Signal Processing: *Image Communication* 23 (2008) 257–268

Contents lists available at ScienceDirect

Signal Processing: *Image Communication*

journal homepage: www.elsevier.com/locate/image

No reference image quality assessment for JPEG2000 based on spatial features

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

Graduate School of Science and Engineering, University of Toyama, 3190 Gofuku, Toyama 930-8555, Japan

ARTICLE INFO

Article history:
Received 6 June 2007
Received in revised form
30 November 2007
Accepted 10 March 2008

Keywords:
Image quality evaluation
JPEG2000
No-reference (NR)
Mean opinion score (MOS)
Mean opinion score prediction (MOSp)
Zero-crossing (ZC)

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 approach for designing a no reference image quality evaluation model for JPEG2000 images in this paper, which uses pixel distortions and edge information. Subjective experiment results on the images are used to train and test the model, which has achieved good quality prediction performance.

1. Introduction

It is an ever increasing requirement to send more multimedia data over tighter bandwidth which has been driven to develop advanced compression technology. Due to the advanced development of different image compression techniques and processing systems, there is a very big concern about the levels of image quality both for providers and users in many image processing applications from compression to printing. Obviously, digital images suffer a wide variety of distortion in these applications and perceptual quality of the images are degraded. Therefore, perceptual image quality measurement is an important problem. Though the subjective test is considered to be the most accurate method since it reflects human perception, it is time consuming and expensive. Furthermore, it cannot be done in real time. As a result, developing objective image quality evaluation methods are getting more attention nowadays. There are three types of methods that are used for objective image quality evaluation: *full-reference* (FR), *reduced-reference* (RR) and *no-reference* (NR). In the FR method, a reference/original image is required to assess the quality of the distorted image. Therefore, it is highly desirable to develop a quality assessment method that does not require full access to the reference images. In the RR method, some extracted features of the reference/original image are required to assess the quality. However, in many practical applications, the reference image is not available and an NR quality assessment approach is desirable.

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, among others things, for not correlation well with perceived quality measurement. In the past, a great deal of effort has been made to develop new objective image/video quality metrics that incorporate perceptual quality measures by considering human visual system (HVS) characteristics [1–6]. Most of the proposed image quality assessment approaches require the original image as a reference.

* Corresponding author. Tel./fax: +81 76 445 6758.
E-mail addresses: d0570004@ems.u-toyama.ac.jp (Z.M. Parvez Sazzad), horita@eng.u-toyama.ac.jp (Y. Horita).

0923-5965/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.image.2008.03.005

Nevertheless, human beings do not need to have access to the reference image to make judgements regarding quality. Human observers can easily assess the quality of distorted images without using any reference image. By contrast, designing objective 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 prior knowledge about the image distortion types is available. Although only a limited number of methods have been proposed in the literatures for objective NR quality assessment, this topic has attracted a great deal of attention recently [7–19]. Since the predominant mode for image and video coding and transmission is using block-based video compression algorithms, blind measurement of the blocking artifact has been the main emphasis of NR quality assessment researches [7–13]. Blockiness, activity and segmentation-based measure are explained either in the spatial domain [7–11] or in the frequency domain [12,13].

However, the above described methods would obviously fail for any other distortion types, such as ringing or blurring introduced by the JPEG2000 image compression algorithm, or the H.264 video compression algorithm. Some researchers have attempted to quantify the blurring and ringing artifacts without reference. In [14], a visible ringing measure (VRM) is proposed that captures the ringing artifact around strong edges. The algorithm is based on constructing an image mask that exposes only those parts of the image that are in the vicinity of strong edges, and the ringing measure is considered to be the pixel intensity variance around the edges in the masked image. Still, the measure was not compared to the human score of quality. In [15,16] an NR blur metric is proposed based on measuring average edge transition widths, and this blur measure was used to predict the quality of JPEG2000 compressed images. In [17], an NR algorithm is proposed based on natural scene statistics. In [18], a principal component analysis is performed on edge points, beforehand classified as distorted or not, in order to measure both blurring and ringing effects, as well as the combination of spatial ringing and blurring measures, which are also presented in [19].

All of the proposed NR perceptual image quality assessment algorithms are implemented according to the predefined specific artifacts of specific coders [7–19]. A lot of NR quality evaluations have already focus on measuring the blocking artifacts, specially for JPEG images with quite sufficient correlation with the subjective scores. However, very few NR evaluations have been performed for JPEG2000 images and these evaluations' generalization ability and performances are not widely verified and well matched to the subjective scores. Whereas, nowadays, the JPEG2000 coder is getting more attention compared to the JPEG coder, due to its high coding performance, although JPEG was previously the standard coder for still image. The JPEG2000 coder served better in many image processing applications, such as digital cameras, 3G mobile phones, video streaming, printers, scanners, high quality frame-based video recording, nonlinear video editing, storage, etc. Specifically, motion JPEG2000 is the leading digital film standard currently supported by digital cinema initiatives for the storage, distribution and exhibition of motion pictures.

In this research, we propose a new method for NR quality evaluation of JPEG2000 images, irrespective of any predefined specific artifacts based on pixel distortions and edge information measure. This type of quality assessment is used to assess image quality and produces results comparable to those of subjective scores. The subjective experiment results on our database (JPEG2000 color images) were used to train and test the model and it achieved a sufficient quality prediction performance. The other database was also used to verify the model's performance. We report that the performance of the model is sufficient and reliable.

2. Subjective experiments

2.1. Our database [22]

We conducted subjective experiments on 24 bits/pixel RGB color images on our database. There were 98 images of size 768×512 in the database of JPEG2000. Out of all, 14 were reference images that are shown in Fig. 1. The rest of the images were JPEG2000 coded. Six compression ratios (CR: 12, 24, 32, 48, 72 and 96) were selected for the JPEG2000 encoder [20]. Single stimulus (SS) adjectival categorical judgement method was used in these subjective experiments. Prior to participating the session all subjects were screened for normal visual acuity with or without glasses, normal color vision and familiarity with language. Sixteen non-expert subjects were shown the

Fig. 1. Reference images (gray scale display only).

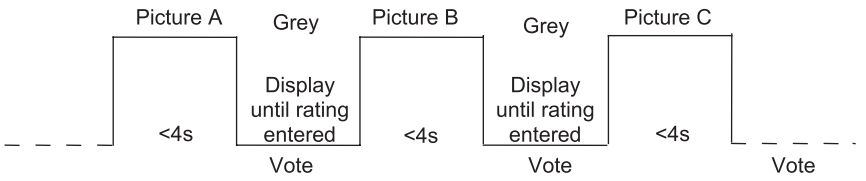

Fig. 2. SS adjective categorical judgement test format.

