FULL-REFERENCE VIDEO QUALITY ASSESSMENT CONSIDERING STRUCTURAL DISTORTION AND NO-REFERENCE QUALITY EVALUATION OF MPEG VIDEO

Ligang Lu[1] *Zhou Wang*[2] *Alan C. Bovik*[2] *and* *Jack Kouloheris*[1]

[1]Multimedia Technologies, IBM T. J. Watson Research Center, Yorktown Heights, NY 10598
[2]Lab for Image and Video Eng., Dept. of ECE, The Univ. of Texas at Austin, Austin, TX 78703-1084
Email: lul@us.ibm.com, zhouwang@ieee.org, bovik@ece.utexas.edu, jacklk@us.ibm.com

ABSTRACT

There has been an increasing need recently to develop objective quality measurement techniques that can predict perceived video quality automatically. This paper introduces two video quality assessment models. The first one requires the original video as a reference and is a structural distortion measurement based approach, which is different from traditional error sensitivity based methods. Experiments on the video quality experts group (VQEG) test data set show that the new quality measure has higher correlation with subjective quality evaluation than the proposed methods in VQEG's Phase I tests for full-reference video quality assessment. The second model is designed for quality estimation of compressed MPEG video stream without referring to the original video sequence. Preliminary experimental results show that it correlates well with our full-reference quality assessment model.

1. INTRODUCTION

Objective image/video quality measures play important roles in various image/video processing applications, such as compression, communication, printing, analysis, registration, restoration and enhancement. Generally speaking, an image/video quality metric can be employed in three ways. First, it can be used to monitor image/video quality for quality control systems. Second, it can be employed to benchmark image/video processing systems and algorithms. Third, it can be embedded into image/video processing systems to optimize algorithms and parameter settings. The video quality experts group (VQEG) [1], [2] was formed to develop, validate and standardize new objective measurement methods for video quality. Although the Phase I test for full-reference (FR) television video quality assessment only achieved limited success, VQEG continues its work on Phase II test for FR quality assessment for television, and reduced-reference (RR) and no-reference (NR) quality assessment for television and multimedia.

The first goal of this paper is to introduce a new FR video quality assessment approach, which incorporates structural distortion measurement. This method is different from traditional image/video quality assessment approaches, which share a common error sensitivity-based framework as shown in Fig. 1 [3], [4]. Although variances exist and the detailed implementations are different for different models, the underlying principles are the same. First, the original and test image/video signals are subject to preprocessing procedures, possibly including alignment, luminance transformation, and color transformation, etc. A channel decomposition method such as DCT transform, wavelet transform and Gabor decomposition, is then applied to the preprocessed signals. The decomposed signal is weighted differently in different channels according to human visual sensitivities measured in the specific channel. The weighted error signals are adjusted by a visual masking effect model. Finally, an error pooling method, typically the Minkowski metric, is employed to supply a single final quality value. The simplest cases (identity transform and constant weighting) of the error sensitivity-based methods are peak signal-to-nose ratio (PSNR) and mean squared error (MSE), which are the most widely used quality/distortion metrics. Many more sophisticated error sensitivity-based methods were proposed to incorporate human visual system (HVS) characteristics [1], [5]–[8]. It has been shown in [3] that error sensitivity-based method implies a number of assumptions, many of which are questionable. In [3], [9], a structural distortion-based method is proposed for still image quality assessment, which achieves very promising results. This paper attempts to apply the structural distortion-based method for video quality assessment.

In many practical video service applications, especially network video communications, the reference sequence is often not available. Therefore, it is useful to develop NR quality measurement algorithms, where access to the reference video sequence is not required. Little has been done in designing NR video quality assessment methods in the literature [10]–[14]. It is believed that effective NR quality assessment is feasible only when the prior knowledge about the image distortion types is available. In [14], an NR MPEG-2 video quality rating method is proposed, which attempted to predict PSNR by taking advantage of the quantization scale parameters available from the MPEG video stream. The second goal of this paper is to develop an objective NR quality assessment algorithm for MPEG video, which is based on 1) an estimation of quantization errors using MPEG quantization scales and a statistics of the DCT coefficients; 2) an NR evaluation of 8×8 and 16×16 blocking effect; and 3) an adaptive combination of the quantization error estimation and the blocking effect evaluation using the MPEG motion vector information.

