

VIDEO QUALITY ASSESSMENT USING LAPLACIAN MODELING OF MOTION VECTOR DISTRIBUTION IN MOBILE VIDEO STREAM

Ho-Sung Han, Rae-Hong Park, Dong-O Kim, and Chang-Dae Jung

Department of Electronic Engineering, Sogang University, Seoul, Korea
pencils@sogang.ac.kr, rhpark@sogang.ac.kr, hyssop@sogang.ac.kr,
jcdlove@sogang.ac.kr

ABSTRACT

Video/Image quality assessment (VQA/IQA) is fundamental in various fields of video/image processing. VQA reflects the quality of a video as most people commonly perceive. This paper proposes a reduced-reference mobile VQA, in which one-dimensional (1-D) motion vector (MV) distributions are used as features of videos. This paper focuses on reduction of data size using Laplacian modeling of MV distributions because network resource is restricted in the case of mobile video. The proposed method is more efficient than the conventional methods in view of the computation time, because the proposed quality metric decodes MVs directly from video stream in the parsing process rather than reconstructing the distorted video at a receiver. Moreover, in view of data size, the proposed method is efficient because a sender transmits only 28 parameters. We adopt the Laplacian distribution for modeling 1-D MV histograms. 1-D MV histograms accumulated over the whole video sequences are used, which is different from the conventional methods that assess each image frame independently. For testing the similarity between MV histogram of reference and distorted videos and for minimizing the fitting error in Laplacian modeling process, we use the chi-square method. To show the effectiveness of our proposed method, we compare the proposed method with the conventional methods with coded video clips, which are coded under varying bit rate, image size, and frame rate by H.263 and H.264/AVC. Experimental results show that the proposed method gives the performance comparable with the conventional methods, especially, the proposed method requires much lower transmission data.

KEYWORDS

H.263, H.264/AVC, mobile video, motion vectors, reduced-reference, video quality assessment

1. INTRODUCTION

Quality metrics have been studied to assess the image and video quality. Subjective quality assessment (QA) is effective since it reflects closely the human perception. However, subjective QA is inefficient in views of the computation time and cost, because it requires various experiments of real-time QA for a large number of images/videos. Thus, instead of subjective QA, objective QA is preferred. Objective QA shows a good performance when its result is similar to that of subjective QA. Objective QA is also needed to meet various requirements such as the computation time and cost.

Development of objective image/video QA (IQA/VQA) coinciding with the subjective image quality is important because people are the ultimate evaluators [1]. Thus, IQA/VQA can be applied to image/video processing, reproduction, and compression systems [2]. In other words, in some restricted applications, IQA/VQA might balance the tradeoff between desired quality and limited resources. For example, QA can be applied to video compression systems such as H.263,

H.264/AVC, and moving picture experts group (MPEG). Generally, the peak to signal to noise ratio (PSNR) or mean squared error (MSE) is used to control the compression ratio, which does not reflect the human visual system (HVS). Objective QA is preferred to effectively consider the HVS, which will enhance the compression efficiency. Practically, video coding system was developed in [3], where a rate control technique was proposed by using a video quality metric.

Moreover, QA can be used to measure the quality of service (QoS) for service provider. For example, service channel provider can estimate channel condition such as the amount of additive channel noise and data loss by measuring quality of contents [4]. Although the same video (or image) is transmitted, users are differently perceived by channel distortions. Thus, service channel provider can charge a rate according to service quality.

For these reasons, many image quality metrics have been developed [5]. Video quality experts group (VQEG) of International Telecommunication Union (ITU) is working for standardization [6, 7]. VQA methods are classified into three types depending on the availability of reference video (or image): full-reference (FR) [8–13] that uses all the information of a reference video, reduced-reference (RR) [14–17] that utilizes features of a reference video, and no-reference (NR) [18–20] that does not employ any information of a reference video. The availability of the reference video is related to transmission data size. From this fact, it is easy to see that transmission data size is an important issue in VQA. Practically, in some applications such as mobile devices and Internet protocol television (IPTV), transmission data size is very important because their network provides limited resources. These applications utilize data compression (H.263, H.264/AVC, and MPEG) to reduce transmission data size. Therefore, the methodology using data compression is of great advantage to VQA in the previously mentioned applications.

This paper especially focuses on RR VQA for mobile videos. Recently, mobile contents are widely used in video applications such as digital multimedia broadcasting (DMB) and video streaming by portable devices. Practically, digital mobile video should be compressed because of the limitation of network bandwidth in mobile network. So, video compression techniques (H.263 and H.264/AVC) are used for mobile video applications. In [21], Lee and Sim proposed an objective video quality evaluation method for digital mobile videos. They used block, edge, gradient features to consider block artifacts caused by video codec. Generally, conventional RR VQAs extract features from the reference video, and transmits them, which is suitable for video, via an ancillary channel. However, to assess quality of mobile video, some conditions should be considered. For example, the size of RR features is very restricted (for example, the bit rate of encoded common intermediate format (CIF) is between 64 to 512 kbps and VQEG (multimedia test plan) recommends 10 to 64 kbps for the size of RR) and video stream (not video) is transmitted. We consider this circumstance, so we focus on reduction of transmission data size and directly utilize video stream information. Specifically, we use motion vectors (MVs) that are directly obtained from video stream as features of video through the parsing process. To reduce the transmission data size, rather than directly comparing MVs of reference and distorted videos, we compare one-dimensional (1-D) MV distributions, several parameters of which are transmitted to a receiver. In this process, we use a Laplacian modeling of MV distribution to reduce transmission data size without sacrificing quality. Note that MV distributions are computed over whole frames of a video, not on each frame. The proposed method uses MV histograms as feature information computed over all frames of a video, because overall impression is more efficient than impression of each frame for VQA.

The rest of the paper is structured as follows. Section 2 presents an overview of QA. Section 3 presents the proposed VQA method using Laplacian modeling of MV distributions. Experimental

results with discussions are shown Section 4. Finally, Section 5 gives conclusions and comments on future works.

2. OVERVIEW OF QA

Traditionally, the quality of a video (or an image) is estimated by calculating the PSNR or mean squared error (MSE). Thus, the PSNR and MSE are used to calculate the difference between compressed data and reference data in the compression algorithms (video coding and image compression). However, they are not enough to reflect the human visual perception due to development of image processing (such as size of an image, compression technology) camera, storage devices, transmission bandwidth, and so on. According to this development, people need higher quality contents to visual satisfaction. Thus, research of QA is needed.