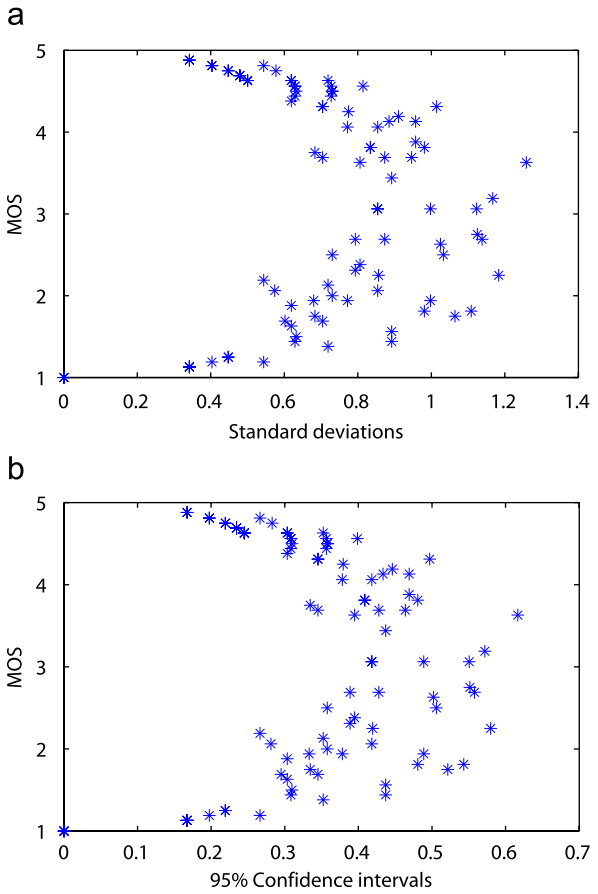

Fig. 3. Standard deviations and 95% confidential intervals of all MOSs. (a) MOS versus STD. (b) MOS versus 95% CI.

database; most of them were college students. 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 BT.500-10 [21]. Note that the numerical values attached to each category were only used for data analysis and were not shown to the subjects. The test format is shown in Fig. 2. At the end of each test presentation, the human judges ("subjects") provided a quality rating using the adjective scale. The test presentation order was randomized according to standard procedure and the raw scores were collected in a data file by computer. Mean opinion scores (MOSs) were then computed for each image after the screening of post-experiment results (most subjects had no outliers) according to

Table 1
Subjective test conditions and parameters

Method	Single stimulus (SS)
Evaluation scales	5 Grades (adjective scales)
Images	24-bits/pixel RGB color (768 × 512)
Reference images	14
Coder	JPEG2000
Coding parameters [20]	6 (CR: 12, 24, 32, 48, 72 and 96)
Viewers	16 (non-expert, college students)
Display	CRT 17-in
Viewing distance	4H (H: picture height)
Monitor resolution	1024 × 768
Room illumination	Low

ITU-R Rec. 500-10. Sixteen scores for each image were averaged to get a final MOS of the image. Standard deviations (STDs) and 95% Confidence intervals (CI) of all MOSs of images are shown in Fig. 3. Details of the database are available online [22]. The subjective test conditions and parameters are summarized in Table 1.

2.2. *LIVE Texas' database [27]*

In the Texas' database, they conducted subjective experiments 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 of 1–100. There were two study groups (group-1 and group-2) and a total of 227 JPEG2000 images with 29 reference images of different sizes. The images were coded at a wide range compression ratio. See database details in [27].

3. Proposed model

Many researches have already established that the main function of the HVS is to extract structural or edge information from the viewing field, and the HVS is highly adapted for this purpose [4,5]. Under the assumption that human visual perception is very sensitive to edge information, natural image signals are highly structured, specifically the samples of the signals have strong dependencies between each other, especially when they are close in space. Therefore, any kind of artifacts creates pixel distortions from neighborhood pixels. As a result, in this research, we want to develop a new NR 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

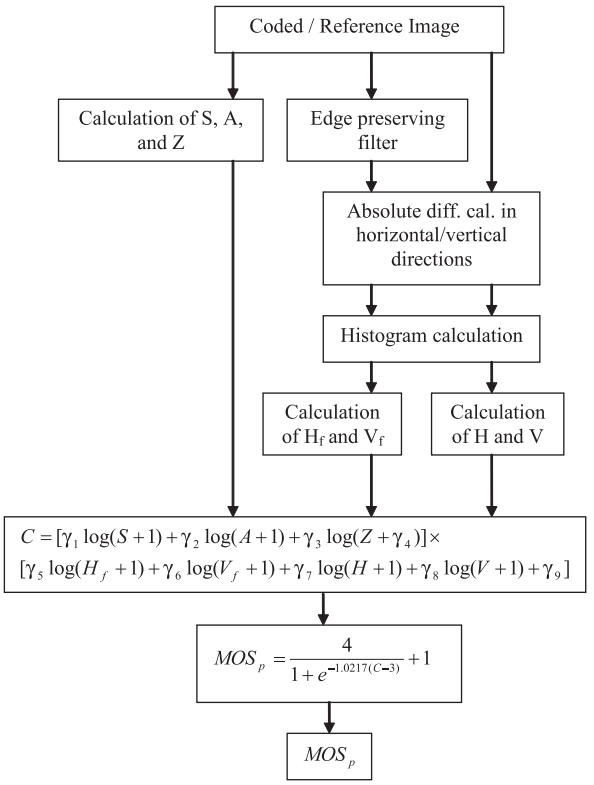

Fig. 4. Proposed model.

domain. We attempt to design a memory efficient and computationally inexpensive feature extraction method. A previous instantiation of this approach was made in [23] and promising results on simple tests were achieved. In this paper, we generalize this algorithm, and provide a more extensive set of validation results. The block diagram of the model is shown in Fig. 4. For simplicity, only the luminance component is considered to make quality prediction of the color images, though this should not be considered generally true for color image quality assessment. We believe that the incorporation of color information could further improve performance [6,11].

3.1. Pixel distortions measure

Pixel distortions are estimated using two features. First, STD of a central pixel is estimated within 5×5 neighborhood pixels which is applied for all available central pixels in the image. STD values are then averaged within a 5×5 partially overlapping block. Let X_{13} be the central pixel of the 5×5 pixels "neighborhood" as shown in Fig. 5(a); also let $\overline{X}$ and "std" be the mean of pixels within the 5×5 pixels, and the standard deviation of X_{13} pixel in the 5×5 pixels. The statistical features can be estimated as follows:

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

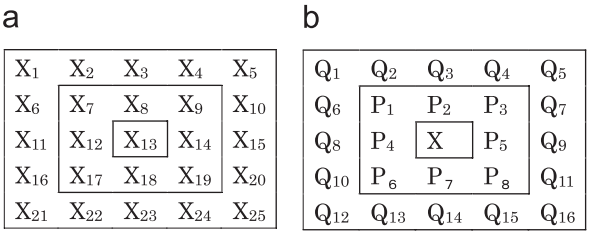

Fig. 5. The 5×5 neighborhood pixels: (a) central pixel X_{13} of the 5×5 pixels. (b) Central pixel X and its first (P_1 to P_8) and second closest (Q_1 to Q_{16}) neighborhoods of the 5×5 pixels.

