Image quality assessment models for JPEG and JPEG2000 compressed color images

Z. M. Parvez Sazzad, and Y. Horita,
Graduate School of Engineering, University of Toyama,
3190 Gofuku, Toyama, 930-8555 Japan
Phone / Fax: +81-76-445-6758
E-mail: horita@eng.u-toyama.ac.jp

Abstract

In this paper, we present a computational and memory efficient no-reference (NR) image quality assessment models for JPEG and JPEG2000 color images and also present the discrimination algorithm of these two types of images, which are applicable to various image processing applications. The proposed models and algorithm are based on blockiness around the block boundary and activity measure of the image signal within block of the image. Subjective experiment results on the two types of images are used to train the models, that achieves good quality prediction performance, and the models are also tested on a test database.

Introduction

Digital images data, stored in images databases and distributed through communication networks, are subject to various kinds of distortions during acquisition, compression, processing, transmission, and reproduction. For example, lossy image compression techniques, which are almost always used to reduce the bandwidth needed to store or transmit image data, may degrade the quality during the quantization process. Therefore any of which may create degradation result of visual quality.

In recent years, there has been an increasing need to develop objective measurement techniques that can predict image/ video quality automatically. Generally speaking, an image/video quality metric can be employed in three ways. First, it can be used to monitor image/video quality for quality control systems. Second, it can be employed to benchmark image/video processing systems and algorithms. Third, it can be embedded into image/video processing systems to optimize algorithms and parameter setting.

Objective image quality metrics can be classified according to the availability of an original image, with which the distorted image is to be compared. Most existing approaches are known as *full-reference*, meaning that a complete reference image is assumed to be known (Figure 1). In many practical applications, however, the reference image is not available, and a *no-reference* or "blind" quality assessment approach is desirable (Figure 3). In a third type of method, the reference image is only partially available, in the form of a set of extracted features made available as side information to help evaluate the quality of the distorted image. This is referred to as *reduced-reference* quality assessment (Figure 2). The most widely used objective image quality/distortion metrics are Peak Signal-to-Noise Ratio (PSNR) and Mean Squared Error (MSE), but they are widely criticized as well for not correlation well with perceived quality measurement and in the past, a great deal of effort has been made to develop new objective image/video quality metrics that incorporate perceptual quality measure by considering Human Visual System (HVS) characteristics [1]-[7]. Most of the proposed image quality assessment approaches require the original image as a reference.

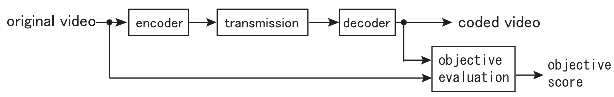

Figure 1. *Full reference model (FR-model).*

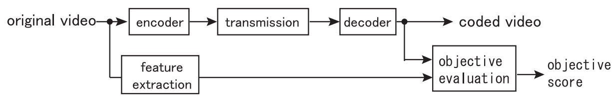

Figure 2. *Reduced reference model (RR-model).*

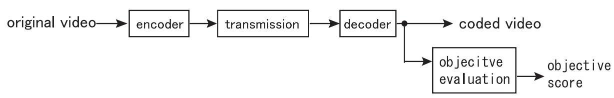

Figure 3. *No reference model (NR-model).*

Without using any reference image human observers can easily assess the quality of distorter images. By contrast, designing objective No-Reference (NR) quality measurement algorithms is a very difficult task. This is mainly due to the limited understanding of the HVS, and it is believed that effective NR quality assessment is feasible only when the prior knowledge about the image distortion types is available. Although only a limited number of methods have been proposed in the literature[8]-[15] for objective NR quality assessment, this topic has attracted a great deal of attention recently.

The blind blocking artifact measurement algorithms are proposed in [8] used a weighted mean squared difference along block boundaries as the blockiness measure. Such kind of methods cannot distinguish how much of the gray level difference between block boundaries is due to real blocking discontinuity or the oscillation of the original signal itself. Even the original image might be evaluated as, to some extent, blocky. The blind blocking effect measurement is proposed in [9] to detect and estimate the power of the blocky signal and combined it with the human visual luminance and texture masking effects. In [10] a DCT-Domain blind measurement of blocking artifacts is proposed in DCT coded images and then an HVS based measurement of blocking artifacts is conducted. Both methods are considered only JPEG gray scale images not true color images and true subjective score. In [12], an NR MPEG-2 video quality rating method is proposed, which attempted to predict PSNR by taking advantages of the quantization scale parameters available from the MPEG video stream. The second goal of this paper is to develop an objective NR quality assessment algorithm for