2.1. Limitation of Conventional QA

The simplest process to calculate the difference between two images is subtraction. For example, the PSNR and MSE are based on subtraction process. The MSE between the reference image I^r and the distorted image I^d is defined as

$$MSE(I^r, I^d) = E \left\{ (I^r - I^d)^2 \right\} \quad (1)$$

where the operation E denotes the expectation. From the MSE, the PSNR is defined as

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (2)$$

where MAX represents the maximum value at a pixel (generally, 255 or 1). From (1) and (2), the PSNR and MSE just calculate the pixel-wise difference value of an image. So, there are limitations to assess the image quality. The limitations can be easily observed from Figure 1, which shows examples of three distorted images that give different visual perception. Figures 1(b)–1(d) are distorted versions of the original image in Figure 1(a), all of which have the similar MSE values (around 57), however each of distorted images is differently perceived by the HVS. Nevertheless they have similar MSE values, each figure shows different quality. This means that the MSE cannot closely reflect the human visual perception. This problem will be more and more important as quality issue. To deal with this problem, there have been many studies on IQA/VQA.



(a)



(b)



Figure 1. Original image (512×768 lighthouse2) and examples of three distorted images that have similar MSE values. (a) Original image, distortion type: (b) Salt & pepper noise (MSE: 57.21), (c) Gaussian noise (MSE: 57.54), (d) JPEG compression (MSE: 57.78).

2.2. Types of Objective QA

FR quality metrics are effective since they use all the information of a reference video. However, they have a limitation on multimedia services through network because they need all of a reference video. The PSNR and MSE are typical FR measures. These methods are computationally simple, so it is easy to implement them in real-time. However, their result is not closely correlated with the HVS. Recently, edge PSNR [17] and structural similarity (SSIM) [10] were proposed. The edge PSNR considers the HVS based on the PSNR, utilizing the fact the HVS is sensitive to change of features such as edges. It was adopted as ITU-Telecommunication (ITU-T) Standardization Sector recommendation J.144: Annex B. Wang *et al.* proposed the SSIM using structure information of images [4]. The SSIM uses the structural difference of images for QA by using the mean, variance, and covariance of reference and distorted images. Also, to enhance the SSIM, a gradient-based structural similarity (GSSIM) [9] was proposed by utilizing the fact that the HVS is sensitive to change of edges. Shnayderman *et al.* proposed the FR quality metric using singular value decomposition (SVD) [11], in which SVD results of each block between reference and distorted images are compared.

NR quality metrics use any information of a reference video. Susstrunk and Winkler proposed a color IQA method by considering compression or transmission distortion [18]. They measured artifacts such as blocking, blurring, and color bleeding, from which they assessed color image quality. Hasler and Susstrunk proposed a color image quality metric using color information of an image [19]. They employed the CIELAB color space to reflect the HVS, in which the mean, standard deviation, and distance information of a color image are used.

RR quality metrics using features of reference images (videos) need less transmission data than FR quality metrics, so they can be practically used as QA measures for multimedia services. Figure 2 illustrates a block diagram of general RR QA. Transmitter sends features of the reference video via ancillary channel. Then, receiver analyzes features of distorted video and received data (features of the reference video) to estimate quality of distorted video. As in Figure 2, the performance of RR quality metrics is dependent on the features extracted from the reference video. Thus, many feature extraction methods have been proposed for IQA or VQA. In [22], an RR quality metric is proposed using edge histogram descriptor (EHD) of MPEG-7 as feature information. Also, Wang and Simoncelli proposed wavelet coefficients as feature information for QA [16]. They reduced the amount of transmission data by comparing computed

wavelet coefficients with the generalized Gaussian density (GGD) model of reference images. If two parameters of the GGD model are given, we can construct the GGD model and compute the Kullback-Leibler divergence (KLD) as a QA metric.

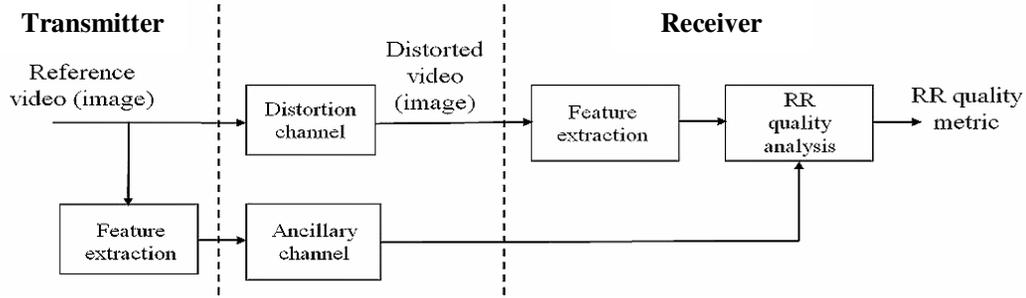


Figure 2. Block diagram of general RR QA.

As previously stated, FR methods are not proper to assess the quality of multimedia contents that are transmitted via network. NR methods have merit in terms of transmission data size, however they cannot guarantee the performance. So RR methods are used for QA of multimedia contents, and transmission data size is an important measure in QA in network services.

This paper focuses on the development of the proposed RR quality metric. The proposed RR quality metrics use MVs as features of reference and distorted videos, noting the fact that MVs of the reference video are changed by distortion or noise. In the proposed algorithm, we compare the 1-D (horizontal and vertical) MV histograms of the reference video with those of the distorted video, instead of directly comparing MVs. Note that MVs represent feature information of videos, and data size is reduced by using two 1-D MV histograms instead of the two-dimensional MV histograms. Practically, in [10], Wang *et al.* proved that MVs are important factors to assess the video quality by showing that combining the SSIM with MVs improves the performance of the VQA algorithm.

2.3. Mean Opinion Score (MOS)

The purposed QA is to estimate the quality score that is close to human perception. In other words, the score data (assessed by human) is needed to evaluate the performance of a QA algorithm. The score data is called MOS. MOS can be calculated to difference MOS (DMOS) by applying double-stimulus continuous quality-scale (DSCQS) [21]. DMOS represents human perception for the difference between reference and distorted videos (images). DMOS is generally used to verify the performance of FR and RR QA algorithms.

2.4. Criteria for Performance Comparison

The performance of a QA metric can be quantitatively defined by calculating the similarity of a QA metric value and the DMOS. In this paper, as similarity criterion, the Pearson correlation coefficient (CC) is adopted, which is recommended by VQEG [4]. The Pearson CC of data **A** and **B** is defined as

$$C(\mathbf{A}, \mathbf{B}) = \frac{(\mathbf{A} - E(\mathbf{A})) \bullet (\mathbf{B} - E(\mathbf{B}))}{\sqrt{\|(\mathbf{A} - E(\mathbf{A}))\|_2} \sqrt{\|(\mathbf{B} - E(\mathbf{B}))\|_2}} \quad (3)$$

where the operation \bullet denotes inner product. Let **A** be the QA metric value and **B** be the DMOS,

then the absolute value of the Pearson CC can be used to represent the performance of a QA metric. That is, higher absolute value of the Pearson CC means that the QA metric is better.