2. FULL-REFERENCE VIDEO QUALITY ASSESSMENT USING STRUCTURAL DISTORTION MEASUREMENT

One of the main features of the error sensitivity-based methods is that they treat any kind of image degradation as certain type of *errors*. However, large errors do not always result in large perceptual distortions. Our new philosophy in designing image quality

The authors would like to thank Dr. Philip Corriveau and Dr. John Libert for providing the Matlab routines used in VQEG Phase I test for the regression analysis of subjective/objective data comparison.

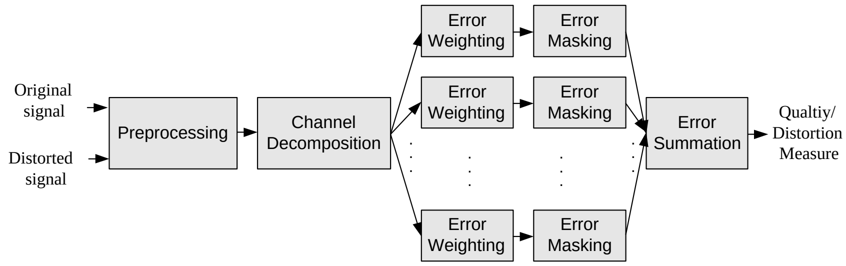


Fig. 1. Error sensitivity-based FR image/video quality measurement system.

metrics is [3], [4]: *The main function of the human eyes is to extract structural information from the viewing field, and the human visual system is highly adapted for this purpose. Therefore, a measurement of structural distortion should be a good approximation of perceived image distortion.* The key point is the switch from *error* measurement to *structural distortion* measurement.

Many different quality assessment methods may be developed using the new philosophy, depending on how the structural distortions are quantified. A simple but effective quality indexing algorithm is proposed in [9], which models any image distortion as a combination of three factors: loss of correlation, luminance distortion, and contrast distortion. More detailed discussion and insights about this new quality index are given in [3], [4], [9]. The new quality index exhibits much more consistency with subjective measures than PSNR. Demonstrative images and an efficient MATLAB implementation of the algorithm are available online at: *http://anchovy.ece.utexas.edu/~zwang/ research/quality_index/demo.html*.

The diagram of the proposed video quality assessment system is shown in Fig. 2. The video quality is first measured frame by frame. For each frame, the corresponding local areas are extracted from the original and the test video sequences, respectively. The local areas are 8×8 blocks randomly selected from the whole picture. In each frame, only a proportion of all possible blocks are selected to reduce computation cost. For each selected local area, statistical features such as mean and variance are calculated and used to classify the local area into smooth region, edge region or texture region. Next, the local quality measure is calculated, which is basically the quality index defined in [9]. The measurement results of all the local areas are averaged to give a quality value of the entire frame. The frame quality value is adjusted by two factors: the blockiness factor and the motion factor. Blocking effect is very common in most image and video coding approaches that use block-DCT transforms and block-based motion estimation/compensation techniques. The blockiness of the frame is measured as a separate procedure on the whole picture. The blockiness measurement method is based on the algorithm introduced in [11], in which the blockiness feature is evaluated in the power spectrum of the image signal. Besides blockiness, the blurring effect is also estimated in the power spectrum, which is characterized by the energy shift from high frequency to low frequency bands. The blockiness measure is used to adjust the overall quality value only if the frame has relatively high quality index value but severe blockiness. This happens frequently in MPEG encoding of large motion frames at low bit rate. Next, we estimate the motion occurred between the current frame and its previous frame. The motion information is obtained by a simple block-based motion estimation algorithm with full pixel resolution. The reason to use motion information is based on the observation that when large motion occurs, the human eyes become less sensitive to the blurring effect. This adjustment is applied only if a frame simultaneously satisfies the conditions of low quality index value, high blurriness and low blockiness, which usually happens when reduced-resolution mode is used in low bit rate MPEG coding.

