.

In our study, it was found that edge, smooth and texture regions have different visual impact. For example, edge component has stronger impact than other two parts. According to this, we use the three-region image model to classify image local regions by their image gradient properties, and then apply BP network to retrieve their subjective scores from their SSIM and PSNR scores. Thus, an improved IQA method called Region-and-Bp-based SSIM/PSNR (RBSSIM/RBPSNR) is proposed.

The work is organized as follows. Section Ⅱ introduces two basic methods and our proposed model. In section Ⅲ, results of experiments conducted on LIVE database are reported. Finally, conclusion is drawn and future work is discussed in section Ⅳ.

II. REGION-BASED METHOD

This section introduces three parts, three-region image model, BP neural network and our proposed region-based measure.

A. Three-Region Image Model

We partition the distorted image into edge, smooth and texture regions using the computed gradient magnitude [11].

$$(E, S, T) = F_R(Ori(x, y), Dis(x, y)) \quad (1)$$

where E, S and T are edge, smooth and texture regions respectively, Ori and Dis indicate the original image and distorted image and F_R denotes dividing Ori and Dis into E, S and T.

The following steps explain the whole process. First, compute the gradient magnitudes using a Sobel operator on the original and the distorted images. Second, determine thresholds $TH_1 = 0.12\ g_{max}$ and $TH_2 = 0.06\ g_{max}$ where g_{max} is the maximum gradient magnitude value computed over the original image. Third, assign pixels as belonging to edge, smooth and texture regions as follows:

R_1: if $p_o(x,y) > TH_1$ or $p_d(x,y) > TH_1$, then the pixel is considered to be an edge pixel.

978-1-4244-9306-7/11/$26.00 ©2011 IEEE

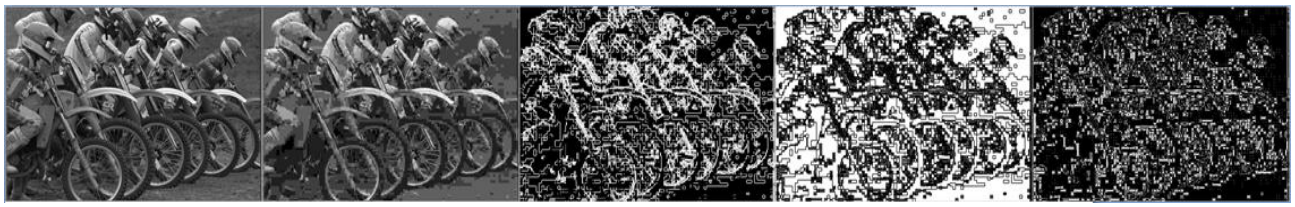

Figure 1. Illustration of three-region image model: original image, distorted image, edge_map image, smooth_map image and texture_map image (from left to right)

R_2: if $p_o(x,y) < TH_2$ and $p_d(x,y) \leqslant TH_1$, then the pixel is regarded as part of a smooth region.

R_3: otherwise, the pixel is regarded as part of a textured region.

$$(x,y) \in \begin{cases} E & p_o(x,y) > TH_1 \text{ or } p_d(x,y) > TH_1 \\ S & p_o(x,y) < TH_2 \text{ and } p_d(x,y) \leq TH_1 \\ T & \text{otherwise} \end{cases} \quad (2)$$

where $p_o(x,y)$ and $p_d(x,y)$ indicate the gradient values at coordinate (i,j) on the original image and distorted image. One example of this model is given in Fig.1.

B. BP Neural Network

The BP network is the most generally used neural network because of its wide application and efficiency. It consists of an input layer, several hidden layers, and an output layer [12]. BP network can approximate any continuous non-linear function with arbitrary accuracy provided that there are enough hidden neurons. Their corresponding role function, in modern research, regularly selects the S function [13], and its expression is:

$$f(x) = \frac{1}{1+\exp(-ax)}. \quad (3)$$

Also there are some other functions, such as:

$$f(x) = \tan\left(\frac{x}{2}\right) = \frac{1-\exp(-ax)}{1+\exp(-ax)}. \quad (4)$$

BP network structure is provided in Fig.2 and its corresponding function is as follows:

$$(Y_1, \ldots, Y_j, \ldots Y_m) \approx F_B(X_1, \ldots, X_i, \ldots X_n) \quad (5)$$

where F_B denotes the relationship between input vector (X_1, … ,X_n) and output vector (Y_1, … ,Y_m).