To verify stability of QA metric, we also adopt outlier ratio (OR) (recommended by VQEG). An outlier signifies an observation that lies outside the expected distribution of a particular data set, for example, does not fall within “2×standard deviation” from the DMOS. Thus, a small OR represents that outlier data is rare, that is, OR is a measure of stability of the QA metric. However, if the performance of a QA metric varies depending on the statistical characteristics of video (small OR value), then the QA metric is unstable.

In summary, an efficient QA metric gives a large Pearson CC and small OR. Both measures are used for performance comparison of QA metrics in Section 4.

3. PROPOSED VIDEO QUALITY METRIC

In this paper, we propose two video quality metrics using MV extraction and MV Laplacian modeling. H.264/AVC basically estimates MV of 1/4 subpixel unit, which enhances the compression efficiency. H.264/AVC uses the variable block size for MV estimation, instead of the fixed block size as in conventional standards such as H.263 and MPEG-2. Figure 3 shows variable block size for each mode. Specifically, five internodes are supported in H.264/AVC: SKIP, 16×16, 16×8, 8×16, and 8×8 (known as macroblock (MB) partition). Note that 8×8 mode can be further divided into 8×8, 8×4, 4×8, and 4×4 (known as MB sub-partitions). H.264/AVC gives high compression efficiency using four MB sub-partitions. Parameter t represents block type of each mode.

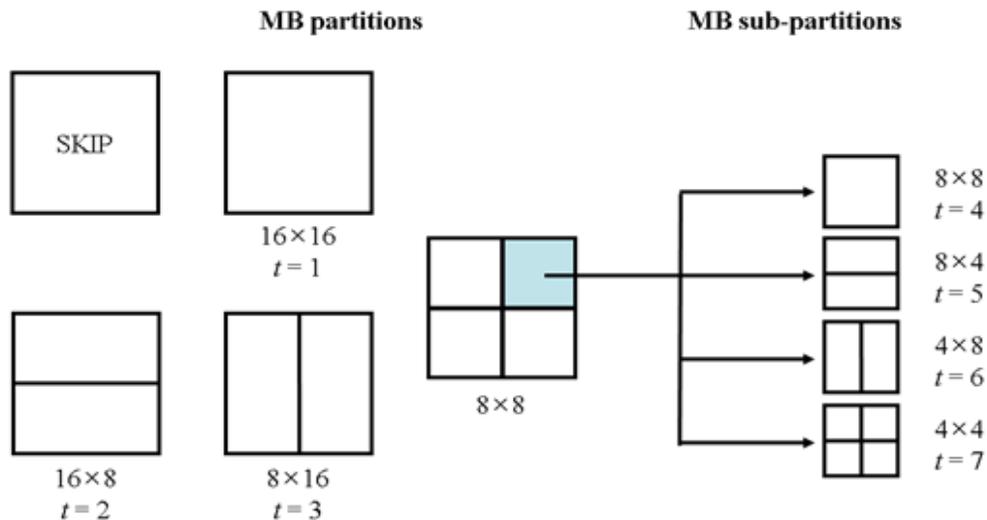


Figure 3. Variable block size in H.264/AVC.

3.1. Laplacian Modeling of MV Distribution

3.1.1. MV Histogram based Quality Metric (MVHQM)

An RR quality metric using 1-D MV histograms of a video was proposed [14]. MVHQM approximates the distribution of MVs as Laplacian distribution, in which parameter in Laplacian

modeling is sent to a receiver for VQA. Generally, an MV extraction process is needed to obtain frame-accumulated 1-D MV histograms. However, in our study, an MV extraction process is not needed because MVs are directly decoded from a parsing process at a receiver. Figure 4 shows a block diagram of the proposed MVHQM. Information to be transmitted is MV distribution of the original video. Since MVs reflect feature information of a video, similarity of MVs between the reference and distorted videos are used to assess the quality of the distorted video. Conventional RR quality metrics usually compare information of each frame and accumulate the quality metric of each frame to assess the video quality. We employ a linear regression with the proposed metric values and the DMOS, and finally get the video quality metric.

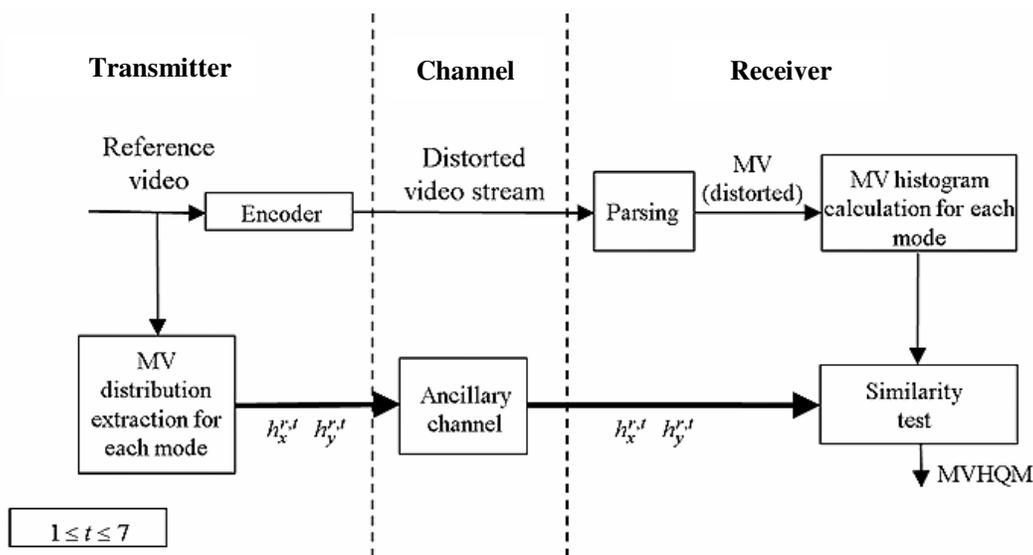


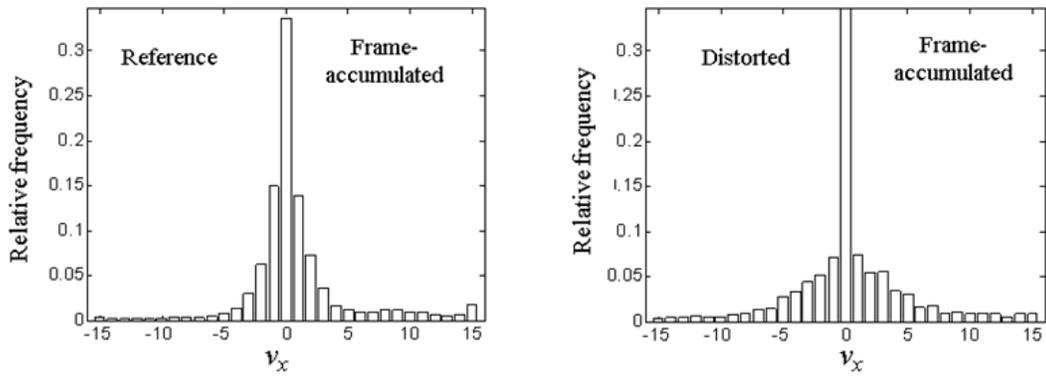
Figure 4. Block diagram of the proposed MVHQM without Laplacian modeling.