MPEG video, which is based on 1) an estimation of quantization errors using MPEG quantization scales and a statistics of the DCT coefficient; 2) an NR evaluation of 8 $\times$ 8 and 16 $\times$16 blocking effect; and 3) an adaptive combination of the quantization error estimation and the blocking effect evaluation using the MPEG motion vector information. In [13], an NR image quality assessment model for JPEG is proposed, which is based on blockiness and average activity measure of the image. Though the algorithms is very interesting, it's used only gray level MOS score and not very well matched to perceived visual quality. In [14], an NR blur metric is proposed, which is based on the analysis of the spread of the edges in an image. But it's studied on the limited number of compressed image and not very well matched to perceived visual quality.

In this paper, we propose no-reference quality assessment models for JPEG and JPEG2000 coded image and also the discrimination of these two types of images. The metrics are defined in the spatial domain and based on the measurement of blocking, blurring and ringing. The features are extracted and combined to constitute a quality prediction models. Though subjective experiment results on JPEG and JPEG2000 compressed images are used to train the model, which achieves very good quality prediction performance, a set of test images are also used to verify the models performance.

Subjective Experiments

The subjective experiments were conducted on 24 bit/pixel RGB color images. In these experiments, a number of human subjects were asked to assign each image with a score indicating their assessment of the quality of that image, defined as the extent to which the artifacts were visible. There were 98 images of size 768 $\times$ 512 in the database for each JPEG and JPEG2000 group. Fourteen of it were original images in each group that are shown in Figure 4. The rest of the images were JPEG and JPEG2000 coded images, i.e., 84 compressed images in each group. The six quality scales, 15, 20, 27, 37, 55 and 79 were selected for JPEG encoder and six compression ratio 12, 24, 32, 48, 72 and 96 were selected for JPEG2000 encoder. All subjects were screened prior to participate the session for normal 20/20 visual acuity with or without glasses, normal color vision and familiarity with the language. Fifteen non-expert subjects were shown the database; most of them were college student. The subjects were asked to provide their perception of quality on a discrete quality score that was divided into five and marked with the numerical value of adjectives " Bad =1," "Poor=2," "Fair=3," "Good=4," and "Excellent=5" under the test conditions of ITU-R Rec. 500-10 [15]. The fifteen scores of each image were averaged to get a final Mean Opinion Score (MOS) of the image with subject reliability of 95% confidence interval.

Image Quality assessment models

Two important image compression algorithms that use very frequently is the JPEG and JPEG2000 compression standard, which are respectively based on the block-based discrete cosine transform (DCT) and wavelet based discrete wavelet transform (DWT). Both blurring and blocking artifacts may be created during quantization of DCT coefficients in JPEG images. The blurring effect is mainly due to the loss of high frequency DCT coefficients, which smooths the image signal within each block. Blocking effect occurs due to the discontinuity at block boundaries, which is generated because the quantization in JPEG is block based and the blocks are quantized independently. In case of JPEG2000, blurring and ringing are the main artifacts. Blur is due to the attenuation of the high spatial frequencies in the image, and ringing is caused by the quantization of high frequency coefficients in wavelet transform coding. Ringing introduces ripples around sharp edges. One effective way to examine both the blurring and blocking effects is to transform the signal into the frequency domain. The blocking effect can be easily identified by the peaks at the several feature frequencies and the blurring effect is also characterized by the energy shifting from high frequency to low frequency bands [9]. A disadvantages of the frequency domain method is the involvement of the Fast Fourier Transform (FFT), which has to be calculated many times for each image, and is therefore expensive. FFT also requires more storage space because it cannot be computed locally.

In this paper, we employ a computationally inexpensive and memory efficient feature extraction method for JPEG and JPEG2000 coded image quality evaluation [15]. The features are calculated horizontally and then vertically on three components of color image Y, Cb, and Cr. The $YCbCr$ color space is used for all calculation. In the case of $YCbCr$ color space, no need any color conversion. Therefore it is used for simplicity. The range of RGB space in our calculation is [0, 255]. For luminance component (Y):

First, the blockiness is estimated as the average differences across block boundaries:

$$B_{h(y)} = \frac{1}{M([N/8]-1)} \sum_{i=1}^{M} \sum_{j=1}^{[N/8]-1} |d_{h(y)}(i,8j)| \tag{1}$$

where we denote the test image signal as $x(m, n)$ for m $\in$ [1, M] and n $\in$ [1, N], and calculate a differencing signal along each horizontal line:

$$d_{h(y)}(m,n) = x(m,n+1) - x(m,n), n \in [1,N-1]. \tag{2}$$

In all calculations we are using the absolute differences therefore the differences of the other side of the blocks, i.e., the pixels respectively located to x(m, n-1) and x(m, n) is same.