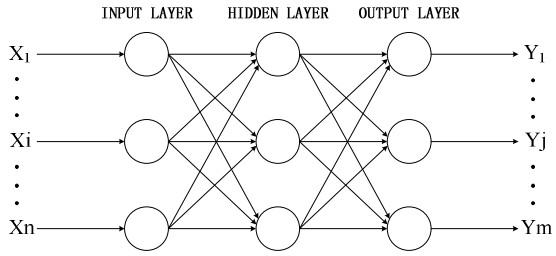

Figure 2. Model structure of artificial neural network

In this paper, we use BP neural network to construct the relationship between three groups of input (edge, smooth and texture) and DMOS output. Replace X_1, X_2, X_3 and Y_1 in Eq. (5) with edge_ssim, smooth_map, texture_map and DMOS, then

$$DMOS \approx F_B(\text{edge\_map}, \text{smooth\_map}, \text{texture\_map}) \quad (6)$$

where DMOS indicates difference mean opinion score. The setting of our BP network is as follows:

(1) Determining unit number of input layer: This article will take 3 vectors (edge, smooth, texture) as 3 nodes of the network input layer.

(2) Determining the units number of hidden layer: The three-layer network can implement the nonlinear mapping from input vector $X = (X_1,\ldots,X_n)$ to output vector $Y = (Y_1,\ldots,Y_m)$. The unit number, denoting q, of hidden layer has yet no sophisticated methods and, in actual operation, the formula $q = \log_2 n$ can be determined. So, nodes number in hidden layer of this model is selected as 2.

(3) Determining unit number of output layer: In our model, the result is DMOS, which can be as outputs of the corresponding network, and apparently m equals to 1.

C. Region-Based Measure

Image can be divided into three regions (edge, smooth and texture) and these three regions are perceived differently by human beings' eyes. Fig.3 presents that smooth distorted image looks worse than texture distorted image, and edge distorted image is the worst of all, when their PSNRs are almost the same. The PSNR values and perception results of these four images in Fig.3 are shown in Table Ⅰ.

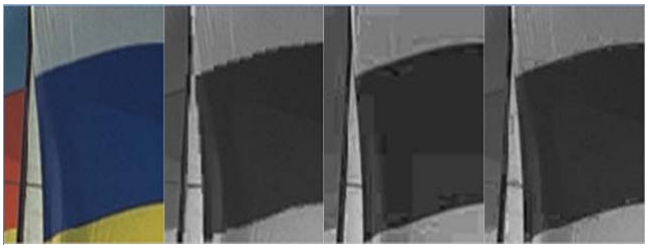

Figure 3. Illustration of different region impact: no distorted image, edge distorted image, smooth distorted image, texture distorted image (from left to right)

TABLE I. THE PSNRS AND PERCEPTION OF FOUR IMAGES IN FIGURE 3

DISTORTION SITUATION	PSNR	PERCEPTION
NO DISTORTED IMAGE	+ ∞ dB	REFERENCE
EDGE DISTORTED IMAGE	33.12 dB	BAD
SMOOTH DISTORTED IMAGE	33.25 dB	POOR
TEXTURE DISTORTED IMAGE	33.04 dB	FAIR

Fig.3 and Table Ⅰ make the qualitative analysis of these three kinds of regions. It is in search of the relationship between DMOS and SSIM/PSNR values on image edge, smooth and texture that further experiments are carried out.