We consider video sequences with three color components: Y, Cr and Cb. The same algorithm is applied to each components independently and the results are averaged (with a weighting of 0.7 to Y, 0.15 to Cr and Cb each) to give the final frame quality index. Finally, all frame quality index values are averaged to a single overall quality value of the test sequence.

The VQEG Phase I test data set for FR video quality assessment (available at *http://www.vqeg.org*) is used to test the system. Figs. 3(a) and (b) show the scatter plots of the subjective/objective comparisons on all test video sequences given by PSNR and the proposed method, respectively. It can be observed that the proposed method has better consistency with the subjective measurements. This is confirmed by Fig. 4, which shows the regression correlation and variance-weighted regression correlation values between the subjective and objective evaluations of all the test video sequences (They are defined as Metric 2 and Metric 1, respectively, in VQEG Phase I test to evaluate the prediction accuracy of the objective model [1]). The 95% confidence interval error bar of each method is also given in the same figure. It can be seen that higher correlation values are achieved by the new system.

3. NO-REFERENCE QUALITY MEASUREMENT OF MPEG VIDEO STREAM

Fig. 5 shows the diagram of the proposed NR MPEG video quality measurement system. The input to the system is compressed MPEG video bitstream. The output quality index value can be evaluated on either a frame or a sequence basis, depending on the application. First, the input MPEG video bitstream is partially decoded and we get 1) the inverse quantized DCT coefficients; 2) the quantization scale; and 3) the motion vector for each block. Second, we estimate the quantization error. A histogram statistics is conducted on the inverse quantized DCT coefficients, which are available from the MPEG decoder. With this histogram, we can estimate the distribution on a piece-wise basis (different from [14]). For a certain DCT coefficient, if the inverse quantized value is L and the quantization scale is q, then the quantization error is estimated as

$$E = \frac{\int_{L-q/2}^{L+q/2} |x-L|^2 p(x)dx}{\int_{L-q/2}^{L+q/2} p(x)dx} \ , \tag{1}$$

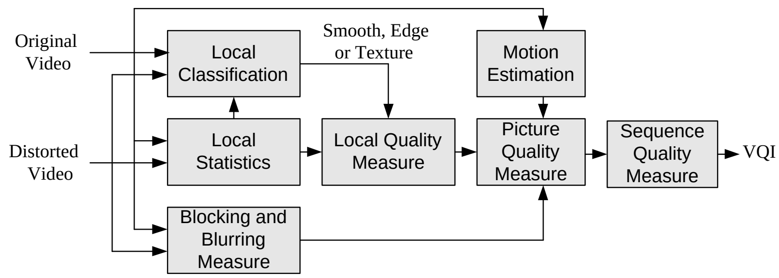
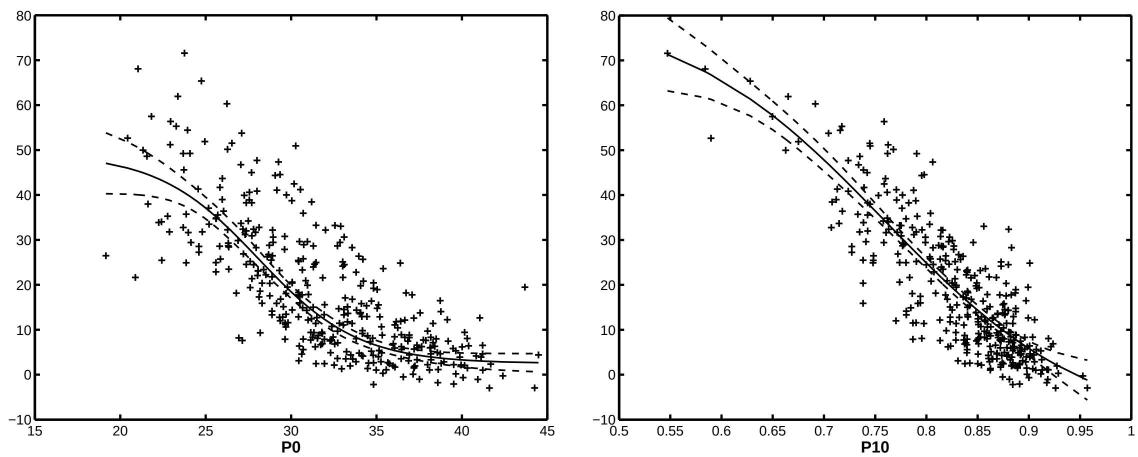