Second, we estimate the activity of the image signal. Although blurring and ringing are difficult to be evaluated without the reference image, it causes the reduction of signal activity, and combining the blockiness and activity measures gives more insight into the relative blurring and ringing in the image. The activity is measured using two factors. The first is the average absolute difference between in-block image samples:

$$A_{h(y)} = \frac{1}{7}\Big[\Big\{ \frac{8}{M(N-1)} \sum_{i=1}^{M} \sum_{j=1}^{N-1} |d_{h(y)}(i,j)| \Big\} - B_{h(y)}\Big] \tag{3}$$

The second activity measure is the zero-crossing (ZC) rate. For horizontal ZC rate:

$$d_{h-sign(y)}(M,N-1) = \begin{cases} 1 & \text{if } d_{h(y)}(m,n) > 0 \\ -1 & \text{if } d_{h(y)}(m,n) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

$$\begin{aligned} d_{h-mul(y)}(M,N-2) &= d_{h-sign(y)}(M,N-2) \\ &\times\ d_{h-sign(y)}(M,2:N-1) \end{aligned} \tag{5}$$

$$z_{h(y)}(M,N-2) = \begin{cases} 1 & \text{if } d_{h-mul(y)}(m,n) < 0 \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

The horizontal ZC rate then can be estimated as:

$$Z_{h(y)} = \frac{1}{M(N-2)} \sum_{i=1}^{M} \sum_{j=1}^{N-2} z_{h(y)}(i,j) \tag{7}$$

Using similar methods, we calculate the vertical features of $B_{v(y)}$, $A_{v(y)}$, and $Z_{v(y)}$. Finally, the overall features B_y, A_y and Z_y are given by:

$$B_y = \frac{B_{h(y)} + B_{v(y)}}{2}, A_y = \frac{A_{h(y)} + A_{v(y)}}{2}, Z_y = \frac{Z_{h(y)} + Z_{v(y)}}{2} \tag{8}$$

Similarly, we calculate the horizontal and vertical features of the color difference components Cb and Cr. That is the value of B_{cb}, A_{cb} and Z_{cb} for Cb component and B_{cr}, A_{cr} and Z_{cr} for Cr component.

There are many different ways to combine the features to constitute a quality assessment model. One method we find that gives good prediction performance is given by

$$S_y = \alpha_1 + \beta_1 B_y^{\gamma_1} A_y^{\gamma_2} Z_y^{\gamma_3} \tag{9}$$

$$S_{cb} = \alpha_2 + \beta_2 B_{cb}^{\gamma_4} A_{cb}^{\gamma_5} Z_{cb}^{\gamma_6} \tag{10}$$

$$S_{cr} = \alpha_3 + \beta_3 B_{cr}^{\gamma_7} A_{cr}^{\gamma_8} Z_{cr}^{\gamma_9} \tag{11}$$

where α_1, α_2, α_3, β_1, β_2, β_3, γ_1, γ_2, γ_3, γ_4, γ_5, γ_6, γ_7, γ_8, and γ_9 are the model parameters that must be estimated with the subjective test data such as Mean Opinion Score(MOS). The prediction equation (9), (10) and (11) are for JPEG images. The same equations with different parameters value are used for JPEG2000.

The quality degradation types between the JPEG and JPEG2000 images are different. So, it's difficult to evaluate different encoded images by using the image quality evaluation model which uses the same combine features of S_y, S_{cb} and S_{cr}. For this reason, the followings combined functions (12) and (13) are used respectively for JPEG and JPEG2000 which gives good prediction performance.

$$S = S_y S_{cb} S_{cr} \tag{12}$$

$$S = S_y S_{cb}^{\theta_1} S_{cr}^{\theta_2} \tag{13}$$

where θ_1 and θ_2 are also the model parameters for JPEG2000 that must be estimated with the subjective test data such as Mean Opinion Score(MOS).