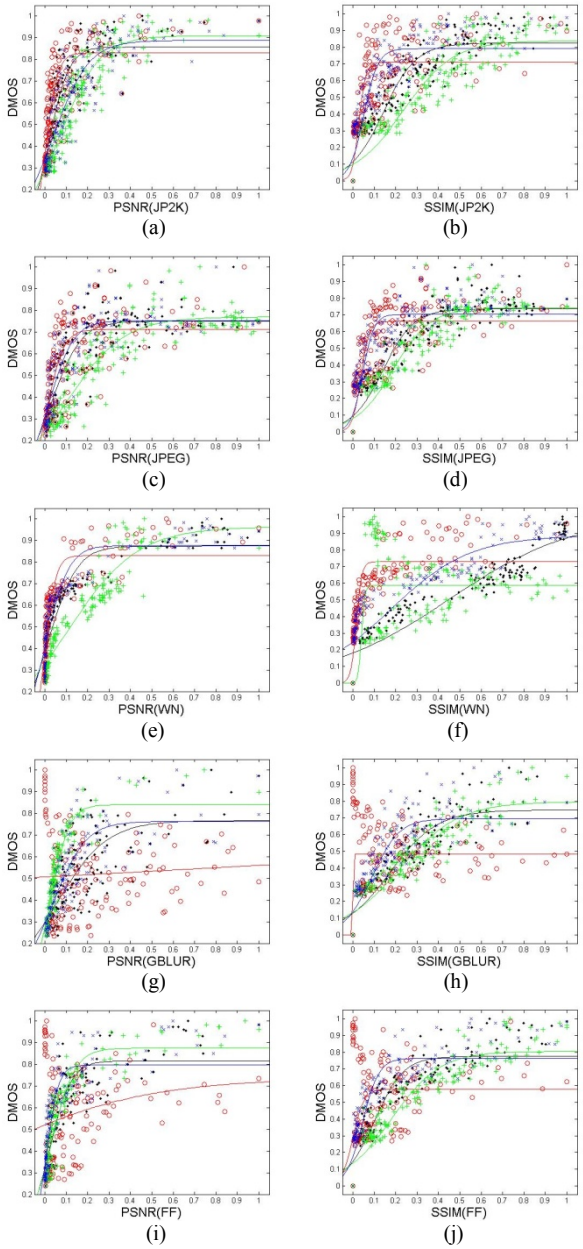

Figure 4. Illustration of scatter plots and their corresponding regression lines of PSNR(black dot), EDGE_PSNR(red circle), SMOOTH_PSNR (green "x"), TEXTURE_PSNR(blue "+"), SSIM(black dot), EDGE_SSIM(red circle), SMOOTH_SSIM(green "x") and TEXTURE_SSIM(blue "+") for JP2K, JPEG, WN, GBLUR and FF

Fig.4 illustrates that the scatter plots between the DMOS and the predictions of eight objective metrics (PSNR, EDGE_PSNR, SMOOTH_PSNR, TEXTURE_PSNR, SSIM, EDGE_SSIM, SMOOTH_SSIM and TEXTURE_SSIM) for JPEG2000 compression (JP2K), JPEG compression (JPEG), white noise (WN), Gaussian blur (GBLUR) and simulated fast fading Rayleigh (wireless) channel (FF). Then, regress analysis is used to each objective metric, shown in Fig.4, by the psychometric function recommended by the Video Quality Expert Group (VQEG) [14] as follows:

$$\mathrm{MOS_p} = \frac{\mathrm{b1}}{1+\exp\left(-\mathrm{b2}*(\mathrm{Q}-\mathrm{b3})\right)} \tag{7}$$

where b1, b2 and b3 are the three parameters of the psychometric function.

TABLE II. THE RMSE VALUES OF PSNR/SSIM BETWEEN THE SCATTER PLOTS AND THEIR CORRESPONDING REGRESSION LINES FOR JP2K, JPEG, WN, GBLUR AND FF

	REGION	JP2K	JPEG	WN	GBLUR	FF
PSNR	EDGE	0.1165	0.1283	0.0947	0.1881	0.2107
	SMOOTH	0.0781	0.0797	0.0750	0.0705	0.0850
	TEXTURE	0.1032	0.0992	0.0845	0.1033	0.1015
	ALL	0.1022	0.1029	0.0752	0.1212	0.1061
SSIM	EDGE	0.1582	0.1387	0.1356	0.2362	0.2613
	SMOOTH	0.1084	0.0913	0.1727	0.1010	0.0987
	TEXTURE	0.1336	0.1186	0.1389	0.1232	0.1257
	ALL	0.1104	0.0955	0.1074	0.1098	0.1090