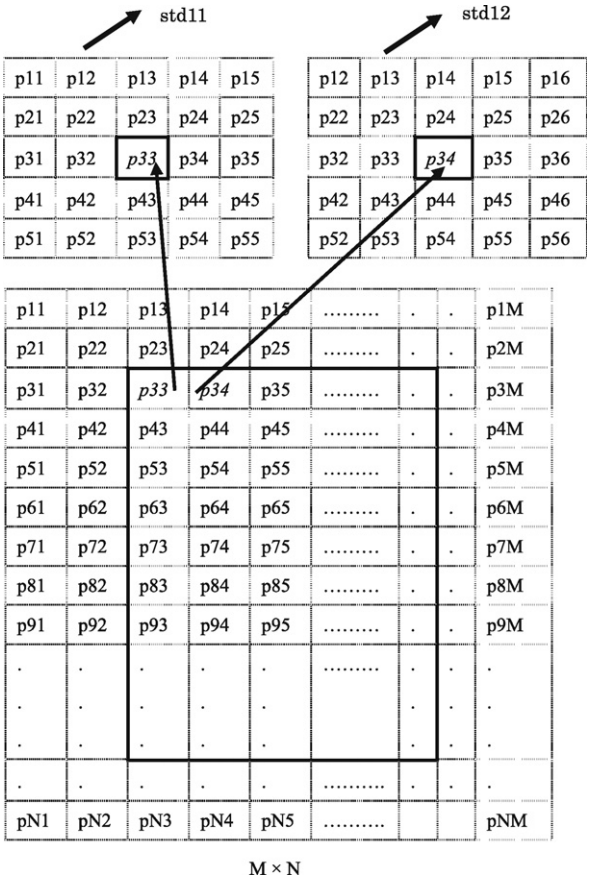

Fig. 6. Image pixels.

where $L=25$ is the total number of pixels within 5×5 neighborhoods.

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

Thus we calculate "std" values of all available central pixels of an image. For a clear understanding, we denote the test image signal as $x(m,n)$ for $m\in[1,M]$ and $n\in[1,N]$ $(M>N)$, where all available central pixels of the image are located within the bolded line boundary as shown in Fig. 6. Consequently, the "std" values of all

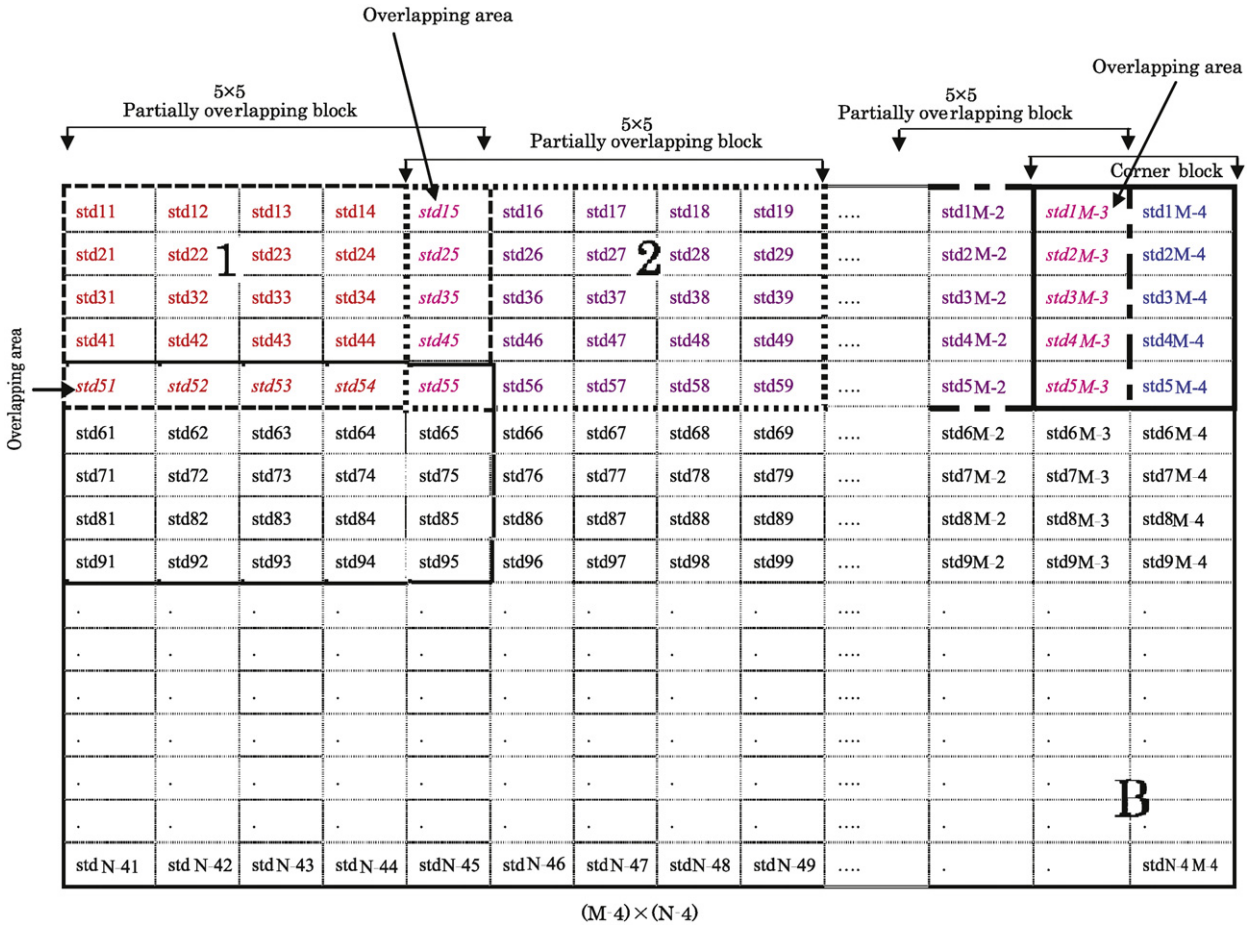

Fig. 7. "std" Values of all available central pixels of an image ("std" plane).