MVHQM compares the 1-D (horizontal and vertical) MV histograms of the reference video with those of the distorted video. It uses MVs as features of the reference and distorted videos, noting that MVs of the reference video are changed by distortion or noise. Figure 5 shows examples of MV histograms of the reference and distorted videos by H.263. Since the distorted video (right in Figure 5(a)) is different from the original video (left in Figure 5(a)), MV histograms are different. Left and right columns in Figures 5(b) and 5(c) show MV histograms of the reference and distorted videos, respectively. It is noted that MV histograms are changed when the video contains distortion caused by H.263.

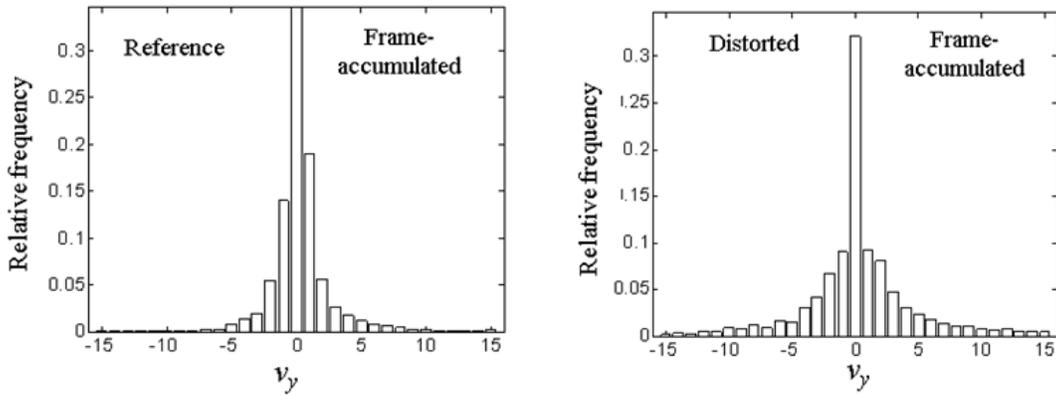
Therefore, the similarity test between MV histograms of the reference and distorted videos can be used as a video quality metric. To compare MV histograms of the reference and distorted videos, two commonly used similarity tests of histograms are considered: the histogram difference and histogram intersection methods. The histogram difference is a simple similarity test that computes the difference between two MV histograms.



(a)



(b)



(c)

Figure 5. Frame-accumulated MV histograms of reference (left) and distorted video (right); (a) 1st frame of reference and distorted videos, (b) Frame-accumulated MV histograms (x direction) of reference video, (c) Frame-accumulated MV histograms (y direction) of distorted video.

The histogram difference, MVH_{diff} is defined as

$$MVH_{diff} = \sum_{v_x} |h_x^r(v_x) - h_x^d(v_x)| + \sum_{v_y} |h_y^r(v_y) - h_y^d(v_y)| \quad (4)$$

where the superscripts r and d represent the reference and distorted videos, respectively. The subscripts x and y denote horizontal and vertical directions of MV projections, respectively. h_x and h_y represent the 1-D horizontal and vertical MV histograms, respectively. Similarly, the histogram intersection, MVH_{int} between two MV histograms is defined as

$$MVH_{int} = \sum_{v_x} \min(h_x^r(v_x), h_x^d(v_x)) + \sum_{v_y} \min(h_y^r(v_y), h_y^d(v_y)) \quad (5)$$

where min denotes the minimum operation that takes the smaller value of the two values. To test the efficiency of the proposed algorithms, the proposed metrics are compared with the MOS. We employ a linear regression with the proposed metrics and the differential MOS (DMOS). The final motion vector histogram based quality metric (MVHQM) is defined as

$$MVHQM_k = a_k \times MVH_k + c_k \quad (6)$$

where the subscript k represents the type (diff or int) of the similarity test of MV histograms used. Constant a and b can be computed by a linear regression between the proposed MVHQM and the DMOS. Two metrics $MVHQM_{int}$ and $MVHQM_{diff}$ give similar results. Thus, two video quality metrics are called the MVHQM.

3.1.2. Laplacian Modeling of MV Distribution

Generally, MV distribution can be well assumed as having a Laplacian distribution. In [22], a Laplacian distribution was used for measuring the motion compensation error, context-adaptive arithmetic coding, or understanding of noise distribution. Therefore, we fit the distribution of MVs with a Laplacian distribution, in which the Laplacian distribution is defined as

$$f(v|\alpha, \beta) = \frac{1}{2\beta} \exp\left(-\frac{|v-\alpha|}{\beta}\right) \quad (7)$$

where α represents the location parameter and β signifies the scale parameter.

To obtain parameters of a Laplacian distribution, we minimize the difference between the MV histogram and the Laplacian distribution, in which we use the chi-square distance [23] as a histogram difference measure. The chi-square distance is defined as

$$d(h_1, h_2) = \sum_v \frac{(h_1(v) - h_2(v))^2}{h_1(v) + h_2(v)} \quad (8)$$

where v represents an index of distribution. Minimization to obtain Laplacian model parameters is defined as

$$\begin{aligned} \alpha_x, \beta_x &= \arg \min_{\alpha, \beta} d(h_x^r, f(v_x | \alpha, \beta)) \\ \alpha_y, \beta_y &= \arg \min_{\alpha, \beta} d(h_y^r, f(v_y | \alpha, \beta)) \end{aligned} \quad (9)$$

where the subscripts x and y represent horizontal and vertical directions of MVs, respectively.

