

IMAGE QUALITY ASSESSMENT USING THE MOTIF SCAN

Z. Cui, D.-O Kim, and R.-H. Park

Department of Electronic Engineering, Sogang University, Seoul, Korea
{Cuizt11694, hyssop, rhpark}@sogang.ac.kr

ABSTRACT

Image quality assessment (IQA) evaluates the quality of an image by computing the difference between the reference and distorted images. This paper proposes an image quality metric (IQM) that uses the motif scan. Since edges often contain much information on an image and the human visual system is highly adapted for extracting structural information from a scene, we propose a new IQM comparing the similarity of 3×3 block motif scans between the reference and distorted images, where the center of the block is located at edge pixel. A 3×3 block is divided into four overlapping 2×2 grids with the center pixel located at four different corners. The modified versions of the proposed IQM using the uniform and Gaussian weights are also proposed. Experiments with LIVE database for five different distortion types of test images show that the proposed metric gives better performance than the conventional IQMs.

KEYWORDS

DMOS, Image Quality Assessment, Logistic Regression, Motif Scan

1. INTRODUCTION

Image quality assessment (IQA) evaluates how good an image is. It can be classified into two types: subjective and objective. The former is better than the latter because the quality of an image or video is eventually assessed according to the human visual perception. However, such subjective quality assessments are troublesome and expensive, thus not suitable for practical use in real applications. Therefore, the latter is desirable as the practical IQA and thus many objective IQA methods have been developed [1–27]. The objective IQA methods are considered better if they are as closer as to the subjective IQA.

The goal of the objective IQA is to evaluate the image quality that is similar to the quality as people perceive. Thus, the mean square error (MSE) and peak signal to noise ratio (PSNR) that use only the intensity difference for assessing the quality cannot effectively reflect the human perception properties and thus cannot assess the image quality coincidentally with the subjective quality like the mean opinion score (MOS). This reason leads to the development of the objective IQA algorithms. Conventional IQA algorithms can be classified into several approaches depending on the usage of the reference image and the kind of the information used for IQA. In this paper, we classify the convention IQA algorithms into three categories as in [1]. Three categories are structural information based [2–14], human perception/visual attention based [15–25], and information theoretical approaches [26, 27].

Motivation of structural information based IQA is that the structural information of an image changes if an image is distorted. The universal quality index (UQI) [2] was presented as a full-reference (FR) IQM using the structural information of an image. The structural similarity (SSIM) [3], a modified version of the UQI, was also developed. However, the SSIM gives poor

performance for badly blurred images. To reduce performance degradation, new methods have been developed [4, 5]. Edge-based SSIM [4] is based on the edge information as the most important image structure information. The gradient-based SSIM has been proposed by considering the fact that the human visual system (HVS) is sensitive to changes of edges [5]. A multi-scale SSIM (MSSSIM) was proposed [6]. An IQA method based on the edge and contrast similarity between the reference and distorted images [7] was proposed by combining the edge similarity with the contrast similarity used in the SSIM. Similarly to the MSSSIM, an IQM based on multi-scale edge representation was proposed [8]. Also, a discrete wavelet transform-based structural similarity for IQA was proposed [9].

Shnayderman *et al.* proposed an image quality metric (IQM) based on singular value decomposition (SVD) [10], which is called the MSVD. The MSVD method used the mean of the differences between SVD values for assessing the image quality. An IQM using LU factorization (MLU) was proposed, where LU factorization was used for representation of the structural information of an image [11]. An IQM based on Harris response (HRQM) was proposed, in which Harris response was computed from the gradient information matrix [12]. In [13], the joint feature-based quality metric was presented, in which image pixels were classified into three structure types to effectively assess the visual quality. Also, feature map based IQM [14] was proposed based on the SSIM. Instead of comparing images directly, the method uses the SSIM in the feature maps (corner, edge, and symmetry maps) between the reference and distorted images.

A human perception based IQM is motivated by the fact that only the image distortions that can be perceived by most people affects the subjective image quality. This means that if the distortion that cannot be perceived by the HVS occurs with regard to the reference image, people may consider that the distorted image is the same as or similar to the reference image. Thus, some IQMs use the just-noticeable-distortion to detect the distortion that human eye can perceive. A sharpness metric based on just-noticeable blurs was proposed to measure blur artifacts [15] and a perceptual IQM for blocking artifacts of joint photographic experts group (JPEG) compressed images was proposed [16]. A distortion measure of the effect of frequency distortion, and a noise quality measure of the effect of additive noise were developed [17]. DCTune is designed in the context of a discrete cosine transform (DCT)-based model of visual quality [18]. Also, ITU recommended four annexes for objective perceptual IQA [19]. A wavelet based visual signal-to-noise ratio (VSNR) for natural images was also developed based on near-threshold and suprathreshold properties of the human vision [20].

Another approach related to the human perception is based on the phase [21–23]. Phase-based IQMs were motivated from the fact that if an image has some structural distortions, structural distortions lead to the consistent phase change. For IQA, the phase congruency was used [21], which is a measure of feature significance in images, a method of edge detection that is particularly robust against changes in illumination and contrast. Zhai *et al.* proposed the log Gabor phase similarity [22], which is a FR IQM based on measuring of similarities between phases in log Gabor transform domain. Also, similarity of phase spectrum was used for evaluating the image quality [23].

Also, visual attention based IQMs were investigated [24, 25]. Most existing IQMs do not take the human attention analysis into account. Attention to particular objects or regions is an important fact of human vision and perception system in measuring perceived image and video qualities. Feng *et al.* presented a saliency-based objective quality assessment metric [24], for assessing the perceptual quality of decoded video sequences affected by packet loss. Also, an approach for extracting visual attention regions based on a combination of a bottom-up saliency model and semantic image analysis was proposed [25].

An information theoretic approach was proposed, which quantifies visual fidelity by means of an information fidelity criterion (IFC) derived based on natural scene statistics [26]. The visual information fidelity (VIF) was also presented [27].

In this paper, we propose a new structural information based IQM, which uses the 3x3 block motif scan of edges in an image. In the proposed metric, edges of the reference image is extracted and then, processed by the 3x3 block motif scanning. The 3x3 block can be divided into four overlapping 2x2 grids where the edge pixel at the center of the 3x3 block is located at the different corners of four grids. The motifs can reflect the structural information of the scanned images. When an image is degraded by noise, compression, and so on, motifs of the image are changed. Therefore, the difference between four motifs of each edge pixel between the reference and distorted images can be used to evaluate the amount of distortion. The difference between the reference and distorted images is computed by comparing whether the motifs of four grids in each 3x3 block are the same or not.