In our experiment, the optimization of models parameters are performed by using the Particle Swarm Optimization(PSO) algorithm [16]. The models are shown in Figure 8. This mentioned models are not taken into account the nonlinearity between the human perception and the physical feature, so our image quality evaluation models consider the logistic function as the nonlinear property. Finally, obtained assessment score MOSp is derived from the following equation.

$$MOSp = \frac{4}{1 + exp[-1.0217(S-3)]} + 1 \tag{14}$$

It is generally acceptable for a quality assessment method to stably predict subjective quality within a nonlinear mapping, since the mapping can be compensated for easily. Thus, in both the VQEG Phase-I and II testing and validation, a nonlinear mapping between the objective and subjective scores was allowed, and all the performance validation metrics were computed after compensating for it. This is true for the result presentation in this work, where seven parameter nonlinearity is used.

The parameters obtained by using the PSO algorithm with all training images are shown in Table 1 and 2 respectively for JPEG and JPEG2000 and also the parameters θ_1= 0.6019 and θ_2= -0.6499 for JPEG2000.

Table 1: JPEG Model Parameters

Y	Cb	Cr
α_1 = 221.5952	α_2 = -5.7676	α_3 = 2.3609
β_1 = -213.8241	β_2 = 4.9364	β_3 = -2.8655
γ_1 = 0.0372	γ_4 = -0.0046	γ_7 = 0.027
γ_2 = -0.0342	γ_5 = 0.0385	γ_8 = 0.0387
γ_3 = -0.0029	γ_6 = 0.0526	γ_9 = -0.0243

Table 2: JPEG2000 Model Parameters

Y	Cb	Cr
α_1 = -391.201	α_2 = -5.9098	α_3 = -3.129
β_1 = 405.2078	β_2 = 6.1502	β_3 = 4.4695
γ_1 = 0.0276	γ_4 = 0.0907	γ_7 = -0.0665
γ_2 = -0.0344	γ_5 = -0.0212	γ_8 = 0.0274
γ_3 = 0.0088	γ_6 = -0.0631	γ_9 = 0.0362

Discrimination algorithm

A new discrimination algorithm is proposed only for JPEG and JPEG2000 coded images based on blockiness (B_y), average absolute difference between in-block image samples(A_y) and zero crossing rate (Z_y) is defined as:

if abs(A_y-B_y) $<$ 0.51 and Z_y $<$ 0.32 then we consider the image is JPEG2000
else if [abs(A_y-B_y) $>$ 0.51 and abs(A_y-B_y) $<$ 1.2] and Z_y $<$ 0.16 then also JPEG2000
else JPEG

In this discrimination algorithm, we have used some threshold values that are obtained by PSO optimization process. By using this discrimination information, we can easily separate the JPEG and JPEG2000 coded images and calculate their objective MOS score by using the corresponding equation and parameters value.

Discrimination accuracy

The discrimination algorithm is applied to our database as training data and the Texas' database as test data to verify it's discrimination performance. In our database, we have 84 coded images for each group of JPEG and JPEG2000. In the Texas database, they have around 190 coded JPEG and 179 coded JPEG2000 images. Considering only coded images, the discrimination accuracy for our and Texas' database are shown in Table 3 and 4.

Table 3: Performance of the discrimination algorithm based on our and Texas database of coded images

Proc.	Image	jp	j2k	err.(jp)	err.(j2k)
Train	168	84	84	2	3
Test	369	190	179	4	23

Table 4: Discrimination accuracy

Process	Accu.(JP)	Accu.(J2K)	Overall accu.
Training	97.62%	96.45 %	97%
Testing	97.89%	87.15 %	92.68%

Results

Though our subjective results are used to train the model, which achieves good quality prediction performance, we also consider another database (including either 768 × 512 or 480 × 720 pixels) of images for the model's generalization ability (Texas' database; Live quality assessment database, http://live.ece.utexas.edu/ research/quality). This database includes 233 JPEG and 227 JPEG2000 images with twenty nine original images. This database also includes some repetition of original images. The images were coded at different compres-