available central pixels are shown in Fig. 7. Then we average the STD values from within each 5×5 partially overlapping block of "std" values of the "std" plane (only corner "std" values are overlapped. "std" plane means a hypothetical plane consists of "std" values of all available central pixels of an image). Fig. 7 portraits the way we consider 5×5 partially overlapping blocks. Let $\overline{\text{std}}$ be the average standard deviation within each 5×5 partially overlapping block of "std" values. The $\overline{\text{std}}$ can be estimated by

$$\overline{\text{std}} = \frac{1}{25}\sum_{i=1}^{25} \text{std}_i \tag{3}$$

However, if one or more rows or columns of the overlapping block (5×5) of "std" plane are located outside the "std" plane ($M-4 \times N-4$), there are several ways to handle this situation. In this research we use a simple method to limit the excursions of the block to the boundary and calculate the average within the available elements of "std" in the block, whereas the block will be considered as a full block. Following, if the total number of 5×5 partially overlapping blocks is "B" (see Fig. 7), the 5×5 partially overlapping STD feature (S) of the image is finally estimated by the following equation:

$$S = \frac{1}{B}\sum_{i=1}^{B} \overline{\text{std}_i} \tag{4}$$

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

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

We then calculate the A_d values of all available central pixels of the image and get "A_d" plane (like "std" plane of Fig. 7). Available pixels are the same with previous calculations of the S measure, located within the bolded line boundary in Fig. 6. The next step is averaging the A_d

values from within each 5×5 partially overlapping block of A_d values of the "A_d" plane (like Fig. 7). Let $\overline{A_d}$ be the average A_d of within each 5×5 partially overlapping block. The $\overline{A_d}$ can be estimated by

$$\overline{A_d} = \frac{1}{25}\sum_{i=1}^{25} A_{d_i} \tag{6}$$

Finally, the 5×5 partially overlapping average absolute difference feature (A) of the image is estimated by the following equation:

$$A = \frac{1}{B}\sum_{i=1}^{B} \overline{A_{d_i}} \tag{7}$$

where B is the total number of 5×5 partially overlapping block (like previous feature S).

3.2. Edge information measure

Edge information is estimated using two features. First, zero-crossing (ZC) rate is estimated in both horizontal and vertical directions of the image.

For horizontal direction: Since the test image signal is $x(m,n)$ for $m \in [1,M]$ and $n \in [1,N]$, a differencing signal along each horizontal line is calculated by

$$d_h(m,n) = x(m,n+1) - x(m,n)$$
$$n \in [1,N-1] \text{ and } m \in [1,M] \tag{8}$$

For horizontal ZC:

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

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

$n \in [1,N-2]$:

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

where the size of $z_h(m,n)$ is $M \times (N-2)$. The horizontal ZC rate, Z_{bh}, is calculated within the 5×5 partially overlapping block of $z_h(m,n)$:

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

Finally, if the total number of 5×5 partially overlapping blocks is "BH" (like Fig. 7), the average horizontal ZC rate, Z_h, is estimated as follows:

$$Z_h = \frac{1}{BH}\sum_{i=1}^{BH} Z_{bh_i} \tag{13}$$

Here the value of BH is different compared to B (Eqs. (4) and (7)) because the size of "std" and A_d planes ($M-4 \times N-4$) are different compared to $z_h(m,n)$ ($M \times N-2$).

For vertical direction: We can also calculate a differencing signal along each vertical line:

$$d_v(m,n) = x(m+1,n) - x(m,n)$$
$$n \in [1,N] \text{ and } m \in [1,M-1] \tag{14}$$

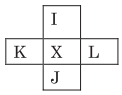

Fig. 8. Masking block of the edge preserving filter.

Similarly, we can calculate the average vertical ZC rate, Z_v and finally the overall feature of ZC rate:

$$Z = \frac{Z_h + Z_v}{2} \tag{15}$$

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 [24]:

$$H_d = K - 2X + L \tag{16}$$

$$V_d = I - 2X + J \tag{17}$$

$$\text{if } (H_d < V_d), \quad X = (K + 2X + L)/4$$
$$\text{else } X = (I + 2X + J)/4 \tag{18}$$

where X is the central pixel and I,J,K and L are the four closest pixels that are shown in Fig. 8. With and without applying the edge preserving filter, the absolute difference calculations are estimated between two neighborhood pixels separately in horizontal and vertical directions. Then we calculate the 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 (see Appendix). 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:

$$H_f = \frac{(h_{f0} + h_{f1} + h_{f2})}{(M-2) \times (N-2)} \tag{19}$$

$$H = \frac{(h_0 + h_1 + h_2)}{M \times N} \tag{20}$$

Similarly, we can calculate the vertical histogram features, V_f and V. Although blurring and ringing of JPEG2000 images are difficult to be evaluated without the reference image, combining all the extracted features measure gives more insight into the relative blurring and ringing in the image.

3.3. Features combination

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

by the following equation [23]:

$$C = [\gamma_1 \log(S+1) + \gamma_2 \log(A+1) + \gamma_3 \log(Z+\gamma_4)] \times [\gamma_5 \log(H_f+1) + \gamma_6 \log(V_f+1) + \gamma_7 \log(H+1) + \gamma_8 \log(V+1) + \gamma_9] \quad (21)$$

where γ_1 to γ_9 are the model parameters that must be estimated with the subjective test data and optimization algorithm [25]. Particle swarm optimization (PSO) is used for these parameters optimization, because it is one of the powerful optimization algorithms. 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, $\mathrm{MOS_p}$, is derived by the following equation [26]:

$$\mathrm{MOS_p} = \frac{4}{1+\exp[-1.0217(C-3)]} + 1 \quad (22)$$

In order to verify extensively the performance of our proposed model, we consider our database (MOS scale, 1–5, see Section 2.1) [22] and LIVE Texas' database (MOS scale, 1–100, see Section 2.2) [27]. Since our proposed model is using two different MOS scale related databases, it is very difficult to develop a mathematical relationship between these two scales. Moreover, it would not be wise to correlate between two different subjective scales because the two different subjective experiments were conducted by two different sets of viewers with different language, culture and behavior, in two different labs, in different continents of the world. Although Pinson and Wolf [28] presented a mapping method to convert one subjective scale to another, the performance was not good for all subjective data sets. Therefore, we do not combine the two different subjective experimental data, but instead consider the real subjective scores of both databases. As a result, we estimate the model's parameter values separately for the two databases.

In order to use our database, we divide the database into two parts for training and testing. The training data set consists of seven randomly selected images (from the total 14) and all of their distorted versions. The testing data set consists of the other seven images and their distorted versions, and also there is no overlapping between training and testing. The model's parameters are obtained by the PSO algorithm with all of our training images are shown in Table 2.