The rest of the paper is structured as follows. In Section 2, we describe the proposed IQM. First, we briefly review on the motifs which are the base of the proposed IQM. Then, the 3x3 block motif scan is described. Next, the proposed IQM and its modified versions are presented. In Section 3, experimental results are shown with discussions. Finally, in Section 4 conclusions and future works are given.

2. PROPOSED IQM USING THE MOTIF SCAN

In this section, we introduce the motif scan which is the main idea of the proposed IQM. Then, the motif scan based quality metric (MSQM) is proposed and its modified versions are also presented.

2.1. Review on Motif Scan

Jhanwar *et al.* proposed an image retrieval method using a 2x2 local feature, which is a motif scan [28]. A motif is a kind of pattern obtained from tracing intensities of pixels in a block and is used as a descriptor of that block. Motifs were used in image retrieval [29–32], since if the images are different, the motifs are also different. Also, if an image is degraded, motifs of that image also change with respect to those of the reference image. Thus, this is a motivation of the proposed IQM, to which a modified motif scan is applied.

In general, 24 different motifs could traverse a 2x2 grid, however, among which in this paper we consider only six motifs that start from the top left corner of the grid. Also an additional motif is defined when the four intensity values of a 2x2 grid are the same. An image is divided using 2x2 grid units and each grid is replaced by a motif. Fig. 1 illustrates seven motifs defined in a 2x2 grid. The relevant motifs minimize the local intensity variation along the scan line.

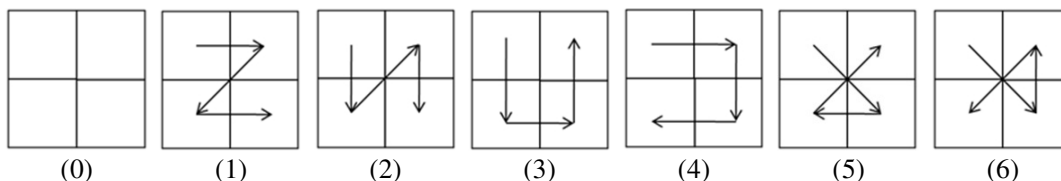


Figure 1. Seven motifs used to traverse a 2x2 grid

Given a 2 x2 grid $\begin{matrix} p_1 & p_2 \\ p_3 & p_4 \end{matrix}$, then δ_q ($1 \leq q \leq 6$) are defined as the sums of absolute differences of intensity values along the q th scan path:

$$\begin{aligned}
 \delta_1 &= |p_1 - p_2| + |p_2 - p_3| + |p_3 - p_4|, \\
 \delta_2 &= |p_1 - p_3| + |p_3 - p_2| + |p_2 - p_4|, \\
 \delta_3 &= |p_1 - p_3| + |p_3 - p_4| + |p_4 - p_2|, \\
 \delta_4 &= |p_1 - p_2| + |p_2 - p_4| + |p_4 - p_3|, \\
 \delta_5 &= |p_1 - p_4| + |p_4 - p_3| + |p_3 - p_2|, \\
 \delta_6 &= |p_1 - p_4| + |p_4 - p_2| + |p_2 - p_3|.
 \end{aligned} \tag{1}$$

Then, the motif of a grid is defined as

$$m = \begin{cases} 0, & \text{if } \min(\delta_q) = 0 \\ \arg \min(\delta_q), & \text{otherwise} \end{cases} \tag{2}$$

where $0 \leq m \leq 6$ represents the index of the seven motifs. In fact, motifs are six types as defined in [29–32]. However, in flat region, a motif cannot be defined. Therefore, we add a motif for representing flat region, resulting in seven motifs in total. In eq. (2), if $\min(\delta_q) = 0$, $\arg \min(\delta_q)$ is not equal to 0 and m is equal to q that gives the minimum of δ_q , $1 \leq q \leq 6$. In other words, if there is at least one δ_q that is equal to zero, it means that the region is flat and m becomes zero. Otherwise, m can have a value (from 1 to 6) among six values, which gives the minimum of δ_q .

Fig. 2 shows an example of the motifs. With an 8×8 image shown in Fig. 2(a), it can be transformed into a corresponding 4×4 image in Fig. 2(b) and also can be represented as Fig. 2(c) using the index defined in Fig. 1. When scanned along the line as in Fig. 2(b), the value of the intensity variation is the minimum among the six motifs.

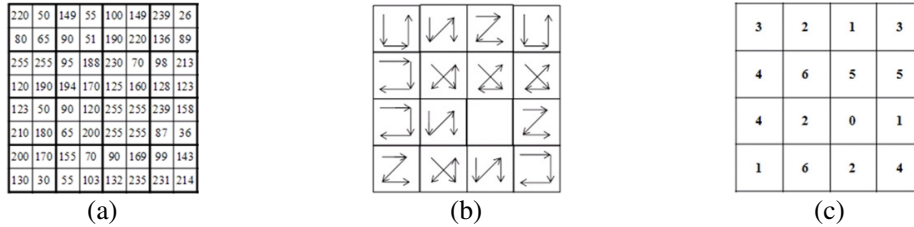


Figure 2. An example of the motifs: (a) An 8×8 image; (b) Motif transformed image of (a); (c) Motifs represented by the index

2.2. 3×3 Block Motif Scan

In this paper, we use a 3×3 block motif scan which was used in image retrieval [31, 32]. In the 3×3 block motif scan, each edge pixel located at the center of the block represents the structural information of an image. Fig. 3 shows an example of the 3×3 block motif scan. Fig. 3(a) represents a 3×3 block that can be divided into four overlapping 2×2 blocks, as shown in Fig. 3(b). The center pixel (i, j) is located at different corners of the four grids as shown in Figs. 3(c)–3(f). In a 2×2 grid, the motif considers only the relationship among the four adjacent pixels, however, in a 3×3 block, it considers the relationship among eight neighboring pixels around the center pixel, where the center pixel is used for constructing four motifs with overlapping 2×2 blocks. Thus, the pixel located at the center produces a set of four adjacent motifs that can

represent the relationship among the neighboring pixels in the 3×3 block and these four motifs in a 3×3 block are used to construct a feature vector.

