SPATIAL FEATURES BASED NO REFERENCE IMAGE QUALITY ASSESSMENT FOR JPEG2000

Z. M. Parvez Sazzad, Y. Kawayoke, and Y. Horita

Graduate School of Engineering, University of Toyama,
3190 Gofuku, Toyama, 930-8555 Japan
Phone / Fax: +81-76-445-6758
E-mail: d0570004@ems.u-toyama.ac.jp; horita@eng.u-toyama.ac.jp

ABSTRACT

Perceptual image quality evaluation has become an important issue due to increasing transmission of multimedia contents over the Internet and 3G mobile networks. Most of the no reference perceptual image quality evaluations traditionally attempted to quantify the predefined artifacts of the coded images. Under the assumption that human visual perception is very sensitive to edge information of an image and any kinds of artifacts create pixel distortion, we propose a new philosophy for designing a no reference image quality evaluation model for JPEG2000 images, which uses pixel distortions and edge information. Subjective experiment results on the images are used to train and test the model, which achieve good quality prediction performance.

Index Terms— No-reference (NR), JPEG2000, Mean opinion score prediction (MOSp), Zero-crossing (ZC).

1. INTRODUCTION

There is a big concern about the levels of image quality both for providers and users in many image processing applications from compression to printing, due to the advanced development of different image compression techniques and processing systems. There is no doubt that the subjective test is the most accurate method for quality evaluation because it reflects the true human perception. However, it is time consuming and expensive. There are three types of methods that are used for objective image quality evaluation, *full-reference (FR)*, *reduced-reference (RR)* and *no-reference (NR)*. Of them, NR method has recently received a great attention because the reference signal is not available in many practical applications or may be too expensive to provide. All of the published NR perceptual image quality assessment algorithms are implemented according to the prior knowledge of the types of image artifacts and few metrics are specialized in the quality assessment of JPEG2000 coded images [1]-[5]. Whereas nowadays, the JPEG2000 coder is getting more attention compared to the JPEG coder, due to its high coding performance, even though previously JPEG was the standard coder for still image. Therefore, in this research, we propose an NR quality evaluation for JPEG2000 coded images, irrespective of any predefined artifacts, based on pixel distortion and edge information.

Thanks to Japan Government (MEXT) for funding.

Table 1. Subjective test conditions and parameters

Method	SS (Adjectival categorical judgement)
Evaluation scales	5 Grades (Adjective scales)
Images	24-bits/pixel RGB color (768× 512)
Number of reference images	14
Coder	JPEG2000
Coding parameters [8]	6 (CR: 12, 24, 32, 48, 72 and 96)
Viewers	16 (Non expert, college students)
Display	CRT 17-inch
Viewing distance	4H (H: Picture hight)
Room illumination	Low

*Single Stimulus (SS), Compression Ratio (CR)

2. SUBJECTIVE EXPERIMENTS

Each observer was shown the images randomly and asked to provide his/her perception of quality on a discrete quality score that was divided into five and marked with the numerical value of the adjectives ("Bad =1", "Poor=2", "Fair=3", "Good=4", and "Excellent=5") under the test conditions of ITU-R Rec. 500-11. All subjects were screened prior to participating in the session for normal visual acuity with or without glasses, normal color vision and familiarity with the language. The subjective test conditions and parameters are shown in Table 1. The sixteen scores of each image were averaged to get a final Mean Opinion Score (MOS) of the image with subject reliability of 95% confidence interval.

3. PROPOSED MODEL

Many researches have already established that the main function of the human visual system is to extract structural or edge information from the viewing field, and the human visual system is highly adapted for this purpose [6]. Under the assumption that human visual perception is very sensitive to edge information and natural image signals are highly structured

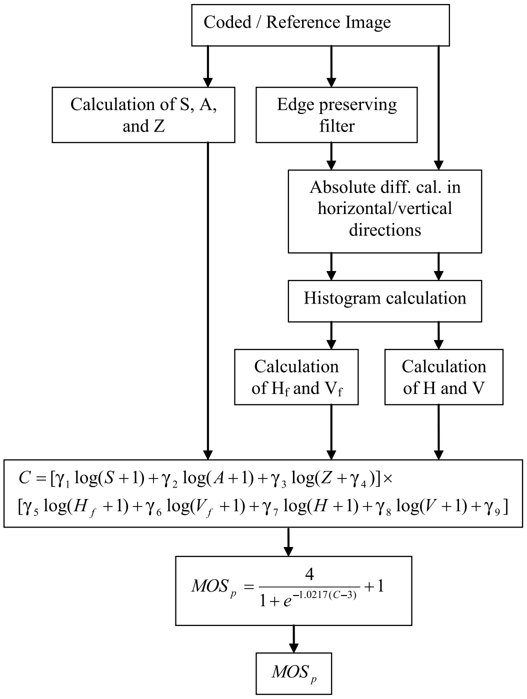
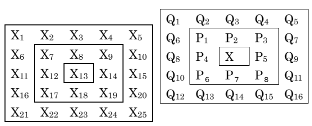