Figure 4. *Original images.*

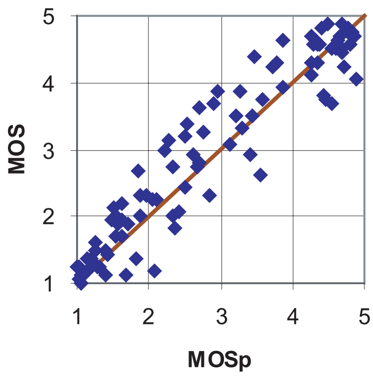

Figure 5. *MOS versus MOS prediction results of the JPEG model (SRC + HRC).*

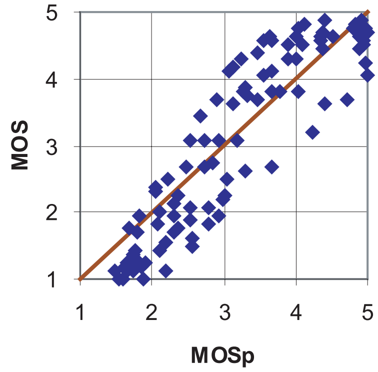

Figure 6. *MOS versus MOS prediction results of the JPEG2000 model (SRC + HRC).*

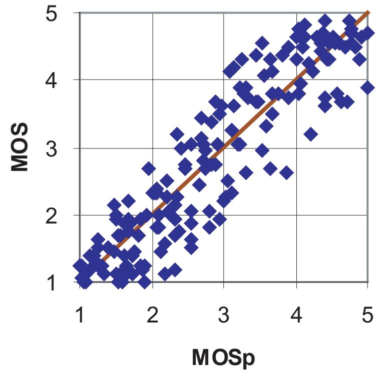

Figure 7. *MOS versus MOS prediction results of the Discrimination model (HRC).*

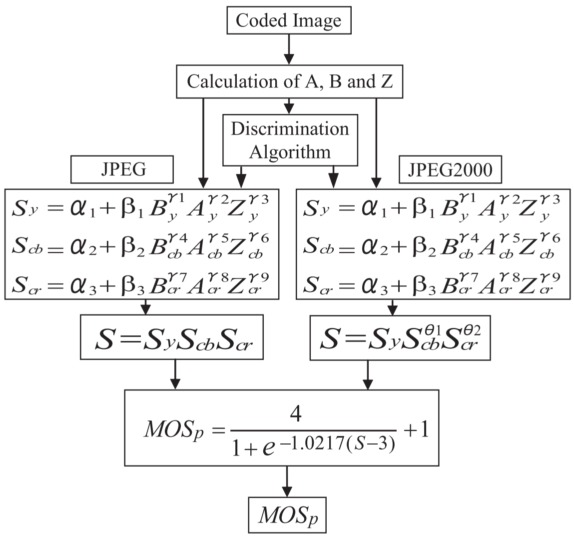

Figure 8. *NR quality assessment model.*

sion ratio. In the Texas database, the subjective experiments were conducted on 24 bit/ pixel color images and subjects were asked to measure the perceived qualities of the viewed images using a continuous linear scale divided into five regions, which was subsequently remapped linearly into the range 1 to 100. In this case, we consider a mapping function, equation (15), to convert their MOS scale, 1 to 100 to our MOS scale 1 to 5. Although this mapping function is not perfect, we try to realize our models' performance to use it.

$$MOS_5 = \frac{10}{1 + exp[-0.0219722(MOS_{100} - 100)]} \quad (15)$$

Proposed JPEG, JPEG2000, and discrimination model performance respectively are shown in Figure 5, 6, and 7. In this case, we consider only our database. Our models' estimation accuracy based on both our database and Texas database are shown in Table 5 and 6. In this case, we calculate correlation coefficient, average error and maximum error for all models.

Table 5: Estimated accuracy based on our database (SRC+HRC)

Model	Corr.	Ave.	Max.
Proposed discrimination (JP, J2K)	0.91	0.415	1.24
Ideal discrimination (JP, J2K)	0.91	0.42	1.24
Proposed JP	0.95	0.33	0.92
Proposed J2K	0.90	0.50	1.06

Table 6: Estimated accuracy based on other (Texas) database (SRC+HRC)

Model	Images	Corr.	Ave.	Max.
Proposed disc. (JP, J2K)	369	0.90	0.69	2.73
Ideal disc. (JP, J2K)	369	0.91	0.62	2.06
Proposed JP	233	0.94	0.65	1.69
Proposed J2K	227	0.94	0.46	1.45