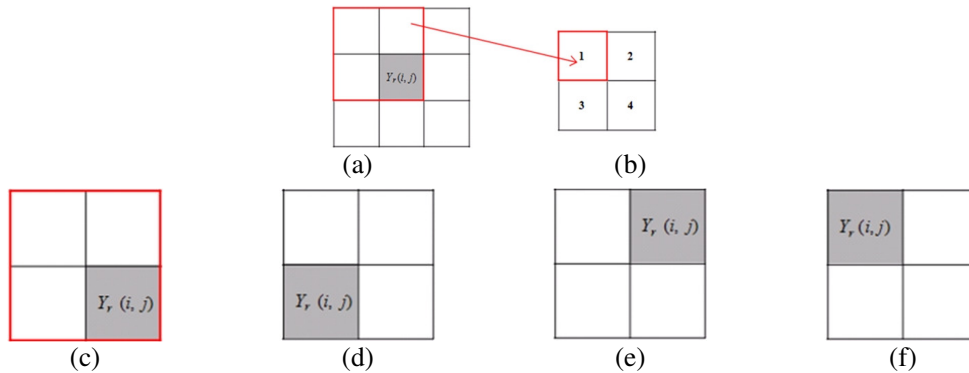


Figure 3. 3×3 block divided into four overlapping 2×2 grids: (a) 3×3 block with center pixel at (i, j) ; (b) Four overlapping blocks with (i, j) centered in the 3×3 block; (c) 2×2 grid of block 1; (d) 2×2 grid of block 2; (e) 2×2 grid of block 3; (f) 2×2 grid of block 4

2.3. Proposed MSQM Using the 3×3 Block Motif Scan

In general, if an image is distorted, the image structure changes. For example, both magnitude and directions of gradients change. Then, scan direction, which defines the motif, will also change. So, motifs can be used for detecting the difference between the reference and distorted images. It is the motivation and the fundamental concept of the proposed MSQM for evaluating the image quality. Also, motif scan was successfully used to retrieve perceptually similar images [31, 32], which means that motifs may be able to detect the perceived difference between two images.

In general, the HVS is sensitive to the changes around edges. Thus, it is possible to assess the image quality although only edge pixels are searched. In practice, many distortions might not be related to edges. In the case of false edges, it is obvious that the image is severely degraded, in which it is not necessary to search all the pixels of the image for assessing the quality. Now, consider some artifacts such as ringing and noise artifacts. Ringing artifacts occur around edges. In IQA, we use dissimilarity values at edge pixels only, however, we also use a 3×3 block for computing the motifs with Gaussian weights. By this process, ringing artifacts can be measured. In the case of noise, noise is widely distributed in the image but not in a particular region. Thus, as in the case of ringing artifacts, noise can also be measured by using only edge pixels. Practically, edge PSNR (EPSNR) [19] used only edge pixels for the video quality assessment. Thus, in this paper, instead of using the whole pixels of an image, only edge pixels of an image are used. To consider the neighboring pixels around edges, a 3×3 block containing four overlapping 2×2 grids is used.

Fig. 4 shows the block diagram of the proposed IQM. First, edges of the reference image are extracted. Then, the 3×3 block motif scan is constructed at each edge pixel in the reference image. In the distorted image, the 3×3 block motif scan is constructed at the same position as the edge position detected in the reference image. The difference between the reference and distorted images is computed by comparing whether the motifs of four grids in each 3×3 block, defined at edge pixel, are the same or not.

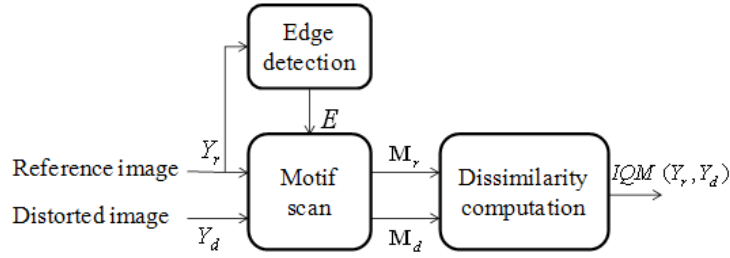


Figure 4. Block diagram of the proposed IQM

For detecting edges, various edge detectors can be used and according to the edge detector used, the performance of the proposed IQM might not be guaranteed. In this paper, the horizontal and vertical Sobel masks are used to obtain the edge information because of their efficiency and simplicity. The edge can be easily detected by

$$E(i, j) = \begin{cases} 1, & |G_x(i, j)| + |G_y(i, j)| > T \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where horizontal and vertical gradients $G_x(i, j)$ and $G_y(i, j)$ are computed by, respectively

$$G_x(i, j) = Y_r(i+1, j-1) + 2Y_r(i+1, j) + Y_r(i+1, j+1) - (Y_r(i-1, j-1) + 2Y_r(i-1, j) + Y_r(i-1, j+1)) \quad (4)$$

$$G_y(i, j) = Y_r(i-1, j+1) + 2Y_r(i, j+1) + Y_r(i+1, j+1) - (Y_r(i-1, j-1) + 2Y_r(i, j-1) + Y_r(i+1, j-1)) \quad (5)$$

with T signifying the threshold and Y_r denoting the intensity of the reference image. Then, the 3×3 block motif scan is obtained at edge pixel of an image. In using the 3×3 grid for scanning the motifs, structural variations of the neighboring pixels at the center pixel are considered by comparing the four motifs. Of course, symmetrical 5×5 and 7×7 grids can be also considered. However, as the grid size increases, the computational complexity increases exponentially. For example, in the case that a 5×5 or 7×7 grids is used for motif scanning, the number of scanning is equal to '16' or '36', respectively. In general, the number of scanning is $(N-1)^2$ for an $N \times N$ grid. Thus, in this paper, a 3×3 grid for scanning the motifs is selected.

The feature vectors \mathbf{M}_r and \mathbf{M}_d at edge pixel (i, j) can be defined as $\mathbf{M}_r(i, j) = [m_r^1, m_r^2, m_r^3, m_r^4]^t$ and $\mathbf{M}_d(i, j) = [m_d^1, m_d^2, m_d^3, m_d^4]^t$, respectively, where subscripts r and d denote the reference and distorted images, respectively, t represents the transpose, and components m_r^u ($1 \leq u \leq 4$) of the column vector \mathbf{M}_r signify the four motifs in a 3×3 block whereas the components m_d^u of the column vector \mathbf{M}_d denote those of the distorted image. Each feature vector includes four motifs defined in a 3×3 block centered at edge pixel (i, j) . Then, the dissimilarity between motifs of the reference and distorted images at edge pixel is defined as

$$D(i, j) = \begin{cases} \frac{1}{4} \sum_{u=1}^4 s^u, & \text{if } E(i, j) = 1 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

with

$$s^u = \begin{cases} 0, & \text{if } m_r^u = m_d^u \\ 1, & \text{otherwise} \end{cases} \quad (7)$$

denoting the dissimilarity between the motif vectors \mathbf{M}_r and \mathbf{M}_d . eq. (7) is defined in [31], in which motifs are used as weights representing whether or not two blocks are the same. The quality measure between the reference image Y_r and distorted image Y_d is defined as

$$MSQM(Y_r, Y_d) = \frac{1}{N} \sum_{i=1}^{w_d} \sum_{j=1}^{h_d} D(i, j) \quad (8)$$

where N denotes the total number of edge pixels of the reference image. W and H signify the width and height of the image, respectively, with both quantities assumed to be even numbers. If major contents reflecting the image quality are found below the threshold in some images, that is, those pixels are not non-edge pixels, our proposed MSQM might not be able to work well. In the case where a white image corrupted with some noise, the original image has no edges. Therefore, our proposed MSQM value is equal to zero. It means that the distorted image is equal to the reference one. Thus, in this paper, we assume that the degradations are widely spread in an image.