MV distribution of the original video is expressed using Laplacian parameters, which are given by

$$\begin{aligned} h_x^m &= f(v_x | \alpha_x, \beta_x) \\ h_y^m &= f(v_y | \alpha_y, \beta_y). \end{aligned} \tag{10}$$

The difference of MV distribution of the original video and modeled Laplacian distribution is computed using the chi-square distance as

$$\begin{aligned} d_x^r &= d(h_x^m, h_x^r) \\ d_y^r &= d(h_y^m, h_y^r). \end{aligned} \tag{11}$$

Figure 6 illustrates an example of Laplacian modeling for horizontal MVs. Left and right columns in Figures 6(a) and 6(b) show MV histograms of the reference and modeled Laplacian distribution, respectively. In Figure 6(b), the distribution of MVs is modeled using two parameters α and β of the Laplacian distribution. Note that $\alpha=0$ is used since MVs have small values close to zero and only β is estimated.

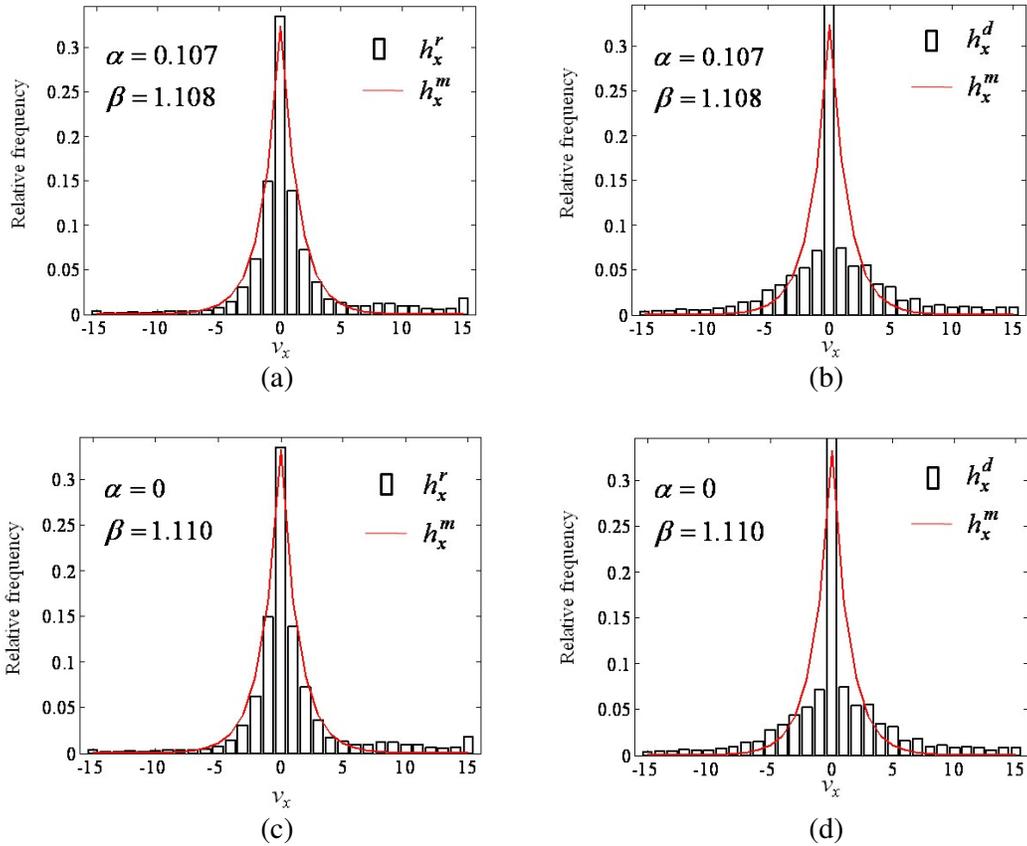


Figure 6. Example of Laplacian modeling for horizontal direction MV; (a), (b) MV histogram of reference video and modeled Laplacian distribution ($\alpha \neq 0$), (c), (d) MV histogram of distorted video and modeled Laplacian distribution ($\alpha = 0$).

Laplacian results are obtained through minimization using the chi-square method. If we know Laplacian parameters, the MV distribution of the original video can be approximately obtained. Through this process we can obtain four parameters: Laplacian modeling parameters of the MV distribution in the horizontal and vertical directions (β_x and β_y , respectively) and the differences of the original MV distribution and the modeled distribution in the horizontal and vertical directions (d_x^r and d_y^r , respectively). The total number of parameters are sent to a receiver as the features, and the final VQA is computed at the receiver. Figure 7 illustrates a block diagram of the proposed RR VQA, in which transmitter and receiver sides are separated.

At the transmitter, h_x^r and h_y^r in the horizontal and vertical directions of the original video can be obtained through MV distribution extraction, and the previously mentioned four parameters estimated through Laplacian modeling are transmitted to the receiver. At the receiver, MVs obtained from the parsing process are used to construct histograms and obtain MV distribution (h_x^d, h_y^d) in the horizontal and vertical directions. The proposed VQA method using Laplacian modeling of MV distribution (VLMVD) is an RR VQA method with the four parameters.

3.2. Proposed VLMVD

In this paper, the proposed VQA transmits the total number of four parameters. The total number of 8-byte meta-data are sent through the ancillary channel, in which two bytes are assigned to each parameter. Our algorithm has an advantage that it requires a smaller amount of meta-data than MVHQM or conventional methods such as FR and RR VQA methods. h_x^m and h_y^m can be obtained from Laplacian parameters β_x, β_y which are transmitted through the ancillary channel.

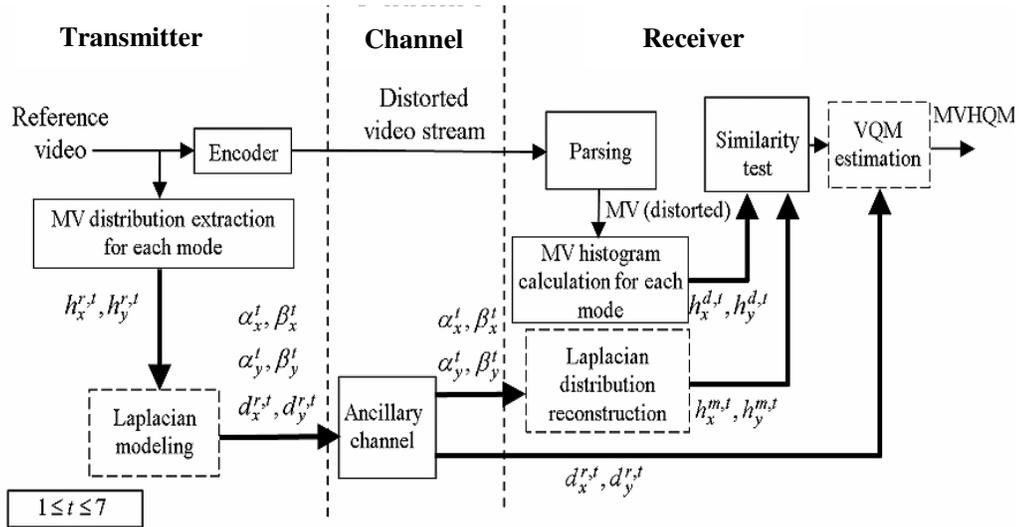


Figure 7. Block diagram of the proposed MVHQM with Laplacian modeling.