In order to use the Texas' database, first we have merged the two study groups' (group-1 and group-2) images into one study group and then randomly divided into two halfs for training and testing. There is also no overlap between the training and testing sets. The new model parameters for the different subjective scale (MOS scale, 1–100) must be estimated with the subjective test data and optimization algorithm. In this case, we consider the following (Eq. (22) for MOS scale 1–5) similar logistic function as the nonlinear property of the model to calculate the final MOS prediction ($\mathrm{MOSp_{(100)}}$) values for Texas' database:

$$\mathrm{MOSp_{(100)}} = \frac{b_1}{1+\exp[-b_2(C-b_3)]} + b_4 \quad (23)$$

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. The model's parameters are obtained by the PSO algorithm with all of Texas training images are shown in Table 3. The obtained parameters of the logistic function are $b_1 = 78.0058$, $b_2 = 1.0346$, $b_3 = 49.6925$ and $b_4 = 2.2622$. Model's performances on both training and testing sets of our and Texas' databases are explained in Section 4.

Table 2
Model parameters for MOS scale, 1–5 (our database)

$\gamma_1 = 34.5354$	$\gamma_2 = -37.5732$	$\gamma_3 = 42.9897$
$\gamma_4 = 1.1934$	$\gamma_5 = -6.0552$	$\gamma_6 = 6.3377$
$\gamma_7 = 6.834$	$\gamma_8 = -6.8069$	$\gamma_9 = 0.8304$

Table 3
Model parameters for MOS scale, 1–100 (Texas' database)

$\gamma_1 = 2.8507$	$\gamma_2 = -3.4735$	$\gamma_3 = 22.1784$
$\gamma_4 = 2.2957$	$\gamma_5 = 0.0096$	$\gamma_6 = 0.3619$
$\gamma_7 = -0.3168$	$\gamma_8 = 0.0452$	$\gamma_9 = 2.7841$

4. Results

In order to verify our proposed model performance against other quality assessment algorithms, we want to consider a general purpose FR model and application specific NR models. These models include MSSIM (general purpose, FR) [6], Sheikh et al. (JPEG2000, NR) [17], and Marziliano et al. (JPEG2000, NR) [16]. Although 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.

To the best of our knowledge, no other NR model performances have been verified on other subjective databases. Most of the authors have used their own databases for training and testing. We use our database (training + testing) and Texas' database (training + testing) to verify extensively the generalization ability and reliability of the proposed model (see Section 3.3). In order to provide quantitative measures on the performance of our proposed NR quality assessment model, we follow the standard performance evaluation procedures employed in the video quality experts group (VQEG) FR-TV Phase II test [29], where mainly three evaluation metrics were used (*Metric* 1, *Metric* 2 and *Metric* 3).

Metric 1: Pearson linear correlation coefficient (CC) between objective (MOSp) and subjective (MOS) scores. It provides an evaluation of the *prediction accuracy*.

Metric 2: Spearman rank-order correlation coefficient (SROCC) between objective (MOSp) and subjective (MOS) scores. It is considered as a measure of *prediction monotonicity*.

Metric 3: Outlier ratio (OR) of "outlier-points" to total points N.

Outlier ratio = (total number of outliers)/N,

where the outlier is defined as the prediction points outside the range of ±2 STDs between subjective scores. It is considered as a measure of *prediction consistency*.

Table 4
Performance evaluation on our database (MOS scale, 1–5)

Model	CC	AAE	MAXE	OR	RMSE	SROCC
Training (49 images)						
Proposed, NR	0.94	0.37	0.90	0.0204	0.46	0.99
MSSIM, FR [6]	0.91	0.46	1.24	0.1020	0.58	0.99
Testing (49 images)						
Proposed, NR	0.93	0.41	0.90	0.0408	0.51	0.99
MSSIM, FR [6]	0.92	0.44	1.26	0.1020	0.54	0.99

Table 5
Performance evaluation on Texas' database (MOS scale, 1–100)

Model	CC	AAE	MAXE	OR	RMSE	SROCC
Training (114 images)						
Proposed, NR	0.93	6.33	26.94	0.0350	8.91	0.96
MSSIM, FR [6]	0.95	5.96	18.61	0.0263	7.35	0.98
Testing (113 images)						
Proposed, NR	0.93	7.06	26.94	0.0442	9.48	0.96
MSSIM, FR [6]	0.97	5.35	18.73	0.0088	6.75	0.98
All images (227 images) (train + test)						
Proposed, NR	0.93	6.69	26.94	0.0396	9.20	0.99
MSSIM, FR [6]	0.96	5.68	18.73	0.0176	7.07	0.99
Sheikh et al., NR [17]	0.93	8.05	N/A	N/A	N/A	N/A
Marziliano et al., NR [16]	0.85	N/A	N/A	N/A	N/A	N/A

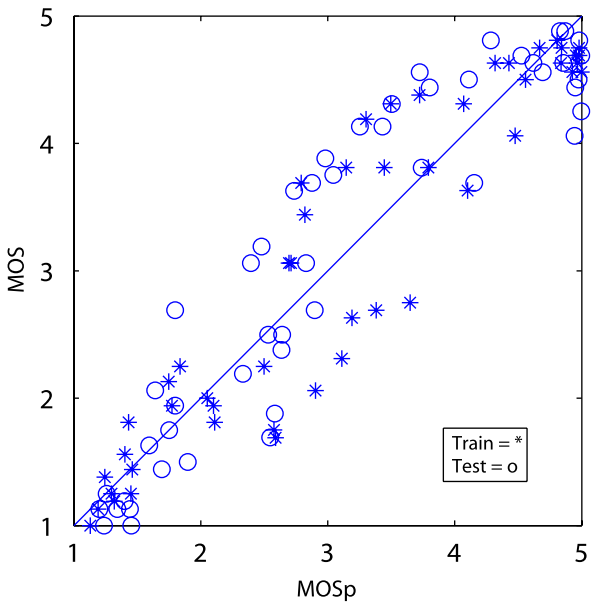

Fig. 9. Proposed model performance on our database (train + test).

More details on these metrics can be found in [29]. In addition to these, we also calculate:

Metric 4: Root mean square prediction error (RMSE) between objective (MOSp) and subjective (MOS) scores.

Metric 5: Average absolute prediction error (AAE) between objective (MOSp) and subjective (MOS) scores.

Metric 6: Maximum absolute prediction error (MAXE) between objective (MOSp) and subjective (MOS) scores.

The evaluation results of our proposed model and JPEG2000 performance on MSSIM (general purpose, FR) method are summarized, respectively, in Tables 4 and 5 for our database and Texas' database. It has been observed from Tables 4 and 5 that the proposed model performances for every one of the evaluation metrics are

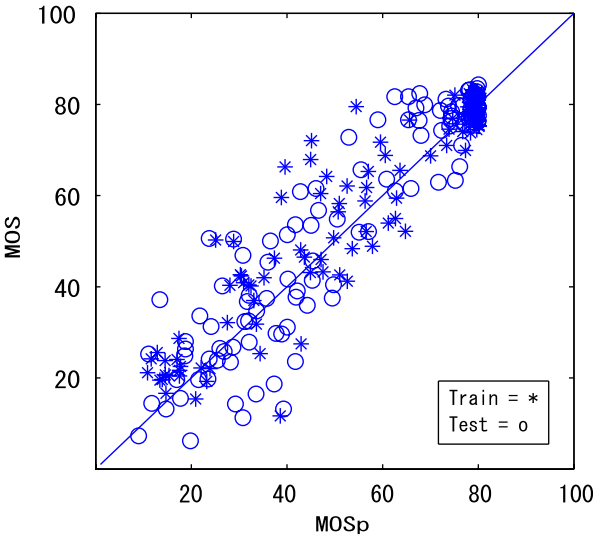

Fig. 10. Proposed model performance on Texas' database (train + test).