In this paper, a logistic regression [33] is performed to describe the relationship between the differential MOS (DMOS) and an IQM. The logistic regression can be written as

$$DMOS = \beta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp(\beta_2(IQM(Y_r, Y_d) - \beta_3))} \right) + \beta_4 \times IQM(Y_r, Y_d) + \beta_5 \quad (9)$$

where the parameters $\beta_1, \beta_2, \beta_3, \beta_4,$ and β_5 are obtained by logistic regression between the quality measure $IQM(Y_r, Y_d)$ and the DMOS [33]. For testing the performance of an IQM obtained after fitting by the logistic regression method, the fitted versions of the IQM are compared with the DMOS, the subjective IQA method, in terms of the performance measure such as the Pearson correlation coefficient (CC), root mean square error (RMSE), and Spearman rank order CC (SROCC), which were recommended by video quality expert group (VQEG) [33] and used in [34].

2.4. Modified MSQM

The proposed IQM described in Sec. 2.3 uses the intensity values at edge pixels of the reference and distorted images, where a 3×3 block motif scan is processed pixel-by-pixel over the edge pixels. One problem with this method is that the 3×3 block motif can only reflect the local feature in a 3×3 block whereas it cannot consider more pixels outside the 3×3 block. In order to solve this problem, we use more information of the neighboring pixels. Thus, instead of the intensity value at pixel, we use the weighted sum of pixel values computed in a prespecified window, for example, in a 5×5 window, which can be written as

$$Y'_r(i, j) = \sum_{k=-2}^2 \sum_{l=-2}^2 Y_r(i+k, j+l) w(k, l) \quad (10)$$

with

$$\sum_{k=-2}^2 \sum_{l=-2}^2 w(k, l) = 1 \quad (11)$$

where $w(k, l)$ denotes the normalized weighting function. In the UQI [2], the local statistics and the UQI were computed in an 8×8 window using the uniform weight. This modified MSQM using the weighted intensity values with a uniform weighting function gives better performance than the MSQM using the original pixel values (with no weight) of an image.

The UQI map often exhibits undesirable “blocking” artifacts, thus to solve the problem, the modified version of this method, the SSIM, was proposed [3] although there are some differences between the SSIM and UQI except for the Gaussian weights. The SSIM uses the Gaussian weight for computing the local statistics and the SSIM index map, which makes the performance of the SSIM better than that of the UQI. Similarly, we use the modified MSQM with Gaussian weight. We get better performance with the Gaussian weight than with the uniform weight, as expected.

The MSQM using the uniform weighing function gives the same importance to all the neighboring pixels around the edge pixel in a 3×3 window. However, the pixels that are far away from the center pixel of the window give little effect on the center pixel. The large window gives blurring and thus degrades the performance of the proposed IQM. Thus, we use a 5×5 circular-symmetric Gaussian weighting function in eq. (10), where $w(k, l)$ denotes the Gaussian weighting function with standard deviation of 0.8 samples. In this paper, we have experimented with three different weighting schemes: MSQM with no weight ($MSQM_N$), MSQM with uniform weight ($MSQM_U$), and MSQM with Gaussian weight ($MSQM_G$).

Fig. 5 shows the reference and three distorted images with their distortion maps. Fig. 5(a) is a reference image and Figs. 5(b), 5(c), and 5(d) are the JPEG 2000 compression images of Fig. 5(a), with the bitrates at 1.3804, 0.3767, and 0.1060 bits/pixel, respectively, and the degree of degradation increases in order of Figs. 5(b), 5(c), and 5(d). The DMOS of Figs. 5(b), 5(c), and 5(d) are 6.9169, 42.4971, and 65.8576, respectively, with the decrease of the bitrates. Figs. 5(e)-(f) represent the distortion map between Figs. 5(a) and 5(b)-(d), respectively, with the range between 0 and 1, at step of 0.25 (see eq. (6)). The red color denotes the value of 1. The more red color pixels the distortion map has, the more the image is degraded. Fig. 5(e) shows the distortion map between Figs. 5(a) and 5(b), with $MSQM_N=11.3345$, $MSQM_U=10.9143$, and $MSQM_G=9.9229$. Fig. 5(f) shows the distortion map between Figs. 5(a) and 5(c), with $MSQM_N=36.5426$, $MSQM_U=37.0516$, and $MSQM_G=39.1152$. Fig. 5(g) shows the distortion map between Figs. 5(a) and 5(d) with $MSQM_N=49.9891$, $MSQM_U=52.9178$, and $MSQM_G=54.4615$. It can be observed that the larger the degree of the degradation becomes, the more red pixels the distortion map has, which represents that degradation of the distorted images becomes larger. For example, Fig. 5(g) has more red pixels than Figs. 5(e) and 5(f) in edge regions of the sail, thus we can say that Fig. 5(d) is more degraded than Figs. 5(b) and 5(c). It can also be observed that the proposed MSQMs can represent the degree of degradation well when compared with the DMOS value.