Therefore, obtained h_x^m and h_y^m from Laplacian distribution and MV distribution h_x^d and h_y^d obtained from distorted video conduct chi-square method to get difference among parameters, in which $d_x^d(h^m, h^d)$ and $d_y^d(h^m, h^d)$ can be obtained in horizontal and vertical directions

$$\begin{aligned} \hat{d}_x &= |d_x^r - d_x^d| \\ \hat{d}_y &= |d_y^r - d_y^d|. \end{aligned} \quad (12)$$

Thus, we can obtain difference values. The final video quality metric VLMVD is defines as

$$VLMVD = \log_2 \left(\frac{1 + (\hat{d}_x + \hat{d}_y)}{d_0} \right) \quad (13)$$

where d_0 is a constant ($d_0 = 0.001$).

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

To evaluate the performance of the proposed metric and conventional video quality metrics, we use a total of 124 mobile videos in our experiments. This section shows the efficiency of the proposed MVHQMs by comparing the performance and transmission data size with conventional VQAs.

4.1. Test Video Data

Generally, the efficiency of the quality metric is dependent on how quality metric is close to the MOS. Therefore, to evaluate the performance of the proposed algorithm, MOS data (scores) for test video clips are needed. We use the MOS data that are obtained from 30 people for test video clips [21]. A total of 124 mobile videos are used as test video data. The test video data consist of 46 mobile videos coded with H.263 (46 clips) and H.264/AVC (78 clips).

Table 1. Test sequences coded with H.263.

Sequence name	Format	Frame rate	Bit rate (Kbps)
Akiyo	QCIF	15	28/33/60/118
Carphone	QCIF	15	63/135
		30	55/64/122/241
Coastguard	QCIF	30	55/63/130/265
Container	QCIF	15	28/33/64/120
		30	32/63/128/236
Foreman	QCIF	15	64/124
		30	65/132/270
Hallmonitor	QCIF	30	32/56/64/127/230
Mother and daughter	QCIF	15	27/32/64/140
Stephen	QCIF	30	130/259
Table tennis	QCIF	30	55/64/127/240
Weather	CIF	15	195

Tables 1 and 2 list the test sequences used in experiments that are coded with H.263 and H.264/AVC, respectively, in which video format, frame rate, and bit rate are listed. The test video clips consist of CIF/ quarter CIF (QCIF) which are coded under varying bit rates/frame rates by H.263 and H.264/AVC. The DMOS can be computed by applying the double-stimulus continuous quality-scale (DSCQS) [21] to the MOS.

4.2. Performance Comparison of VQA

As a similarity measure for VQA, we use the Pearson correlation coefficient (CC), which is one of quality measures recommended by the VQEG [6], between the quality metric and the DMOS. The higher the Pearson CC is, the better the video quality metric is. Additionally, we use the outlier ratio (OR) for performance comparison of video quality metrics [6], in which the quality metric with a lower OR is better.

The performance of the proposed method is compared with that of conventional methods; FR VQA methods such as edge PSNR [4], video version of SSIM (VSSIM) [10], video version of the image quality metric using singular value decomposition (VMSVD) [12], and Czenakowski distance (CZD) [24], as well as RR VQA methods such as MVHQM [14], estimated PSNR (EPSNR) [17], and local harmonic strength (LHS) with motion consideration factor (MCF) [25]. CZD is a downloaded software, in which only the first 25 frames are used. Table 3 shows the computed Pearson CC values of 46 videos that are classified according to video characteristics. Underlined values are the lowest ones among the methods compared, which means that the performance is the worst. Generally, RR VQA methods give worse performance than FR VQA methods because the number data used in RR VQA methods for representing original image information is much smaller than that used in the FR VQA methods.

Table 2. Test sequences coded with H.264/AVC.

Sequence name	Format	Frame rate	Bit rate (Kbps)
Akiyo	QCIF	15	12/16/27/34/60
Carphone	QCIF	15	12/16/34/62
		30	24/32/59/69/135
Coastguard	QCIF	30	25/34/54/64/133
Container	CIF	15	31/61/124/197
	QCIF		16/32/60
Foreman	CIF	30	62/99/121/259/397
	QCIF		24/32/55/62/132
Hallmonitor	QCIF	30	23/32/56/62/133
Mother and daughter	QCIF	15	12/16/30/33/65
Stephen	QCIF	30	26/31/53/60/123
Table tennis	CIF	30	96/122/242/374
	QCIF		24/32/58/66/123
Weather	CIF	15	66/97/122/259

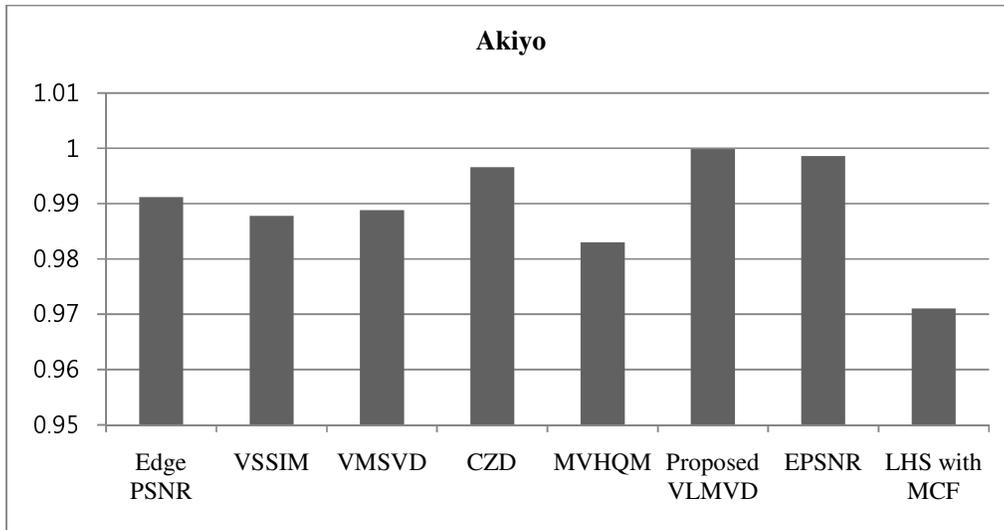
Figure 8 shows that edge PSNR shows the unstable performance among FR quality metrics and that MVQHM and VLMVD show fluctuating performance among RR quality metrics. The proposed method gives the best performance for the Akiyo sequence whereas the worst for the Foreman sequence. The proposed method as a RR VQA method shows the same performance as or slightly worse performance than the FR VQA method.