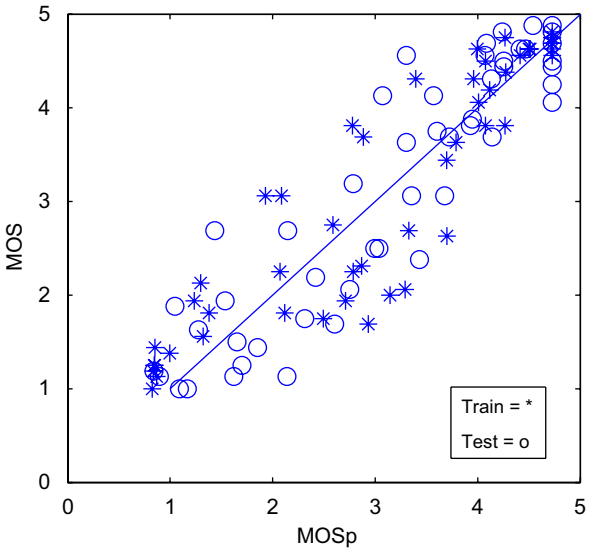

Fig. 11. MSSIM method performance on our database (train + test).

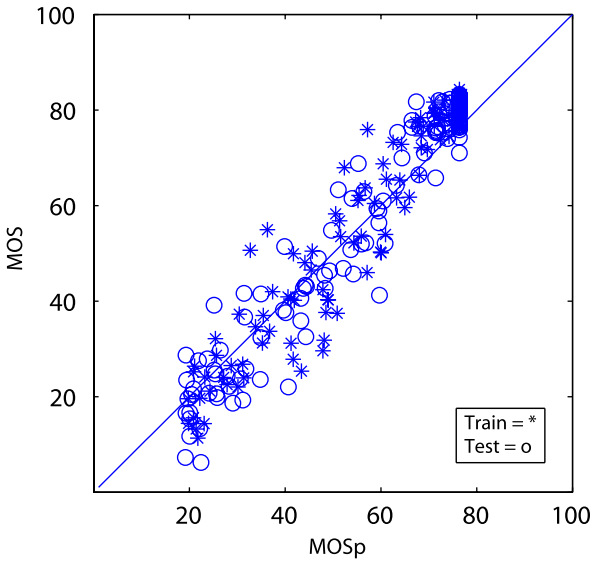

Fig. 12. MSSIM method performance on Texas' database (train + test).

sufficient not only for our database but also for Texas' database. Although MSSIM's performance is better on Texas' database, the performance is inferior on our database. Whereas the proposed model performance is sufficient and stable both for the databases compared to MSSIM method, which proves sufficient generalization ability. It has also been observed from Tables 4 and 5 that our proposed model provides sufficient prediction accuracy (higher CCs), sufficient prediction monotonicity (higher SROCCs) and sufficient prediction consistency (lower outlier ratios). In comparison, we can compare the performance of our proposed model against the NR perceptual blur metric presented in [16]. Although this blur metric was not specifically designed for JPEG2000, the reported CC on the Texas' database was 0.85, whereas our proposed model CC on the same database is 0.93. Another insightful comparison of our proposed model against recently published NR metric for JPEG20000 based on natural scene statistics in [17] is that the reported CC and absolute average error (AAE) on the Texas' database were, respectively, 0.93 and 8.05. Our proposed model's CC and AAE on the same database are 0.93 and 6.69, which is better compared to [17]. The MOS versus MOS prediction (MOSp) of the proposed model and MSSIM method for our database and Texas' database are, respectively, shown in Figs. 9–12. It is clear from these figures, Tables 4 and 5, that our proposed model performances are sufficient for both databases and have better generalization ability compared to MSSIM [6], as well as better performance compared to [16,17].

5. Conclusions

In this paper, 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 approach of image quality assessment model of JPEG2000 based on pixel distortions and edge information. The proposed model had been given good agreement with the MOS. Although 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.

Acknowledgments

The authors would like to thank Dr. H.R. Sheikh for supplying the LIVE Quality Assessment Database (http://live.ece.utexas.edu/research/quality). The authors would also like to thank Prof. Murat Kunt and Prof. Touradj Ebrahimi for their higher motivation about the image quality evaluation methodology when he stayed at EPFL for visiting researcher supported by SNSF & JSPS.

Appendix

In this appendix, you can see the histogram plots of reference and corresponding coded images with and without the edge preserving filter. We have considered one reference image with a variety of textures and its coded image. In the filter mask shown in Fig. 8, the pixel at the center of the mask is multiplied by a higher value than others, thus giving this pixel more importance in the calculation of the average. The basic strategy of the filter is to preserve edge in the image. With and without applying the edge preserving filter, the absolute difference calculations are estimated between two neighborhood pixels separately in horizontal and vertical directions by the following equations:

With filter:

$$D_{fh}(m,n) = |x_f(m,n+1) - x_f(m,n)|, \quad n \in [1, N-3] \tag{24}$$

$$D_{fv}(m,n) = |x_f(m+1,n) - x_f(m,n)|, \quad m \in [1, M-3] \tag{25}$$

where the filtered image signal is $x_f(m,n)$ for $m \in [1, M-2]$ and $n \in [1, N-2]$.

Without filter:

$$D_h(m,n) = |x(m,n+1) - x(m,n)|, \quad n \in [1, N-1] \tag{26}$$

$$D_v(m,n) = |x(m+1,n) - x(m,n)|, \quad m \in [1, M-1] \tag{27}$$

where the image signal is $x(m,n)$ for $m \in [1, M]$ and $n \in [1, N]$.