According to these findings, we propose a new image region model, based on three-region image model and BP neural network. By inserting this model into SSIM, RBSSIM method is developed:

$$\mathrm{RBSSIM} = \mathrm{F_{RSB}}(\mathrm{Ori},\mathrm{Dis}) = \mathrm{F_B}(\mathrm{F_S}(\mathrm{F_R}(\mathrm{Ori},\mathrm{Dis}))) \tag{8}$$

where Fs indicates SSIM[3] method. Combine the Eq. (1) and Fs into Eq. (8):

$$\mathrm{RBSSIM} = \mathrm{F_B}(\mathrm{edge\_map},\mathrm{smooth\_map},\mathrm{texture\_map}) \tag{9}$$

where

$$\mathrm{region\_map} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\{\,\mathrm{region\_ssim\_map(x,y)}\,\}}{\sum_{x=1}^{M}\sum_{y=1}^{N}} \tag{10}$$

and "region" indicates "edge", "smooth" or "texture", and region_ssim_map is called between the corresponding region of the distorted image and original image using SSIM metric. Then, combine the Eq. (6) into Eq. (9):

$$\mathrm{DMOS} \approx \mathrm{RBSSIM}. \tag{11}$$

Similarly:

$$\mathrm{RBPSNR} = \mathrm{F_B}(\mathrm{edge\_map},\mathrm{smooth\_map},\mathrm{texture\_map}) \tag{12}$$

where

$$\mathrm{region\_map} = \frac{\sum_{x=1}^{M}\sum_{y=1}^{N}\{\,\mathrm{region\_psnr\_map(x,y)}\,\}}{\sum_{x=1}^{M}\sum_{y=1}^{N}} \tag{13}$$

and region_psnr_map is called between the corresponding region of the distorted image and original image using PSNR metric, and

$$\mathrm{DMOS} \approx \mathrm{RBPSNR}. \tag{14}$$

The concrete steps of RBPSNR and RBSSIM are as follows: 1. Calculate PSNR/SSIM map. 2. Segment the

original and distorted images into three categories of regions (edge, smooth and texture) as described in Section 2.1 and get six groups of values, PSNR_EDGE, PSNR_SMOOTH, PSNR_TEXTURE, SSIM_EDGE, SSIM_SMOOTH and SSIM_TEXTURE. 3. Train the DMOS values and these three groups of values by BP network and get their weights. 4. Pool the weighted PSNR/SSIM values and define a single quality index (RBPSNR/RBSSIM) for the image. The corresponding strategy is depicted in Fig.5.

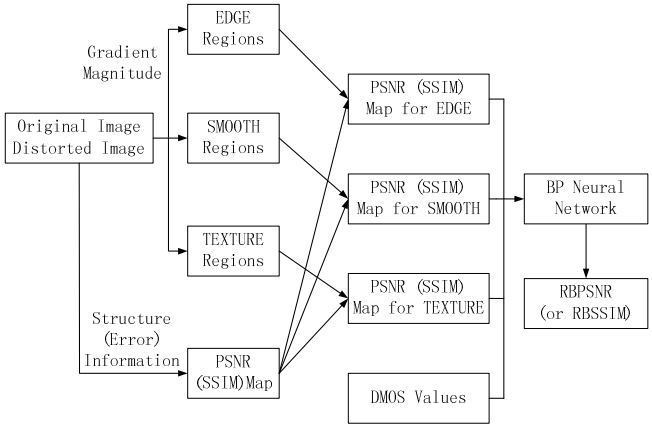

Figure 5. Diagram for RBPSNR/RBSSIM's strategy

III. Experimental Results

The experiments were carried out for the whole LIVE database, including 29 original images and 779 images distorted with JP2K, JPEG, WN, GBLUR and FF. The original images are divided into two groups, as shown in Fig.6. Group 1 is the train group, which is used to train the BP network, and Group 2 is the test group for testifying. The whole DMOS values were derived for each of 779 distorted images by an extensive subjective quality assessment study [10].

(a) (b)

Figure 6. Illustration of the train group and test group: (a) Group 1; (b) Group 2.

These ten metrics, PSNR, WPSNR, RBPSNR-1, RBPSNR-2, RBPSNR-3, SSIM, WSSIM, RBSSIM-1, RBSSIM-2 and RBSSIM-3 are applied to the 779 distorted images in LIVE database (RBPSNR-1/RBSSIM-1 and RBPSNR-2/RBSSIM-2 are calculated from Group 1 and Group 2 respectively and RBPSNR-3/RBSSIM-3 is derived from all the 779 distorted images). Fig.7 presents that the scatter plots between the DMOS and the predictions of six objective metrics (PSNR, WPSNR, RBPSNR-3, SSIM, WSSIM, and RBSSIM-3).