Fig. 2. Proposed FR video quality assessment system.



(a) (b)

Fig. 3. Comparison on VQEG test data set. Vertical and horizontal axes are for subjective and objective measurements, respectively. Each sample point represents one test video sequence. (a) PSNR; (b) Proposed FR method.

where $p(x)$ is the estimated probability density distribution of the DCT coefficients. The quantization errors of all the DCT coefficients are then averaged together as an estimate of the overall quantization error of the frame. Next, we evaluate the blocking effect using a simplified implementation of the idea first introduced in [11]. We also evaluate the motion information using the motion vectors extracted from the MPEG bitstream. Currently, only the magnitude of the motion vectors is calculated and used by our algorithm. Finally, we adaptively combine the quantization error estimation with the blocking effect estimation. Usually, we use a simple linear combination of these two factors and normalize it to generate a single overall quality measure of the frame. The motion information is used to adjust the evaluation of the frames with large motion. The frame quality values are averaged to provide a quality measurement of a group of pictures or the whole video sequence.

We used 704×480 video sequences to test the new approach and compare with our FR quality measurement approach introduced in Section 2. The video sequences were compressed with a MPEG-2 encoder at 1.2 Mega-bits/sec (Mbps), 2.4 Mbpt and 4.8 Mbps, respectively. Table 1 compares the measurement results of the FR and NR approaches. The linear correlation coefficient between the two data sets is 0.9671.

4. CONCLUSIONS

In this paper, we first introduced a new FR objective video quality assessment system. The key feature of the proposed method is the use of structural distortion measurement. Experiments on VQEG Phase I test data set for FR video quality assessment show that it has better correlation with perceived video quality than the proposed methods in VQEG's Phase I test. A new approach for NR quality measurement of MPEG video is also proposed by combining quantization error estimation, blocking effect estimation and the motion information. Our preliminary experiments show that it correlates well with the proposed FR video quality index. In the future, more experiments are needed to fully validate the methods. Furthermore, we are using these quality measures to optimize MPEG encoders to provide better quality video services.

5. REFERENCES

[1] VQEG, "Final report from the video quality experts group on the validation of objective models of video quality assessment," *http://www.vqeg.org/*, Mar. 2000.

[2] P. Corriveau, *et al.*, "Video quality experts group: Current results and future directions," *Proc. SPIE Visual Comm. and Image Processing*, vol. 4067, June 2000.