After that, we calculate the histogram features. Horizontal and vertical histogram plots are shown in Figs. 13 and 14. It has been observed from Figs. 13 and 14 that most of the absolute difference pixels (ADP) amplitudes are in the low level range of the horizontal scale and these ADP amplitudes are slightly diverged to a high level range of horizontal scale compared to those without a

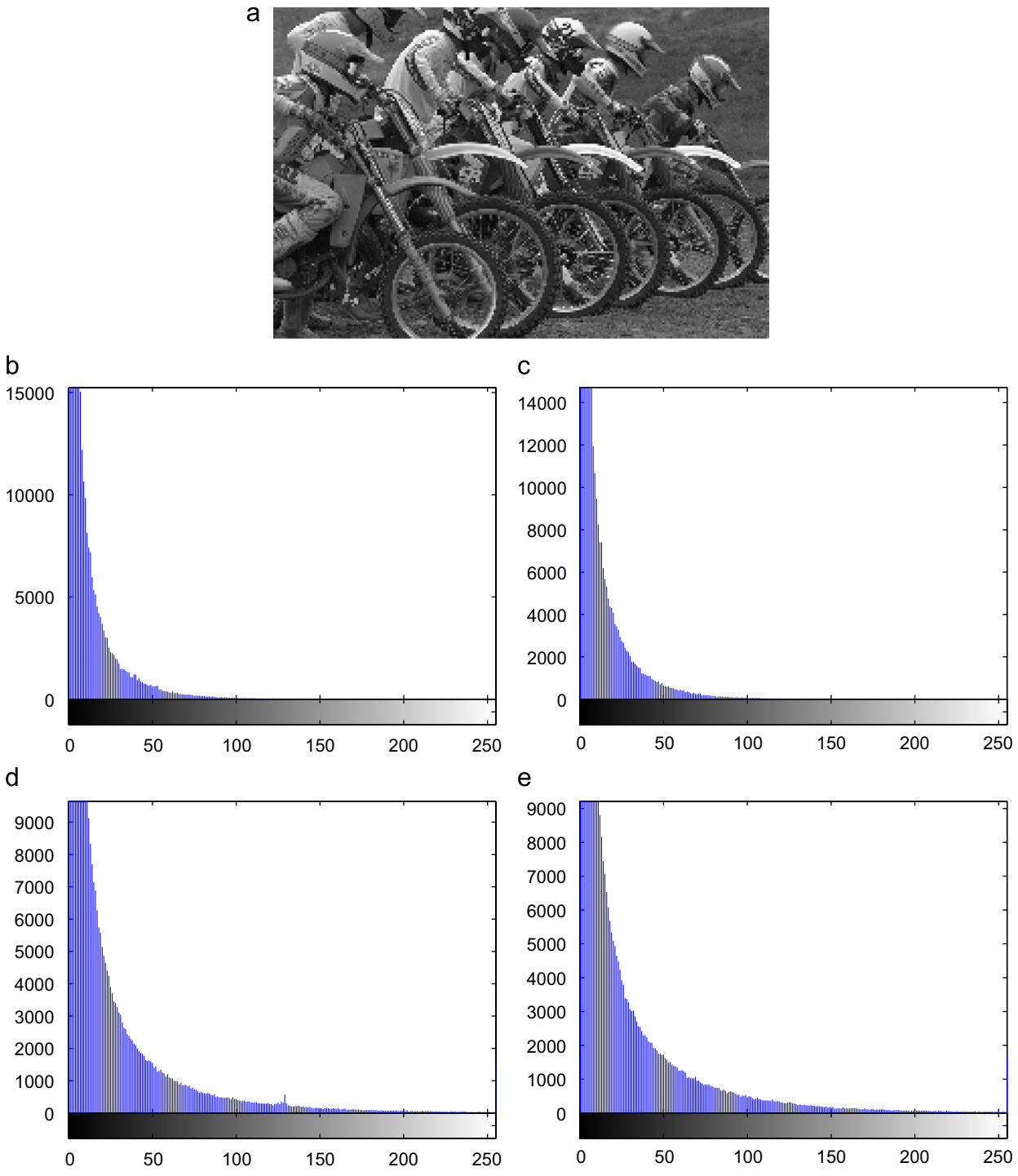

Fig. 13. Histogram plots of the (a) reference image; without filter (b) horizontal, (c) vertical, and with filter (d) horizontal, (e) vertical.

filter. Because of the multiplication and averaging of filter operations, the ADP amplitude of edge and texture pixels are increased and diverged to a high level range of the horizontal scale. Therefore, considering only low level horizontal scale value of the histogram with and without filter gives relatively more insight into the edge information and texture content of the image. We consider the histogram values of horizontal scale level of 0, 1 and 2. Due to page limitations, histogram plots of other images are not provided.