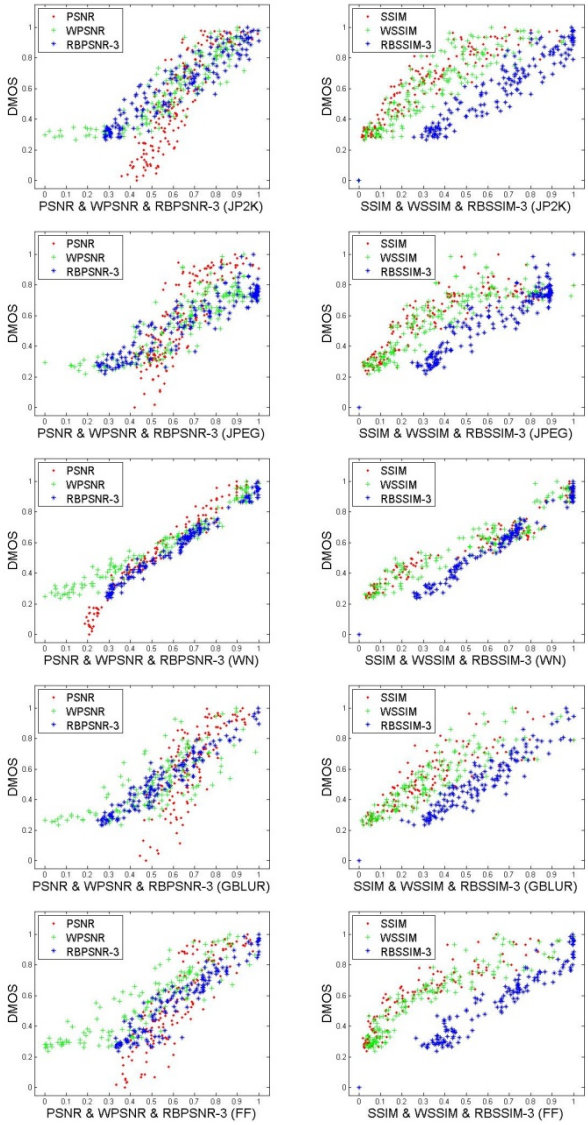

Figure 7. Scatter plots of DMOS vs. six metrics PSNR, WPNSR, RBPSNR-3, SSIM, WSSIM, RBSSIM-3 for JP2K, JPEG, WN, GBLUR and FF, respectively.

Fig.8 shows the corresponding correlation coefficients of PSNR, WPSNR, RBPSNR-1, RBPSNR-2, RBPSNR-3, SSIM, WSSIM, RBSSIM-1, RBSSIM-2 and RBSSIM-3. It illustrates there is really a gain in performance with WPSNR method. Furthermore, there is still an obvious improvement from WPSNR/WSSIM to RBPSNR/RBSSIM method. But the improving performance is dependent on the metrics. For example, the gain of RBPSNR over PSNR/WPSNR for JPEG is greatly different from that for WN, and the situation of RBSSIM over SSIM/WSSIM for JPEG and WN has a similar result. This phenomenon is mainly due to different importance of these three regions in JP2K, JPEG, WN, GBLUR and FF images.

In addition, RBPSNR-3/RBSSIM-3 is higher than RBPSNR-2/PBSSIM-2, and RBPSNR-1/PBSSIM-1 has the highest value, because RBPSNR-1/PBSSIM-1, RBPSNR-2/PBSSIM-2 and RBPSNR-3/PBSSIM-3 are derived from train group, test group and the whole images, respectively.