Fig. 1. Proposed model



(a) (b)

Fig. 2. 5×5 pixel block (a) Central pixel, X_{13} of the block (b) Central pixel, X and it's 1st and 2nd closest neighborhood.

specifically the samples of the signals have strong dependencies between each other, especially when they are close in space. Therefore, any kind of artifacts create pixel distortions from neighborhood pixels. As a result, in this research, we want to develop a new no reference image quality assessment model based on edge information and pixel distortions. The model is proposed mainly for JPEG2000 coded images and the features are calculated in a spatial domain. We attempt to design a memory efficient and computationally inexpensive feature extraction method. The block diagram of the model is shown in Figure 1.

3.1. Pixel distortions measure

In all calculations, we consider only the luminance part of the image. Pixel distortions are estimated using two features. First, standard deviation of a central pixel is estimated within 5×5 neighborhood pixels which is applied for all available pixels in the image. And then average the standard deviation values within 5×5 partially overlapping block. Let X_{13} is the central pixel of the 5×5 block that is shown in Fig.2(a), and also let $\overline{X}$, S_{std}, and $\overline{S_{std}}$ be the mean of pixels within the block, the standard deviation of X_{13} pixel in the block, and the average standard deviation within 5×5 partially overlapping block. The statistical features can be estimated as follows:

$$\overline{X} = \frac{1}{L}\sum_{i=1}^{L} X_i \tag{1}$$

where L = 25 is the total number of pixels in the block.

$$S_{std} = \sqrt{\frac{1}{L-1}\sum_{i=1}^{L}(\overline{X} - X_i)^2} \tag{2}$$

$$\overline{S_{std}} = \frac{1}{L}\sum_{i=1}^{L} S_{std_i} \tag{3}$$

If the total of row and column of the image are respectively M and N (M<N), then size of all available central pixels (within 5×5 neighborhood pixels) in the image is $M' \times N'$, where M' = M-4 and N' = N-4. And also if the $M' \times N'$ image is divided into 5×5 partially overlapping blocks (only corner pixels are overlapped) and each block is identified by a number, i. Finally, the 5×5 partially overlapping standard deviation feature (S) of the image is estimated by the following equation:

$$S = \frac{1}{\frac{M'-3}{4} \times \frac{N'-3}{4}} \sum_{i=1}^{\frac{M'-3}{4} \times \frac{N'-3}{4}} \overline{S_{std_i}} \tag{4}$$

The second pixel distortion measure is the absolute difference measure of a central pixel from the 2nd closest neighborhood pixels which is applied for all available pixels in the image. Let X be the central pixel and $Q_1, Q_2, ..., Q_{16}$ are the 2nd closest neighborhood pixels of the 5×5 block which is shown in Fig.2(b), and also let A_d be the absolute difference of a central pixel from the 2nd closest neighborhood pixels, and it is calculated by the following equation:

$$A_d = \frac{1}{16}\sum_{i=1}^{16} |X - Q_i| \tag{5}$$

Similarly, A_d values are calculated for all available pixels and averaged within 5×5 partially overlapping block by the following equation:

$$\overline{A_d} = \frac{1}{L}\sum_{i=1}^{L} A_{d_i} \tag{6}$$

where L = 25 is the total number of pixels in the block. Finally, the 5×5 partially overlapping average absolute difference feature (A) of the image is estimated by the following equation:

$$A = \frac{1}{\frac{M'-3}{4} \times \frac{N'-3}{4}} \sum_{i=1}^{\frac{M'-3}{4} \times \frac{N'-3}{4}} \overline{A_{d_i}} \tag{7}$$

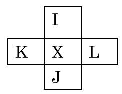


Fig. 3. Masking block of the edge preserving filter

3.2. Edge information measure

Edge information is estimated using two features. First, zero-crossing (ZC) rate is estimated both in horizontal and vertical direction of the image. We denote the test image signal as x(m,n) for m ∈ [1, M] and n ∈ [1, N] and calculate a differencing signal along each horizontal line:

$$d_h(m,n) = x(m,n+1) - x(m,n), \tag{8}$$

$$n \in [1, \text{N-1}] \quad \text{and} \quad m \in [1, \text{M}]$$

For horizontal ZC :

$$d_{h-sign}(m,n) = \begin{cases} 1 & \text{if } d_h(m,n) > 0 \\ -1 & \text{if } d_h(m,n) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$\begin{aligned} & d_{h-mul}(m,n) \\ & = \ d_{h-sign}(m,n) \times d_{h-sign}(m,n+1) \end{aligned} \tag{10}$$