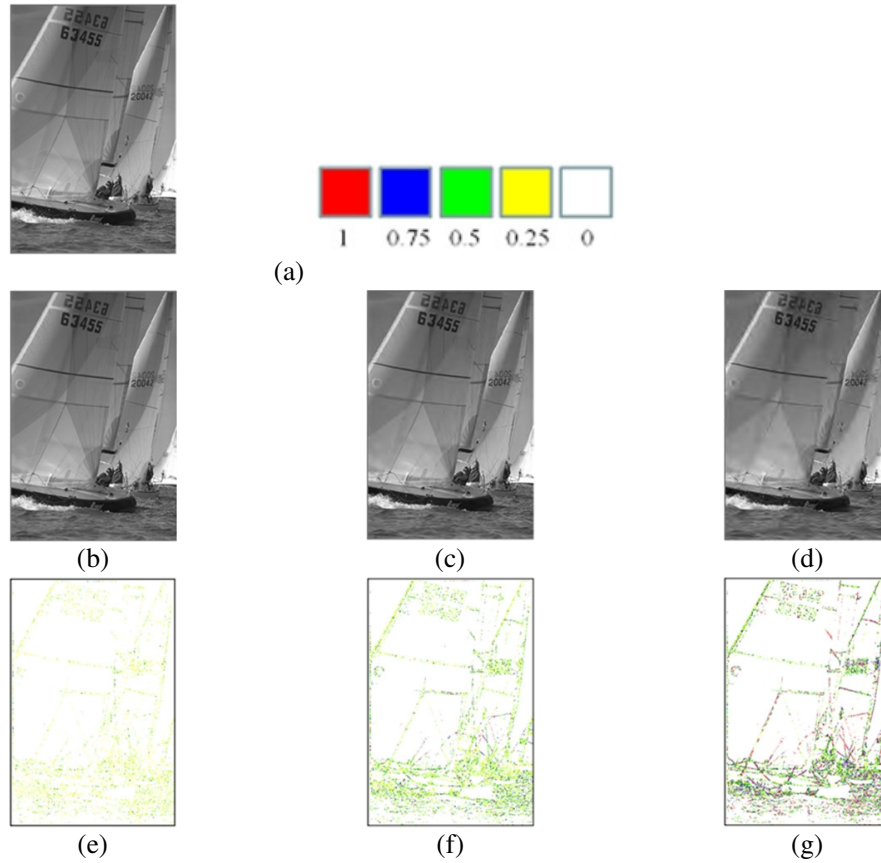


Figure 5. Reference and distorted images with their distortion maps: (a) Reference image; (b) JPEG 2000 image (compressed at 1.3804 bits/pixel, DMOS=6.9169); (c) JPEG 2000 image (compressed at 0.3767 bits/pixel, DMOS=42.4971); (d) JPEG 2000 image (compressed at 0.1060 bits/pixel, DMOS=65.8576); (e) Distortion map between (a) and (b) ($MSQM_N=11.3345$, $MSQM_U=10.9143$, $MSQM_G=9.9229$); (f) Distortion map between (a) and (c) ($MSQM_N=36.5426$, $MSQM_U=37.0516$, $MSQM_G=39.1152$); (g) Distortion map between (a) and (d) ($MSQM_N=49.9891$, $MSQM_U=52.9178$, $MSQM_G=54.4615$)

3. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we evaluate the performance of the proposed $MSQM_G$ through performance comparison with conventional IQMs for the LIVE images [34]. For all the images, the edge is detected using the threshold $T=69$ in eq. (3) and the motifs are obtained by using the Gaussian weight of intensity of an image. We experimentally select the threshold value 69 when the Pearson CC between the $MSQM$ and MOS values on the LIVE dataset is highest. Fig. 6 shows the graph of the performance of the proposed $MSQM$ as a function of the threshold T . The normalized weight with a 5×5 circular-symmetric Gaussian weighting function is used, with the standard deviation of 0.8. The conventional IQMs such as the PNSR, MSVD [10], MSSIM [3], IFC [26], VIF [27], MLU [11], and HRQM [12] are used for performance comparison.

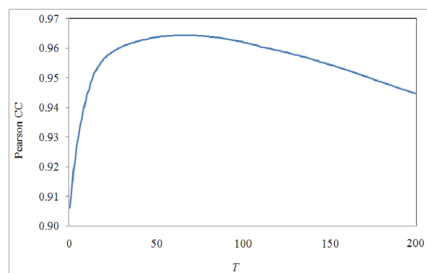


Figure 6. Graph of the performance of the proposed MSQM as a function of the threshold T .

3.1. Images Used in Experiments

In this paper, the LIVE data images [34] are used for performance comparison of the proposed $MSQM_G$ and conventional IQMs. There are total of 982 images consisting of 779 distorted and 203 reference images. The entire images are derived from 29 high-resolution and high-quality color images. The images are various, such as faces, people, animals, natural scenes, and so on. There are five distortion types: JPEG 2000 compression (JPEG 2000), JPEG compression (JPEG), white noise (WN), Gaussian blur (GBLur), and fast fading (FF). These distortions reflect a wide range of image degradations, from smoothing to structured distortion, image dependent distortion, and random noise. The level of distortion was varied to generate images with a wide range of quality, from imperceptible levels to high levels [35].

3.2. Performance Comparison

In order to evaluate the performance of the proposed MSQMs, we compare them with the seven conventional IQMs such as the PSNR, MSVD, MSSIM, IFC, VIF, MLU, and HRQM in terms of the Pearson CC, RMSE, and the Spearman rank order CC (SROCC) which are used as the criteria for the performance evaluation [33]. The performance comparisons of eight IQMs for each of five different distortion types and for the entire images are given.

Table 1 shows the performance of the proposed MSQMs ($MSQM_N$, $MSQM_U$, and $MSQM_G$) in terms of the Pearson CC, RMSE, and SROCC. It can be observed that the proposed $MSQM_G$ gives better performance than the other two MSQMs for entire images (All data) and five distortion types. The $MSQM_U$ gives better performance than $MSQM_N$ whereas the $MSQM_G$ better than $MSQM_U$. Experimental results in Table 1 support the description in Sec. 2.4. Thus, afterward, we regard the proposed $MSQM_G$ as our final proposed IQM and compare the $MSQM_G$ with conventional IQMs. In Tables 1–4, numbers underlined represent the best performance measures considered among IQM methods compared.

Table 1. Performance comparison of the proposed MSQMs (Pearson CC, RMSE, and SROCC).

		JPEG 2000	JPEG	WN	GBLur	FF	All data
Pearson CC	$MSQM_N$	0.9560	0.9681	0.9443	0.9577	0.9545	0.9486
	$MSQM_U$	0.9571	0.9647	0.9734	0.9467	0.9607	0.9584
	$MSQM_G$	<u>0.9630</u>	<u>0.9718</u>	<u>0.9826</u>	<u>0.9676</u>	<u>0.9630</u>	<u>0.9644</u>
RMSE	$MSQM_N$	7.9142	9.1934	9.8963	7.2995	8.7084	8.6509
	$MSQM_U$	7.6134	8.9415	7.2815	6.4857	8.2422	7.8025
	$MSQM_G$	<u>7.5478</u>	<u>8.0140</u>	<u>6.4235</u>	<u>5.6254</u>	<u>7.9936</u>	<u>7.2285</u>
SROCC	$MSQM_N$	0.9496	0.9703	0.9785	0.9502	0.9486	0.9497
	$MSQM_U$	0.9546	0.9673	0.9745	0.9321	0.9580	0.9572
	$MSQM_G$	<u>0.9630</u>	<u>0.9774</u>	<u>0.9821</u>	<u>0.9662</u>	<u>0.9609</u>	<u>0.9637</u>