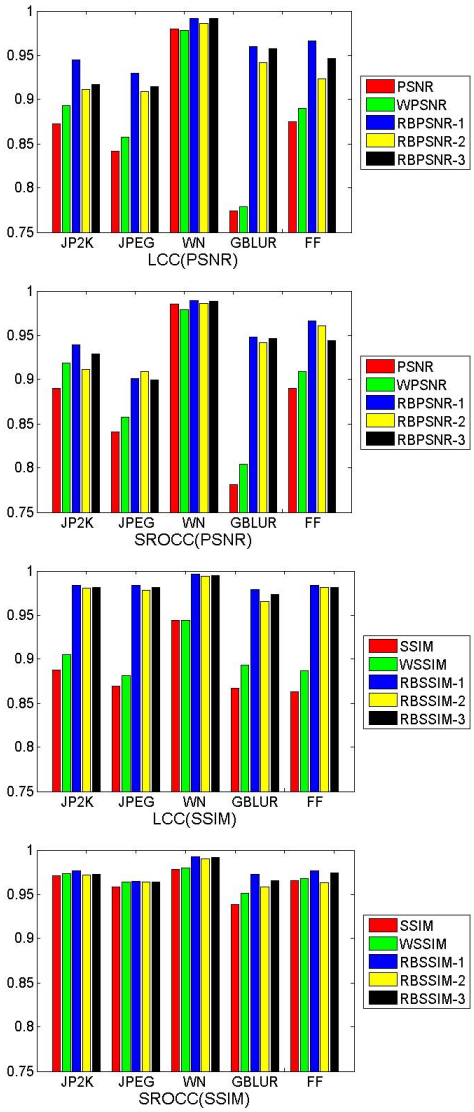

Figure 8. Diagram for Pearson and Spearman rank order correlation coefficients of ten metrics PSNR, WPNSR, RBPSNR-1, RBPSNR-2, RBPSNR-3, SSIM, WSSIM, RBSSIM-1, RBSSIM-2 and RBSSIM-3 for JP2K, JPEG, WN, GBLUR and FF, respectively.

IV. Conclusion and Future Work

In this paper, we propose a new IQA metric based on image region model. This model includes three-region image model and BP neural network. The results of experiments prove that the performance of RBPSNR/RBSSIM is really better than WPSNR/WSSIM and, moreover, further better than PSNR/SSIM.

Future work will involve three respects: 1. Change the constant threshold values 0.12 and 0.06 in equations (TH_1 = 0.12 g_{max} and TH_2 = 0.06 g_{max}) with self-adaptive threshold adjustment technology. 2. Develop the strategy of combining VA model and three-region image model. 3. Extend this approach to Video Quality Assessment considering other features such as movement.

References

[1] A. M. Eskicioglu and P. S. Fisher, “Image quality measures and their performance,” IEEE Trans. Commun., vol. 43, pp. 2959–2965, Dec. 1995.

[2] Z. Wang and A. C. Bovik, Modern Image Quality Assessment. Morgan & Claypool Publishers, Mar. 2006.

[3] 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.

[4] M.Miyahara, K.Kotani and V.Algazi, "Objective Picture Quality Scale (PQS) for Image Coding", Tech.Rep. CIPIC,University of California, Davis, 1996.

[5] Claudio M privitera and Lawrence W.stark, "Algorithms for Defining Visual Regions of Interest Comparison with Eye Fixation," IEEE trans on PAMI, Vol.22, No.9, pp.970-980, 2000.

[6] N. G. Sadaka, L. J. Karam, R. Ferzli, and G. P. Abousleman, "A no-reference perceptual image sharpness metric based on saliency-weighted foveal pooling," in Proc. IEEE Int. Conf. ICIP, pp. 369-372, Oct. 2008.

[7] N. Ouerhani, R. V. Wartburg, H. Hugli, and R. Muri, “Empirical Validation of the Saliency-based Model of Visual Attention,” Electronic Letters on Computer Vision and Image Analysis, 3(1): 13-24, 2004.

[8] A. Ninassi, O. L. Meur, P. L. Callet, and D. Barba, “Does where you Gaze on an Image Affect your Perception of Quality? Applying Visual Attention to Image Quality Metric,” in Proc. IEEE Int. Conf. ICIP, pp. 169-172, Oct. 2007.

[9] Hantao Liu, Ingrid Heynderickx, “Studying the added value of visual attention in objective image quality metrics based on eye movement data,” in Proc. IEEE Int. Conf. ICIP, pp. 3097-3100, Oct. 2009.

[10] H.R. Sheikh, Z.Wang, L. Cormack, and A.C. Bovik, “LIVE Image Quality Assessment Database Release 2,” http://live.ece.utexas.edu/research/quality.

[11] J. L. Li, G. Chen, and Z. R. Chi, “Image coding quality assessment using fuzzy integrals with a three-component image model,” IEEE Trans. Fuzzy Syst. 12(1), 99-106 (2004).

[12] P. S. Neelakanta, 1994. Neural network modelling: statistical mechanics and cybernetic perspectives, Boca Raton, Florida; London: CRC Press.

[13] Guo Yajun, Zhang Shichang, “Mine safety evaluation model based on Wavelet Neural Network”. Northeastern University (Natural Science Edition), Vol.27, No.6, pp. 702-705, 2006.

[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/.