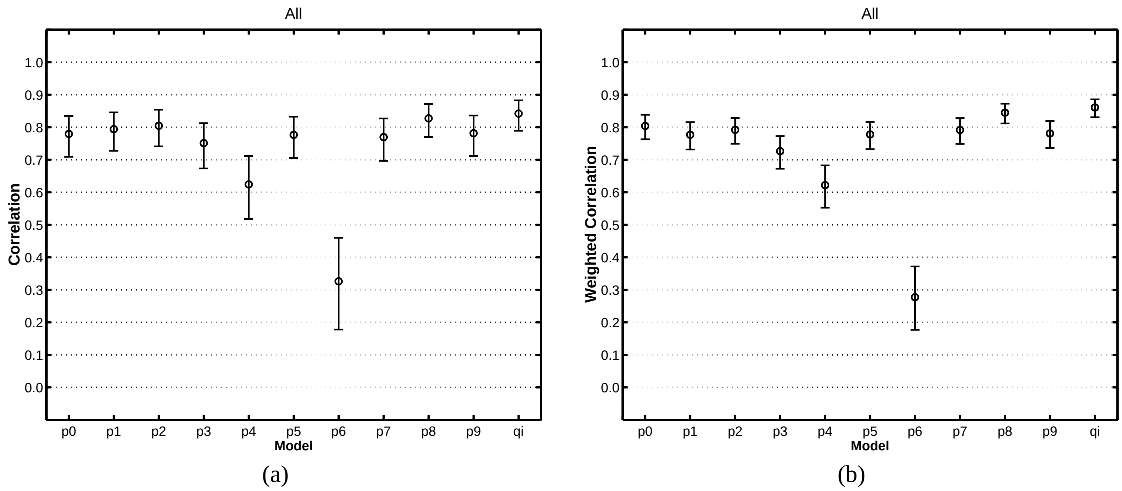


Fig. 4. Regression correlation comparisons. p0∼p9: Indices of the proponents in VQEG Phase I test [1]. qi: the proposed FR video quality index. The error bars represent 95% confidence intervals. (a) Regression correlation; (b) Variance-weighted regression correlation.

Table 1. FR/NR MPEG Video Quality Measurement Comparison

	FR Quality Index			NR Quality Index		
Sequence	1.2Mbps	2.4 Mbps	4.8 Mbps	1.2 Mbps	2.4 Mbps	4.8 Mbps
Marb	0.4801	0.6290	0.7970	0.4724	0.6099	0.7480
Cheerleader	0.5259	0.6854	0.8083	0.5116	0.6053	0.7183
Hockey	0.5836	0.7435	0.8611	0.4897	0.6658	0.7978
Speedbike	0.6609	0.7569	0.8244	0.5706	0.6978	0.7936

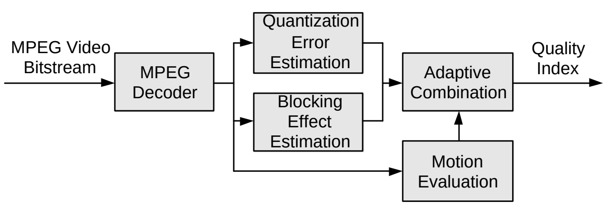


Fig. 5. Proposed NR MPEG video quality measurement system.

[3] Z. Wang, A. C. Bovik, and L. Lu, “Why is image quality assessment so difficult?,” in *Proc. IEEE Int. Conf. Acoust., Speech, and Signal Processing*, (Orlando), May 2002.

[4] Z. Wang, *Scalable foveated image and video communications*. PhD thesis, Dept. of ECE, The University of Texas at Austin, Dec. 2001.

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

[6] J. Lubin, “A visual discrimination mode for image system design and evaluation,” in *Visual Models for Target Detection and Recognition* (E. Peli, ed.), pp. 207–220, Singapore: World Scientific Publishers, 1995.

[7] S. Daly, “The visible difference predictor: An algorithm for the assessment of image fidelity,” in *Proc. SPIE*, vol. 1616, pp. 2–15, 1992.

[8] A. B. Watson, J. Hu, and J. F. III. McGowan, “Digital video quality metric based on human vision,” *Journal of Electronic Imaging*, vol. 10, no. 1, pp. 20–29, 2001.

[9] Z. Wang and A. C. Bovik, “A universal image quality index,” *IEEE Signal Processing Letters*, vol. 9, pp. 81–84, Mar. 2002.

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

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

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

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

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