Table 2 shows the performance comparison of different IQMs in terms of the Pearson CC after logistic regression. The Pearson CC between the IQMs and the DMOS is calculated to evaluate the prediction accuracy. A high value of the Pearson CC signifies that the performance of the IQM is good. As shown in Table 2, for the entire images (All data) the proposed MSQM_G gives better results than the other conventional IQMs, which means that the proposed MSQM_G is most similar to the DMOS (high Pearson CC) among eight IQMs. For the fast fading (FF) distorted images, the MSQM_G gives better performance than the other IQMs. For the other four distortion types except for FF, the VIF gives better performance than the other IQMs. For the JPEG 2000 and JPEG distortion types, the MSSIM gives the performance comparable to the proposed MSQM_G. The proposed MSQM_G gives the performance similar to the VIF in white noise (WN). It gives better performance than the other conventional IQMs except for the VIF.

Table 2. Performance comparison of different IQMs (Pearson CC).

	JPEG 2000	JPEG	WN	GBlur	FF	All data
PSNR	0.8770	0.8675	0.9795	0.7746	0.8762	0.8591
MSVD	0.9580	0.9594	0.9730	0.7801	0.9255	0.8773
MSSIM	0.9657	0.9770	0.9612	0.9388	0.9520	0.9446
IFC	0.9195	0.9436	0.9582	0.9541	0.9542	0.9262
VIF	<u>0.9720</u>	<u>0.9849</u>	<u>0.9868</u>	<u>0.9742</u>	0.9620	0.9604
MLU	0.9205	0.9330	0.8932	0.8301	0.8798	0.8086
HRQM	0.9070	0.9308	0.9393	0.9371	0.9268	0.9051
MSQM _G	0.9630	0.9718	0.9826	0.9676	<u>0.9630</u>	<u>0.9644</u>

Table 3 lists the performance comparison of different IQMs in terms of the RMSE. The RMSE quantifies the amount of the distance between the true values and the estimates. A small value of the RMSE denotes that the performance of the IQM is good. The proposed MSQM_G gives better performance than the conventional IQMs for the entire images (All data), which signifies that the distance between the MSQM_G and DMOS is small. For the three distortion types except for JPEG2000 and WN, the proposed MSQM_G gives the best results, and for JPEG2000 and WN, it gives performance similar to the MSSIM and VIF, respectively. Anyhow, the proposed MSQM_G gives better performance than other conventional IQMs for the entire images and for the different distortion types.

Table 3. Performance comparison of different IQMs (RMSE).

	JPEG 2000	JPEG	WN	GBlur	FF	All data
PSNR	12.7036	18.7445	9.0113	12.7156	13.9818	13.9836
MSVD	8.7476	14.5393	8.5229	19.5291	11.1854	13.1106
MSSIM	<u>7.1817</u>	10.8857	9.9196	7.3768	8.7327	8.9693
IFC	12.3814	11.7226	8.2212	9.1620	8.6057	10.3014
VIF	8.6126	8.0872	<u>6.4199</u>	6.0392	8.2541	7.6110
MLU	10.0915	11.9185	25.2846	15.8980	14.2361	16.0740
HRQM	12.3901	13.8434	10.2022	9.5922	10.8922	11.6163
MSQM _G	7.5478	<u>8.0140</u>	6.4235	<u>5.6254</u>	<u>7.9936</u>	<u>7.2285</u>

Fig. 7 shows the scatter plots of the proposed MSQM_G and the VIF versus the DMOS for JPEG 2000 and Gaussian blur distortions. In case of the Gaussian blur, the Pearson CC of the VIF is larger than that of the proposed MSQM_G, as shown in Table 2, whereas, the RMSE of the proposed MSQM_G is smaller than that of the VIF, as shown in Table 3. The reason is that the

distribution of the points in the proposed $MSQM_G$ is closer to the red line than that of the VIF for the JPEG 2000 compression distortion as shown in Fig. 7(a). The red line signifies the Pearson CC of 1. For the Gaussian blur distortion, the reason is the same as the JPEG 2000 compression distortion case, as shown in Fig. 7(b). It is also true for the JPEG compression and FF cases.

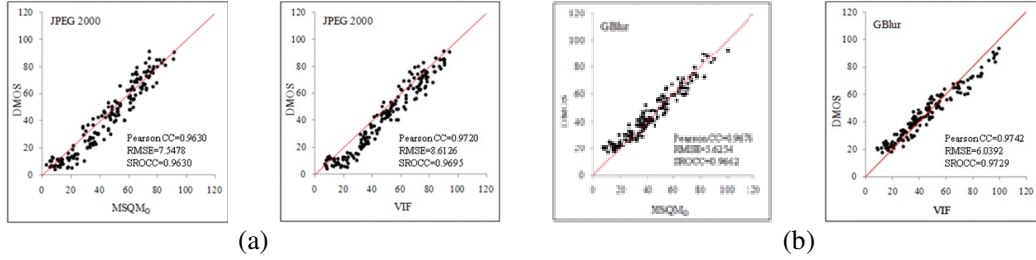


Figure 7. Scatter plots of the proposed $MSQM_G$ and VIF versus DMOS for JPEG 2000 and Gaussian blur distortion types: (a) JPEG 2000; (b) Gaussian blur.

Table 4 lists the performance comparison of different IQMs in terms of the SROCC. The SROCC can reflect the monotonicity of the IQM. A high value of the SROCC signifies that the performance of the IQM is good. It can be observed that the proposed $MSQM_G$ gives the better performance than other IQMs for entire images (All data) and the similar results to the VIF for WN. In WN, the PSNR gives the results comparable to the VIF.

Table 4. Performance comparison of different IQMs (SROCC).

	JPEG 2000	JPEG	WN	GBlur	FF	All data
PSNR	0.8954	0.8804	0.9853	0.7820	0.8906	0.8754
MSVD	0.9574	0.9504	0.9786	0.7792	0.9269	0.8772
MSSIM	0.9600	0.9753	0.9748	0.9474	0.9537	0.9478
IFC	0.9104	0.9439	0.9378	0.9524	0.9599	0.9254
VIF	0.9695	0.9845	0.9859	0.9729	0.9651	0.9636
MLU	0.9290	0.9493	0.9229	0.8164	0.8870	0.8103
HRQM	0.9032	0.9314	0.9287	0.9473	0.9374	0.9056
$MSQM_G$	0.9630	0.9774	0.9821	0.9662	0.9609	0.9637

As shown in Tables 2–4, the proposed $MSQM_G$ gives good performance for the entire images (All data), which signifies that the proposed $MSQM_G$ gives the trend similar to the DMOS (high Pearson CC), the distance between the $MSQM_G$ and DMOS is small, and the rank according to the $MSQM_G$ is similar to that according to the DMOS.