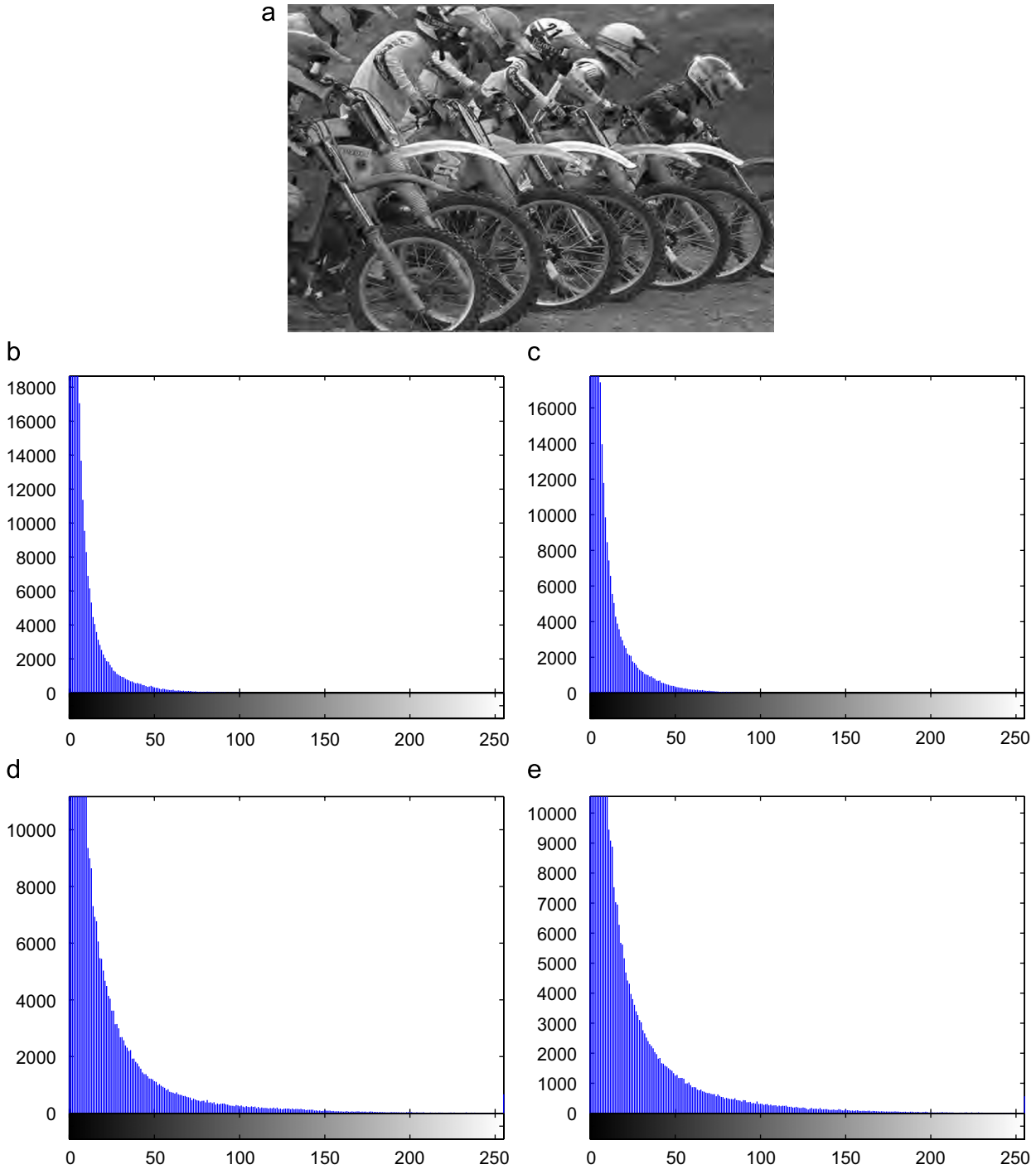

Fig. 14. Histogram plots of the (a) coded image; without filter (b) horizontal, (c) vertical, and with filter (d) horizontal, (e) vertical.

References

[1] Z. Wang, A.C. Bovik, A universal image quality index, IEEE Signal Process. Lett. 9 (March 2002) 81–84.

[2] A.M. Eskicioglu, P.S. Fisher, Image quality measures and their performance, IEEE Trans. Commun. 43 (December 1995) 2959–2965.

[3] B. Girod, What's wrong with mean-square error, in: A.B. Watson (Ed.), Digital Image and Human Vision, MIT Press, Cambridge, MA, 1993, pp. 207–220.

[4] Z. Wang, Rate scalable foveated image and video communications, Ph.D. Thesis, Department of ECE, The University of Texas at Austin, December 2003.

[5] Z. Wang, A.C. Bovik, L. Lu, Why is image quality assessment so difficult?, in: Proceedings of the IEEE International

Conference on Acoustics, Speech, and Signal Processing, Orland, May 2002.
[6] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: from error visibility to structural similarity, IEEE Trans. Image Process. 13 (4) (April 2004) 600–612.
[7] H.R. Wu, M. Yuen, A generalized block-edge impairment metric for video coding, IEEE Signal Process. Lett. 4 (November 1997) 317–320.
[8] Z. Wang, H.R. Sheikh, A.C. Bovik, No-Reference perceptual quality assessment of JPEG compressed images, in: Proceedings of the IEEE ICIP, September 2002, pp. I-477–I-480.
[9] F. Pan, X. Lin, S. Rahardja, W. Lin, E. Ong, S. Yao, Z. Lu, X. Yang, A locally adaptive algorithm for measuring blocking artifacts in images and videos, Signal Process.: Image Commun. 19 (6) (2004) 499–506.
[10] Z.M. Parvez Sazzad, Y. Horita, Image quality assessment models for JPEG and JPEG2000 compressed color images, in: Proceedings of CGIV, Leeds, UK, June 2006.
[11] Y. Horita, M. Sato, Y. Kawayoke, Z.M. Parvez Sazzad, K. Shibata, Segmentation and local features based image quality evaluation, in: Proceedings of the IEEE ICIP, Atlanta, US, October 2006.
[12] Z. Wang, A.C. Bovik, B.L. Evans, Blind measurement of blocking artifacts in images, in: Proceedings of the IEEE ICIP, vol. 3, September 2000, pp. 981–984.
[13] A.C. Bovik, S. Liu, DCT-domain blind measurement of blocking artifacts in DCT-coded images, in: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, vol. 3, May 2001, pp. 1725–1728.
[14] S.H. Oguz, Y.H. Hu, T.Q. Nguyen, Image coding ringing artifact reduction using morphological post-filtering, in: Proceedings of the IEEE 2nd Workshop Multimedia Signal Processing, 1998, pp. 628–633.
[15] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, A no-reference perceptual blur metric, in: Proceedings of the International Conference on Image Processing, Rochester, NY, vol. 3, 2002, pp. 57–60.
[16] P. Marziliano, F. Dufaux, S. Winkler, T. Ebrahimi, Perceptual blur and ringing metrics: applications to JPEG2000, Signal Process.: Image Commun. 19 (2) (2004) 163–172.
[17] H.R. Sheikh, A.C. Bovik, L. Cormack, No reference quality assessment using natural scene statistics: JPEG2000, IEEE Trans. Image Process. 14 (11) (November 2005) 1918–1927.
[18] H. Tong, et al., No reference quality assessment for JPEG2000 compressed images, in: Proceedings of Image Processing, Singapore, 2004.
[19] R. Barland, A. Saadane, Reference free quality metric for JPEG2000 compressed images, in: Proceedings of ISSPA, Sydney, Australia, 2005.
[20] JasPer version 1.701.0 software: ⟨http://www.ece.uvic.ca/~mdadams/jasper/⟩.
[21] ITU-R Recommendation BT.500-10, Methodology for the subjective assessment of the quality of television pictures.
[22] Y. Horita, Y. Kawayoke, Z.M. Parvez Sazzad, Image quality evaluation database ⟨ftp://guest@mict.eng.u-toyama.ac.jp/⟩.
[23] Z.M. Parvez Sazzad, Y. Kawayoke, Y. Horita, Spatial features based no reference image quality assessment for JPEG2000, in: Proceedings of the IEEE ICIP, Texas, US, September 2007.
[24] A. Ito, T. Murakami, Image quality adjustment control in video encoding, in: PCSJ1986, 1986, pp. 75–76.
[25] J. Kennedy, R. Eberhart, Particle swarm optimization, in: Proceedings of the IEEE ICNN, Perth, Australia, November 1995, pp. 1942–1948.
[26] Y. Horita, M. Miyahara, T. Murai, Estimation improvement in picture quality scale of monochrome still picture, IEICE Trans. j80 (B-I) (June 1997) 505–514.
[27] H.R. Sheikh, Z. Wang, L. Cormack, A.C. Bovik, LIVE image quality assessment database, 2003, Available: ⟨http://live.ece.utexas.edu/research/quality⟩ [online].
[28] M. Pinson, S. Wolf, An objective method for combining multiple subjective data sets, in: Proceedings of SPIE Video Communications and Image Processing, Lugano, Switzerland, July 2003.
[29] VQEG: Final Report from the video quality experts group on the validation of objective models of video quality assessment, FR-TV Phase II, August 2003 ⟨http://www.vqeg.org/⟩.