We define for n ∈ [1, N-2]:

$$z_h(m,n) = \begin{cases} 1 & \text{if } d_{h-mul}(m,n) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

Then the horizontal zero-crossing rate, Z_{bh}, is calculated within 5×5 partially overlapping block and then the average horizontal zero-crossing rate (Z_h) is estimated of the image as follows:

$$Z_{bh} = \frac{1}{25} \sum_{i=1}^{5} \sum_{j=1}^{5} z_h(i,j) \tag{12}$$

$$Z_h = \frac{1}{\frac{M-3}{4} \times \frac{N-5}{4}} \sum_{i=1}^{\frac{M-3}{4} \times \frac{N-5}{4}} Z_{bh_i} \tag{13}$$

Similarly, we can calculate the average vertical zero-crossing rate (Z_v) of the image and finally the overall feature of zero-crossing rate is given by:

$$Z = \frac{Z_h + Z_v}{2} \tag{14}$$

The second edge information measure is the histogram measure with and without edge preserving filter. The edge preserving filtering algorithm is calculated by the following equations [7]:

$$H_d = K - 2X + L \ \ and \ \ V_d = I - 2X + J \tag{15}$$

$$\begin{aligned} if(H_d < V_d), \quad & X = (K + 2X + L)/4 \\ else \quad & X = (I + 2X + J)/4 \end{aligned} \tag{16}$$

where X is the central pixel and I, J, K and L are the four closest pixels of its that are shown in Fig. 3. With and without applying the edge preserving filter, the absolute differences calculations are estimated between two neighborhood pixels separately in horizontal and vertical directions. Then we calculate the histogram features and observe that most of the histogram values are in the lowest pixels amplitude. Here we consider only three lowest absolute difference pixel amplitudes of 0, 1 and 2 to get the major information of the image. Let h_{f0}, h_{f1}, h_{f2} and h_0, h_1, h_2 respectively be the number of absolute difference amplitude pixels with and without the edge preserving filter that have been lied on position 0, 1 and 2 on the histogram and also let H_f and H be respectively the horizontal histogram features of the image of size $M \times N$ with and without the filter, then the horizontal histogram features can be estimated as follows:

$$H_f = \frac{(h_{f0} + h_{f1} + h_{f2})}{(M-2) \times (N-2)} \tag{17}$$

$$H = \frac{(h_0 + h_1 + h_2)}{M \times N} \tag{18}$$

Similarly, we can calculate the vertical histogram features, V_f and V. Though blurring and ringing of JPEG2000 images are difficult to be evaluated without the reference image, combining the all extracted features measure gives more insight into the relative blurring and ringing in the image. There are many different ways to combine the spatial features that constitute a quality assessment model. One method that gives good prediction performance is given by the following equation:

$$\begin{aligned} C = & [\gamma_1 log(S+1) + \gamma_2 log(A+1) + \gamma_3 log(Z + \gamma_4)] \\ & \times [\gamma_5 log(H_f + 1) + \gamma_6 log(V_f + 1) + \gamma_7 log(H+1) \\ & + \gamma_8 log(V+1) + \gamma_9] \end{aligned} \tag{19}$$

where γ_1 to γ_9 are the model parameters that must be estimated with the subjective test data and optimization algorithm. In the combine equation, to avoid "log(1)", we introduce an additional one ("1") to almost all features. We consider a logistic function as the nonlinearity property between the human perception and the physical features. Finally, the obtained MOS prediction score, MOS_p, is derived by the following equation:

$$MOSp = \frac{b_1}{1 + exp[-b_2(S - b_3)]} + b_4 \tag{20}$$

where b_1, b_2, b_3, and b_4 are the parameters of the logistic function that also estimated by the PSO algorithm with the subjective test data.

4. RESULTS

To evaluate and compare the performances of our proposed model, we have considered our database and Texas database (Live Image Quality Assessment Database) and randomly divide the databases into two parts for training and testing without overlapping. In the Texas' database there are two study

Table 2. Performance evaluation on Our database (MOS scale, 1-5)

Model	*Training*				
	CC	Ave.	Max.	OR	RMSE
Proposed, NR	0.94	0.37	0.90	0	0.46
MSSIM, FR[9]	0.91	0.46	1.24	0	0.58
	Testing				
Proposed, NR	0.93	0.41	0.90	0	0.51
MSSIM, FR[9]	0.92	0.44	1.26	0	0.54