Fig. 8 shows scatter plots of eight IQMs (seven conventional IQMs and the proposed $MSQM_G$) versus the DMOS. As shown in Fig. 7, the proposed $MSQM_G$ shows similar performance to the VIF, where the VIF gives better performance than the other conventional IQMs. It can be observed that the scatter plot of the proposed $MSQM_G$ is clustered more closely to the red line than other IQMs, and that the distance between the line and the points is smaller than those of other IQMs. Thus, the proposed $MSQM_G$ gives better performance for entire images (All data) than other IQMs in terms of the Pearson CC and RMSE as shown in Tables 2 and 3.

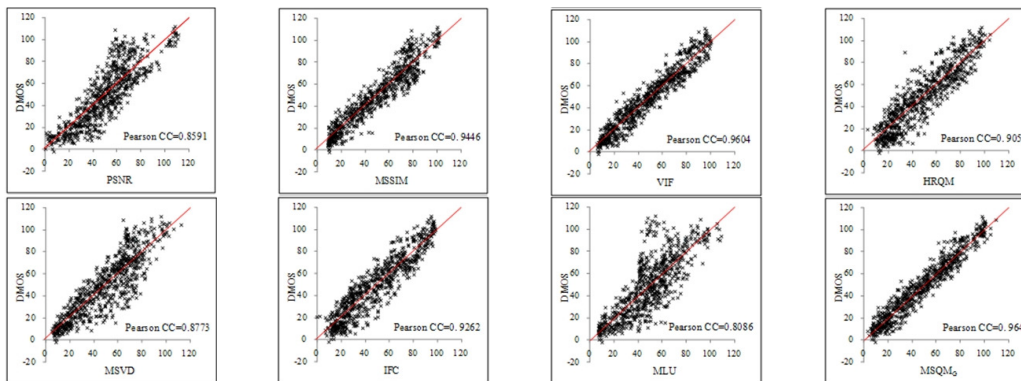


Figure 8. Scatter plots of different IQMs versus DMOS

Fig. 9 shows the scatter plots of five distortion types of the proposed $MSQM_G$ versus the DMOS. It can be observed that the distribution of the points is closer to the red line for the white noise and Gaussian blur distortions than for the other three distortion types. Thus, the RMSE values of the two distortion types are smaller than those of the other three distortion types, especially, for the Gaussian blur distortion.

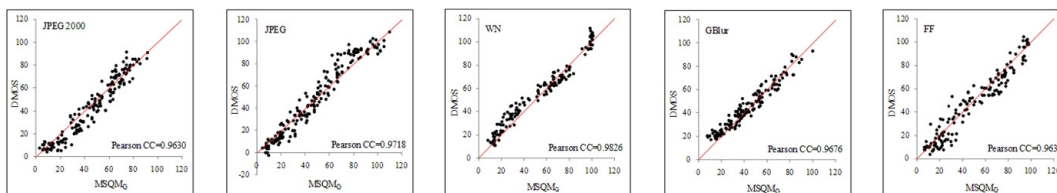


Figure 9. Scatter plots of five distortion types of the proposed $MSQM_G$ versus DMOS

In summary, it can be observed that the performance of the proposed $MSQM_G$ is better than those of conventional IQMs. Figs. 7–9 support the experimental results shown in Tables 2–4.

4. CONCLUSIONS

In this paper, we propose a new structural information based IQM, which uses the 3×3 block motif scan of edge pixels of an image. Instead of using the whole pixels of an image, only edge pixels of an image extracted by Sobel masks are used. Three different types of MSQMs are described and their performance is compared: $MSQM_N$, $MSQM_U$, and $MSQM_G$. Through experiments with the LIVE images, it can be observed that the proposed $MSQM_G$ gives the better performance than other conventional IQMs in terms of the Pearson CC, RMSE, and SROCC. Future work will focus on the extension of the proposed $MSQM_G$ that uses color information.

ACKNOWLEDGEMENTS

This work was supported by the Second Brain Korea 21 Project.

REFERENCES

- [1] Z. Wang, A. C. Bovik, Mean squared error: love it or leave it?– A new look at signal fidelity measures, *IEEE Signal Processing Magazine* 26 (2009) 98–117.
- [2] Z. Wang and A. C. Bovik, A universal image quality index, *IEEE Signal Processing Letters* 9 (2002) 81–84.
- [3] Z. Wang and A. C. Bovik, Image quality assessment: From error visibility to structural similarity, *IEEE Transaction on Image Processing* 13 (2004) 600–612.