Conclusions

New no-reference (NR) image quality assessment models for JPEG and JPEG2000 images and also the discrimination algorithm of these two types of images have been presented in this paper. The proposed models and algorithm are based on blockiness around the block boundary, average absolute difference between adjacent pixels within block, and zero crossing rate within block of the image. Though our subjective results are used to train the models, that achieves good quality prediction performance, we have shown that the models' generalization ability is also good for any other databases. Around two hundred images for each compression group have been used to verify the models. The proposed models have been given good agreement with MOS. The advantages of these models are low computational complexity and the performances are independent of the image content. Results based on other database are not so good, because the mapping function cannot transfer exactly the MOS scale, 1-100 to 1-5. The relation between the two scales are not linear, therefore it is very difficult to develop the mathematical relationship between the two scales. The discrimination algorithm can successfully discriminate almost all JPEG images from JPEG2000 images except some heavily compressed images that are used in the Texas database. In order to improve the proposed models, future research need to classify different types of blocks (flat, edge and texture) based on the value of A, B and Z into different groups and need to develop different parameter values for these groups. Beside that the correlation between the compression rate and the value of A, B and Z also need to be developed.

References

[1] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in Proc. SPIE, vol. 2179, pp. 127-141, 1994.

[2] A. M. Eskicioglu and P. S. Fisher, "Image quality measures and their performance," IEEE Trans. Communications, vol. 43, pp. 2959-2965, Dec. 1995.

[3] B. Girod, "What's wrong with mean-square error," in Digital image and Human vision (A. B. Watson, ed.), pp. 207-220, the MIT press, 1993.

[4] Z. Wang and A. C. Bovik, "A universal image quality index," IEEE Signal Processing Letter, vol. 9, pp. 81-84, Mar. 2002.

[5] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," http://www.vqeg.org/, Mar. 2002.

[6] C. J. van den Branden Lambrecht, Ed., "Special issue on image and video quality metrics," Signal Processing, vol. 70, Nov. 1998.

[7] Z. Wang, A. C. Bovik, and L. Lu, "Why is image quality assessment so difficult?," in Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing, May 2002.

[8] H. R. Wu and M. Yuen, " A generalized block-edge impairment metric for video coding," IEEE Signal Processing Letters, vol. 4, pp. 317-320, Nov. 1997.

[9] Z. Wang, A. C. Bovik, and B. L. Evans, " Blind measurement of blocking artifacts in images," in Proc. IEEE Int. Conf. Image Proc., vol. 3, pp. 981-984, Sept. 2000.

[10] A. C. Bovik and S. Liu, " DCT- domain blind measurement of blocking artifacts in DCT-coded images," in Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing, vol. 3, pp. 1725-1728, May 2001.

[11] P. Gastaldo, S. Rovetta, and R. Zunino, "Objective assessment of MPEG-video quality: a neural-network approach," in Proc. IJCNN, vol. 2, pp. 1432-1437, 2001.

[12] M. Knee, " A robust, efficient and accurate signal ended picture quality measure for MPEG-2," available at http://www-ext.crc.ca/vqeg/frames.html, 2001.

[13] Z. Wang, H. R. Sheikh, and A. C. Bovik , " No-Reference perceptual quality assessment of JPEG compressed images," in Proc. IEEE ICIP, pp. I-477-480, Sept. 2002.

[14] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," in Proc. of the Int. Conf. on Image Processing, vol. 3, pp. 57-60, Rochester, NY, 2002.

[15] ITU-R Recommendation BT.500-10, "Methodology for the subjective assessment of the quality of television pictures"

[16] J. Kennedy and R. Eberhart, "Particle Swarm Optimization," Proc. IEEE ICNN, Perth, Australia, pp. 1942-1948, Nov. 1995.

Author Biography

Z. M. Parvez Sazzad received his B.Sc(Hons.) and M.Sc degree respectively in 1993 and 1994 from the department of Applied Physics and Electronics, University of Dhaka, Bangladesh. He is currently working toward Ph.D degree in the Course of System Science and Engineering, University of Toyama, Japan. He served as lecturer (1997-1999) in the *department of Computer Science and Engineering, Darul Ihsan University, Dhaka and (1999-2003) and (2003-Sept. 2004) respectively lecturer and assistant professor in the the department of Applied Physics and Electronics, University of Dhaka, Bangladesh.*

Y. Horita received the B.E., M.E. and Ph.D degree in Electrical and Electronics Engineering from Nagaoka University of Technology, Nagaoka, Japan, in 1984, 1986 and 1995, respectively. In 1995, he joined the Department of Electronics and Computer Science, Faculty of Engineering, University of Toyama, Toyama, where he is currently a Professor. He is engaged in research on objective quality estimation of coded picture, computer vision and image processing. He is a member of IS&T, IEICE, IIITE, IPSJ, IEEE, OSA and EURASIP.