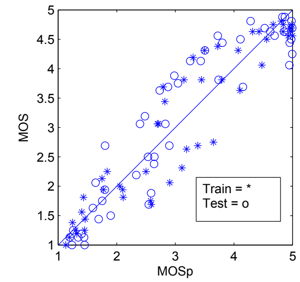
CC = Linear correlation coefficient, Ave. = Absolute average error, Max. = Absolute maximum error, OR = Outlier ratio [9], RMS = Root mean square error

Table 3. Performance evaluation on Texas' database (MOS scale, 1-100)

Model	*Training*				
	CC	Ave.	Max.	OR	RMSE
Proposed, NR	0.93	6.33	26.94	0	8.91
MSSIM, FR[9]	0.95	5.96	18.61	0	7.35
	Testing				
Proposed, NR	0.93	7.06	26.94	0	9.48
MSSIM, FR[9]	0.97	5.35	18.73	0	6.75
	All images (Train + Test)				
Proposed, NR	0.93	6.69	26.94	0	9.20
MSSIM, FR[9]	0.96	5.68	18.73	0	7.07
Sheikh *et al.*, NR[3]	0.93	8.05	N/A	N/A	N/A
Marziliano *et al.*, NR [2]	0.85	N/A	N/A	N/A	N/A

groups (group-1 and group-2) and a total of 227 JPEG2000 images with 29 reference images of different sizes. Since MOS scale of these two databases are different (MOS scale of Our database, 1-5 and Texas' database, 1-100) it is very difficult to develop the mathematical relationship between these two scales. Therefore in order to verify the generalization ability of our proposed model on the other database (Texas' database), we have considered different sets of model parameters and logistic function's parameters values. To compare our proposed model performances, we want to consider the MSSIM (general purpose, FR) [9], H. R. Sheikh *at al.* (JPEG2000, NR) [3], and P. Marziliano *at al.* (JPEG2000, NR) [2] methods. Though such comparison is unfair to one method or another in different aspects, it provides a useful indication about the relative performance of the proposed model. The performance evaluation results of our proposed model and JPEG2000 performance on MSSIM method are summarized respectively in Table 2 and 3 for our database and Texas database. The reported results on the Texas' database of [2], [3] are also shown in Table 3. It has been observed from Tables 2 and 3 that the generalization ability of proposed model is better compared to MSSIM method and performances are sufficient and also better compared to [2], [3]. The MOS versus MOS prediction (MOSp) of the proposed model on our database is shown in Figure 4. The model's parameters obtained with all of our training images are γ_1 = 34.5354, γ_2 = -37.5732, γ_3 = 42.9897, γ_4 = 1.1934, γ_5 = -6.0552, γ_6 = 6.3377, γ_7 = 6.834, γ_8 = -6.8069, and γ_9 = 0.8304 and also

Fig. 4. MOS vs MOSp results on our database (Train + Test)



the logistic parameters are b_1 = 4, b_2=1.0217, b_3=3, b_4=1.

5. CONCLUSIONS

We proposed a no-reference image quality assessment model irrespective of any predefined specific artifacts of JPEG2000 images. We claimed that any kinds of artifacts create pixel distortions and human visual perception is very sensitive to edge information. Therefore we presented a new philosophy of image quality assessment model of JPEG2000 based on pixel distortions and edge information. Though the approach is used only for JPEG2000 images, future research can be extended to generalize the approach irrespective of any kind of artifacts of different coded and distorted images.

6. REFERENCES

[1] Y. Horita, M. Sato, Y. Kawayoke, Z. M. Parvez Sazzad, and K. Shibata,"Segmentation and local features based image quality evaluation," in Proc. ICIP, Oct., 2006.

[2] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, " Perceptual blur and ringing metrics: Applications to JPEG2000," Signal Proc.:Image Commu., 19(2) pp.163-172, 2004.

[3] H. R. Sheikh, A. C. Bovik, and L. Cormack, "No reference quality assessment using natural scene statistics: JPEG2000," IEEE Trans. image processing,vol. 14, no. 11, pp. 1918-1927, Nov. 2005.

[4] Tong H., at all, "No reference quality assessment for JPEG2000 compressed images," in Proc. on image processing, 2004, Singapore.

[5] Barland, R., and A. Saadane, "Reference free quality metric for JPEG2000 compressed images," in Proc. ISSPA, 2005, Sydney, Australia.

[6] Z. Wang, "Rate scalable foveated image and video communications," PhD thesis, Dept. of ECE, The University of Texas at Austin, Dec. 2003.

[7] A. Ito and T. Murakami, "Image quality adjustment control in video encoding," PCSJ1986 pp. 75-76, 1986.

[8] Kakadu v2.2 software: http://www.kakadusoftware.com

[9] 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.