- [4] G.-H. Chen, C.-L. Yang, and S.-L. Xie, Edge-based structural similarity for image quality assessment, in: Proceedings of International Conference on Acoustics, Speech, and Signal Processing, Toulouse, France, 2006, pp. 14–19.
- [5] G.-H. Chen, C.-L. Yang, and S.-L. Xie, Gradient-based structural similarity for image quality assessment, in: Proceedings of International Conference on Image Processing, Atlanta, GA, 2006, pp. 2929–2932.
- [6] Z. Wang, E. P. Simoncelli, and A. C. Bovik, Multiscale structural similarity for image quality assessment, in: Proceedings of IEEE Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, 2003, pp. 1398–1402.
- [7] F. Wei, X. Gu, and Y. Wang, Image quality assessment using edge and contrast similarity, in: Proceedings of IEEE International Joint Conference on Neural Networks, Hong Kong, China, 2008, pp. 852–855.
- [8] G. Zhai, W. Zhang, X. Yang, and Y. Xu, Image quality assessment metrics based on multi-scale edge presentation, in: Proceedings of IEEE Workshop Signal Processing System Design and Implementation, Athens, Greece, 2005, pp. 331–336.
- [9] C.-L. Yang, W.-R. Gao, and L.-M. Po, Discrete wavelet transform-based structural similarity for image quality assessment, in: Proceedings of IEEE International Conference on Image Processing, San Diego, CA, 2008, pp. 377–380.
- [10] A. Shnayderman, A. Gusev, and A. M. Eskicioglu, An SVD-based grayscale image quality measure for local and global assessment, IEEE Transaction on Image Processing 15 (2006) 422–429.
- [11] H.-S. Han, D.-O Kim, and R.-H. Park, Structural information-based image quality assessment using LU factorization, IEEE Transaction on Consumer Electronics 55 (2009) 165–171.
- [12] D.-O Kim and R.-H. Park, New image quality metric using the Harris response, IEEE Signal Processing Letters 16 (2009) 616–619.
- [13] D.-O Kim and R.-H. Park, Joint feature-based visual quality assessment, Electronics Letters 43 (2007) 1134–1135.
- [14] L. Cui and A. R. Allen, An image quality metric based on corner, edge and symmetry maps, in: Proceedings of British Machine Vision Conference, Leeds, UK, 2008.
- [15] R. Ferzli and L. J. Karam, A no-reference objective image sharpness metric based on just-noticeable blur and probability summation, in: Proceedings of International Conference on Image Processing, San Antonio, TX, 2007, pp. 445–448.
- [16] Z. Wang, H. R. Sheikh, and A. C. Bovik, No-reference perceptual quality assessment of JPEG compressed images, in: Proceedings of International Conference on Image Processing, Rochester, NY, 2002, pp. 477–480.
- [17] N. Damera-Venkata, T. Kite, W. Geisler, B. Evans, and A. C. Bovik, Image quality assessment based on a degradation model, IEEE Transaction on Image Processing 9 (2000) 636–650.
- [18] A. B. Watson, DCTune: A technique for visual optimization of DCT quantization matrices for individual images, in: Digest of Technical Papers of Society for Information Display, Seattle, WA, 1993, pp. 946–949.
- [19] ITU-R Recommendation J.144, Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference, International Telecommunication Union, 2004.
- [20] D. M. Chandler and S. S. Hemami, VSNR: A wavelet-based visual signal-to-noise ratio for natural images, IEEE Trans. Image Processing 16 (2007) 2284–2298.
- [21] Z. Liu and R. Laganiere, On the use of phase congruency to evaluate image similarity, in: Proceedings of International Conference on Acoustics, Speech, Signal Processing, Toulouse, France, 2006, pp. 937–940.
- [22] G. Zhai, W. Zhang, Y. Xu, and W. Lin, LGPS: Phase based image quality assessment metric, in: Proceedings of IEEE Workshop Signal Processing Systems, Shanghai, China, 2007, pp. 605–609.
- [23] P. Skurowski and A. Gruca, Image quality assessment using phase spectrum correlation, Lecture Notes in Computer Science, Computer Vision and Graphics, Eds. G. Goos *et al.*, Springer-Verlag Berlin Heidelberg, (2008) 80–89.
- [24] X. Feng, T. Liu, D. Yang, and Y. Wang, Saliency based objective quality assessment of decoded video affected by packet losses, in: Proceedings of International Conference on Image Processing, San Diego, CA, 2008, pp. 2560–2563.
- [25] Z. You, A. Perkis, M. M. Hannuksela, and M. Gabbouj, Perceptual quality assessment based on visual attention analysis, in: Proceedings of ACM International Conference on Multimedia, Beijing, China, 2009, pp. 561–564.

- [26] H. R. Sheikh, A. C. Bovik, and G. de Veciana, An information fidelity criterion for image quality assessment using natural scene statistics, *IEEE Transaction on Image Processing* 14 (2005) 2117–2128.
- [27] H. R. Sheikh and A. C. Bovik, Image information and visual quality, *IEEE Transaction on Image Processing* 15 (2006) 430–444.
- [28] N. Jhanwar, S. Chaudhuri, G. Seetharaman, and B. Zavidovique, Content based image retrieval using motif cooccurrence matrix, *Image and Vision Computing* 22 (2004) 1211–1220.
- [29] T.-X. Lin and C.-S. Hung, Quadrant motif approach for image retrieval, in: *Proceedings of the 14th International Conference on Central Europe on Computer Graphics, Visualization and Computer Vision*, Plzen, Czech Republic, 2006, pp. 209–215.
- [30] A. Hafiane, S. Chaudhuri, G. Seetharaman, and B. Zavidovique, Region-based CBIR in GIS with local space filling curves to spatial representation, *Pattern Recognition Letters* 27 (2006) 259–267.
- [31] C.-H. Lin, R.-T. Chen, and Y.-K. Chan, A smart content-based image retrieval system based on color and texture feature, *Image and Vision Computing* 22 (2008) 1211–1220.
- [32] Y.-F. Chen, Y.-K. Chan, G.-U. Chang, M.-C. Tsao, Y.-J. Syu, and C.-H. Lin, Image retrieval using modified color variation co-occurrence matrix, in: *Proceedings of 21th International Conference on Industrial, Engineering and Other Application of Applied Intelligent System*, Wroclaw, Poland, 2008, pp. 42–51.
- [33] VQEG, Draft final report from the video quality experts group on the validation of objective models of multimedia quality assessment, phase I, <http://www.vqeg.org/>, 2008.
- [34] 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>
- [35] H. R. Sheikh, M. F. Sabir, and A. C. Bovik, A statistical evaluation of recent full reference quality assessment algorithms, *IEEE Transaction on Image Processing* 15 (2006) 3440–3451.

AUTHORS

Z. Cui received the M.S. degree in electronic engineering from Sogang University in 2009. Her current research interests are image quality assessment and image processing.

Dong-O Kim received the B.S. and M.S. degrees in electronic engineering from Sogang University, Seoul, Korea, in 1999 and 2001, respectively. Currently, he is working toward the Ph.D. degree in electronic engineering at Sogang University. His current research interests are image quality assessment and physics-based computer vision for computer graphics.



Rae-Hong Park was born in Seoul, Korea, in 1954. He received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Seoul, Korea, in 1976 and 1979, respectively, and the M.S. and Ph.D. degrees in electrical engineering from Stanford University, Stanford, CA, in 1981 and 1984, respectively. In 1984, he joined the faculty of the Department of Electronic Engineering, School of Engineering, Sogang University, Seoul, Korea, where he is currently a Professor. In 1990, he spent his sabbatical year as a Visiting Associate Professor with the Computer Vision Laboratory, Center for Automation Research, University of Maryland at College Park. In 2001 and 2004, he spent sabbatical semesters at Digital Media Research and Development Center, Samsung Electronics Co., Ltd. (DTV image/video enhancement). His current research interests are computer vision, pattern recognition, and video communication. He served as Editor for the Korea Institute of Telematics and Electronics (KITE) *Journal of Electronics Engineering* from 1995 to 1996.



Dr. Park was the recipient of a 1990 Post-Doctoral Fellowship presented by the Korea Science and Engineering Foundation (KOSEF), the 1987 Academic Award presented by the KITE, and the 2000 Haedong Paper Award presented by the Institute of Electronics Engineers of Korea (IEEK), the 1997 First Sogang Academic Award, and the 1999 Professor Achievement Excellence Award presented by Sogang University.