Table 3 compares the performance of VQA methods in terms of the Pearson CC according to the video type. However, the Pearson CC should be compared over the whole video sequences for a reliable comparison. Table 4 shows the performance comparison of the proposed and conventional methods for the whole video sequences. Bold numbers in Table 4 represent the methods with the best performance.

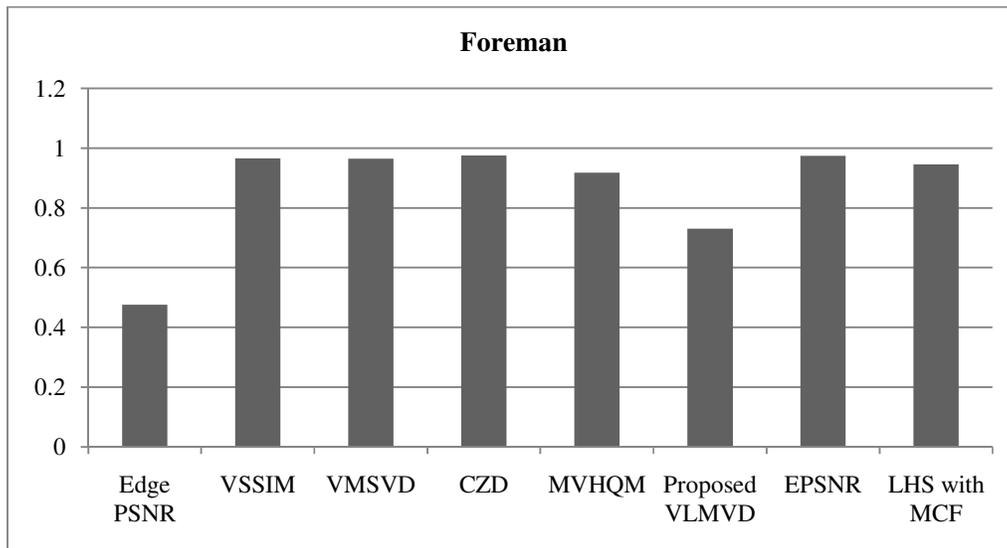
Table 3. Comparison of the Pearson CC of the proposed MVHQM and conventional methods for different video clips

	FR VQA				RR VQA			
	Edge PSNR	VSSIM	VMSVD	CZD	MVHQM	Proposed VLMVD	EPSNR	LHS
Akiyo	0.9912	0.9878	0.9888	0.9966	0.9830	0.9999	0.9986	0.9710
Carphone	0.9990	0.9913	0.9875	0.9844	0.9990	<u>0.7806</u>	0.9944	0.9755
Coast guard	0.9894	0.9880	0.9807	0.9973	0.9598	0.9565	0.9993	0.9696
Container	0.6652	0.9398	0.9098	0.9581	0.8441	0.8160	0.9715	0.9314
Foreman	<u>0.4748</u>	0.9657	0.9644	0.9748	0.9174	<u>0.7303</u>	0.9741	0.9443
Hall monitor	0.8385	0.9510	0.9196	0.9804	<u>0.7839</u>	0.9910	0.9925	0.9177
Mother and daughter	<u>0.5702</u>	0.9951	0.9967	0.9902	0.8060	0.9900	0.9685	0.9962
Silent	0.9543	0.9873	0.9816	0.9950	0.9446	0.9860	0.9677	0.9046
Table tennis	0.9906	0.9882	0.9585	0.9970	0.9977	0.9985	0.9925	0.9755

Generally, FR VQA methods give better results than RR VQA methods in terms of the Pearson CC. However, the proposed VLMVD gives almost the same performance as the FR VQA methods. The proposed method gives the best performance in terms of the OR (the smaller, the better). The proposed VLMVD can give more stable results than the other VQA methods. When compared in terms of the amount of data required for FR VQA methods, a factor of 1.5 is multiplied with the number of frames T for CIF/QCIF video with 4:2:0 YCbCr, because FR VQA methods use both the luminance and (subsampling) chrominance channels. The proposed LHS method with MCF (LHS only) and the proposed metric with MCF (LHS+MCF) are RR VQA methods, with the amount of data required equal to only 1/1024. Note that FR VQA methods need a large number of bits because they require the whole original video. Each frame of CIF and QCIF sequences requires 100 and 25 bytes, respectively. EPSNR requires 68 and 17 bytes per frame for CIF and QCIF sequences [17], respectively. In FR VQA methods, meta-data to be sent increases as the number of the frames increases. However, MVHQM requires 520 bytes per sequence regardless of the number of frames in the sequence, greatly reducing the amount of data required compared with FR VQA methods. The proposed VLMVD requires a total of 8 bytes, greatly reducing the meta-data that should be sent compared with FR as well as RR VQA methods.



(a)



(b)

Figure 8. Pearson CC of two test sequences of the conventional methods and proposed method, (a) Akiyo, (b) Foreman.

In summary, experiments with various test sequences show that the proposed VLMVD has an advantage that it requires a small amount of data, which is an important factor in RR VQA methods. Also the proposed VLMVD gives results similar to the conventional FR and RR VQA methods even though it requires less amount of data for meta-data.

Table 4. Performance comparison of the proposed MVHQM and conventional quality metrics in terms of the Pearson CC, outlier ratio, and data size for the sequences coded with H.263/264/AVC

	VQA	Pearson CC	OR	Data size (byte)
FRVQA	Edge PSNR	0.5510	0.3261	352×288×T×1.5 (CIF) 176×144×T×1.5 (QCIF)
	VSSIM	0.6538	0.3043	
	VMSVD	0.7118	0.3261	
	CZD (result for first 25 frames)	0.4903	0.3043	
RRVQA	MVHQM	0.7022	0.3696	520 (for a video clip)
	Proposed VLMVD	0.6817	0.2391	2×4 (for a video clip)
	EPSNR	0.5938	0.2826	68×T (CIF) 17×T (QCIF)
	LHS with MCF	0.7016	0.3043	100×T (CIF) 25×T (QCIF)

5. CONCLUSIONS

This paper proposes two RR algorithms; MVHQM with and without Laplacian modeling. Conventional FR VQA or RR VQA methods are not suitable for measuring the quality of multimedia services over the network because of a large amount of data required. The proposed method focuses on reducing the amount of data while ensuring the performance for required services. The proposed method approximates the MV distribution of the video using Laplacian modeling. Experimental results with various test sequences show that the proposed method gives the performance comparable to that of the conventional methods, while greatly reducing the amount of data to be transmitted, which is desirable for a quality metric of multimedia services. Another advantage is to get the MV in the parsing process without MV extraction process at the receiver.

The proposed method requires a greatly small number of bits, which is suitable for mobile services. Further research will focus on the extension of the proposed method to high-definition video or video containing texture.

REFERENCES

- [1] Pinson, M. H., and Wolf, S., "A new standardized method for objectively measuring video quality," *IEEE Trans. Broadcasting*, Sept. 2004, vol. 50, no. 3, pp. 312–322.
- [2] Farias, M. C. Q., Carli, M., and Mitra, S. K., "Objective video quality metric based on data hiding," *IEEE Trans. Consumer Electronics*, Aug. 2005, vol. 51, no. 3, pp. 983–992.
- [3] Yasakethu, S. L. P., Fernando, W. A. C., Adedoyin, S., and Konoz, A., "A rate control technique for off line H.264/AVC video coding using subjective quality of video," *IEEE Trans. Consumer Electronics*, Aug. 2008, vol. 54, no. 3, pp. 1465–1473.
- [4] ITU-T Recommendation J.144, "Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference," International Telecommunication Union, Mar. 2004.
- [5] Sheikh, H. R., Wang, Z., Cormack, L., and Bovik, A. C., "LIVE image quality assessment database release 2," <http://live.ece.utexas.edu/research/quality>.

- [6] VQEG, "Draft final report from the video quality experts group on the validation of objective models of multimedia quality assessment, phase I," 2008, <http://www.vqeg.org/>.
- [7] ITU-T Recommendation BT.500-11, "Methodology for the subjective assessment of the quality of television pictures," International Telecommunication Union, Jan. 2002.
- [8] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Processing*, Apr. 2004, vol. 13, no. 4, pp. 600–612.
- [9] Chen, G.-H., Yang, C.-L., and Xie, S.-L., "Gradient-based structural similarity for image quality assessment," in: Proc. Int. Conf. Image Processing, Oct. 2006, Atlanta, GA, pp. 2929–2932.
- [10] Wang, Z., Lu, L., and Bovik, A. C., "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, Feb. 2004, vol. 19, no. 2, pp. 121–132.
- [11] Shnayderman, A., Gusev, A., and Eskicioglu, A. M., "An SVD-based grayscale image quality measure for local and global assessment," *IEEE Trans. Image Processing*, Feb. 2006, vol. 15, no. 2, pp. 422–429.
- [12] Tao, P. and Eskicioglu, A. M., "Video quality assessment using M-SVD," in: Proc. SPIE-IS&T Electronic Imaging, Jan. 2007, vol. 6494, San Jose, CA, pp. (08)1–10.
- [13] Han, H.-S., Kim, D.-O, and Park, R.-H. "Structural information-based image quality assessment using LU factorization," *IEEE Trans. Consumer Electronics*, Feb. 2009, vol. 55, no. 1, pp. 165-171.
- [14] Han, H.-S., Kim, D.-O, Park, R.-H., and Sim, D.-G., "Visual quality metric using one-dimensional histograms of motion vectors," in: Proc. SPIE Image Quality and System Performance V, Jan./Feb. 2008, San Jose, CA, vol. 6808, pp. (52)1–8.
- [15] Kim, D.-O, Park, R.-H., and Sim, D.-G., "Joint feature-based visual quality assessment," *Electron. Lett.*, Oct. 2007, vol. 43, no. 21, pp. 1134–1135.
- [16] Wang, Z. and Simoncelli, E. P., "Reduced-reference image quality assessment using a wavelet-domain natural image statistic model," in: Proc. SPIE Human Vision and Electronic Imaging X, Jan. 2005, San Jose, CA, vol. 5666, pp. 149-159.
- [17] Yamada, T., Miyamoto, Y., Serizawa, M., and Harasaki, H., "Reduced-reference based video quality metrics using representative-luminance values," in: Proc. Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, AZ, Jan. 2007.
- [18] Susstrunk, S. and Winkler, S., "Color image quality on the Internet," in: Proc. IS&T/SPIE Electronic Imaging 2004: Internet Imaging V, Jan. 2004, San Jose, CA, vol. 5304, pp. 118–131.
- [19] Hasler, D. and Susstrunk, S., "Measuring colorfulness in natural images," in: Proc. IS&T/SPIE Electronic Imaging 2003: Human Vision and Electronic Imaging VIII, Jan. 2003, San Jose, CA, vol. 5007, pp. 87–95.
- [20] Wang, Z., Sheikh, H. R., and Bovik, A. C., "No-reference perceptual quality assessment of JPEG compressed images," in: Proc. IEEE Int. Conf. Image Processing, Sept. 2002, Rochester, NY, vol. 1, pp. 477–480.
- [21] Lee, S.-O., and Sim, D.-G., "New full-reference visual quality assessment based on human visual perception," in: Proc. Int. Cont. Consumer Electronics, Jan. 2008, Las Vegas, NV, pp. 1–2.
- [22] Kuo, T.-Y., "Fast variable size motion estimation for H.264 using likelihood and correlation of motion field," *IEEE Trans. Circuits and Systems for Video Technology*, Oct. 2006, vol. 16, no. 10, pp. 1185–1198.
- [23] Brunelli, R. and Mich, O., "On the use of histograms of image retrieval," in: Proc. IEEE Int. Conf. Multimedia Computing and Systems, June 1999, Florence, Italy, vol. 2, pp. 143–147.
- [24] <http://www.semaca.co.uk/homepage.php>.
- [25] Gunawan, I.P. and Ghanbari, M., "Reduced-reference video quality assessment using discriminative local harmonic strength with motion consideration," *IEEE Trans. Circuits and Systems for Video Technology*, Jan. 2008, vol. 18, no.1, pp. 71-83.

Authors

Ho-Sung Han received the B.S. degree in electronic engineering from Sogang University in 2006. Currently he is working toward the M.S. degree in electronic engineering from Sogang University. His current research interests are image/video quality assessment.

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), Suwon, Korea. In 2012, he spent a sabbatical year in Digital Imaging Business (R&D Team) and Visual Display Business (R&D office), Samsung Electronics Co., Ltd., Suwon, Korea. His current research interests are video communication, computer vision, and pattern recognition. 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. He is a co-recipient of the Best Student Paper Award of the IEEE International Symposium on Multimedia (ISM 2006) and IEEE International Symposium on Consumer Electronics (ISCE 2011).

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.

Chang-Dae Jung received the B.S. degree in electronic engineering from Namseoul University, Cheonan, Korea, in 2012. Currently he is working toward the M.S. degree in electronic engineering from Sogang University. His current research interests are image/video quality assessment.