VIDEO QUALITY ASSESSMENT USING STRUCTURAL DISTORTION MEASUREMENT

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

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

ABSTRACT

Objective image/video quality measures play important roles in various image/video processing applications, such as compression, communication, printing, analysis, registration, restoration and enhancement. Most proposed quality assessment approaches in the literature are error sensitivity-based methods. In this paper, we follow a new philosophy in designing image/video quality metrics, which uses structural distortion as an estimation of perceived visual distortion. We develop a new approach for video quality assessment. Experiments on the video quality experts group (VQEG) test data set shows that the new quality measure has higher correlation with subjective quality measurement than the proposed methods in VQEG's Phase I tests for full-reference video quality assessment.

1. INTRODUCTION

There has been an increasing need recently to develop objective quality measurement techniques that can predict perceived image/video quality automatically. These methods are useful in various image/video processing applications, such as compression, communication, printing, displaying, analysis, registration, restoration and enhancement. Generally speaking, these methods can be employed in three ways. First, they can be used to monitor image/video quality for quality control systems. Second, they can be employed to benchmark image/video processing systems and algorithms. Third, they can also 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.

Most of the proposed objective image/video quality assessment approaches in the literature share a common error sensitivity-based philosophy [3]. The framework of a typical error sensitivity-based approach is shown in Fig. 1 [3]. 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 is then applied to the preprocessed signals. There are many choices for channel decomposition, such as identity transform, wavelet transforms, and Gabor decompositions. The decomposed signal is treated differently in different channels according to human visual sensitivities measured in the specific channel. The errors between the two signals in each channel are calculated and weighted. The weighted error signals are adjusted by a visual masking effect model, which reflects the reduced visibility of errors presented on the background signal. 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 have been proposed to incorporate human visual system (HVS) characteristics [1], [4]–[7].

It has been shown in [3] that error sensitivity-based method implies a number of assumptions, many of which are questionable. In [3], [8], [9], a structural distortion-based method is proposed for still image quality assessment, which achieves very promising results. In this paper, we apply the structural distortion-based method for video quality assessment.

2. STRUCTURAL DISTORTION-BASED METHOD

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 metrics is [3], [9], [10]: *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 [8]. Let $\mathbf{x} = \{\, x_i \,|\, i = 1, 2, \cdots, N\}$ and $\mathbf{y} = \{\, y_i \,|\, i = 1, 2, \cdots, N\}$ be the original and the test image signals, respectively. The proposed quality index is defined as

$$Q = \frac{4\,\sigma_{xy}\,\bar{x}\,\bar{y}}{(\sigma_x^2 + \sigma_y^2)\,[(\bar{x})^2 + (\bar{y})^2]}\ , \tag{1}$$

where $\bar{x}$, $\bar{y}$, σ_x^2, σ_y^2 and σ_{xy} are the estimates of the mean of **x**, the mean of **y**, the variance of **x**, the variance of **y**, and the covariance of **x** and **y**, respectively. The dynamic range of Q is

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.

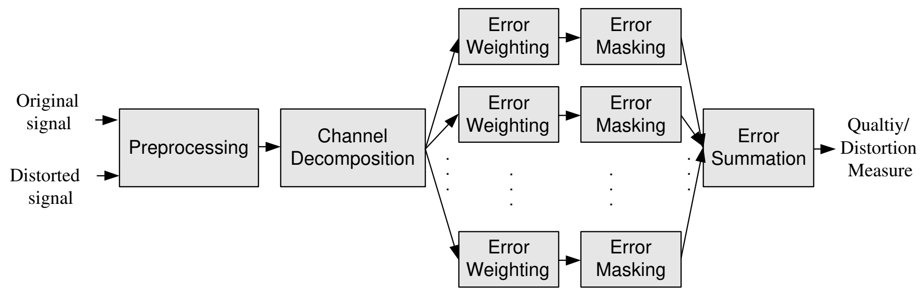

Fig. 1. Error sensitivity-based image/video quality measurement system.

$[-1, 1]$. The best value 1 is achieved if and only if $y_i = x_i$ for all $i = 1, 2, \cdots, N$. More detailed discussion and insights about this new quality index are given in [3], [8], [9].

The quality index is applied to natural images using a sliding window approach, with a window size of 8×8. The quality indices are calculated within the sliding window, leading to a quality map of the image. The overall quality index value is the average of the quality map. Some test images are shown in Fig. 2, where the original images are distorted by blurring and JPEG compression. Both MSE and Q are calculated for the distorted images. It can be observed that the images with nearly identical MSE values may have drastically different visual quality. The new quality index exhibits much more consistency with subjective measures. More demonstrative images and an efficient MATLAB implementation of the proposed algorithm are available online at: *http://anchovy.ece.utexas.edu/~zwang/ research/quality_index/demo.html*.

3. VIDEO QUALITY ASSESSMENT

The diagram of the proposed video quality assessment system is shown in Fig. 3 [10]. 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 (1). 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. Except for 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. 4(a), (b), (c) and (d) show the scatter plots of the subjective/objective comparisons on all test video sequences given by PSNR, the Sarnoff/Tektronix model, the Swisscom/KPN model, 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. 5, 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.

4. CONCLUSIONS

We designed a new 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 good correlation with perceived video quality. More extensive experiments are needed to further improve and fully test the system.

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.

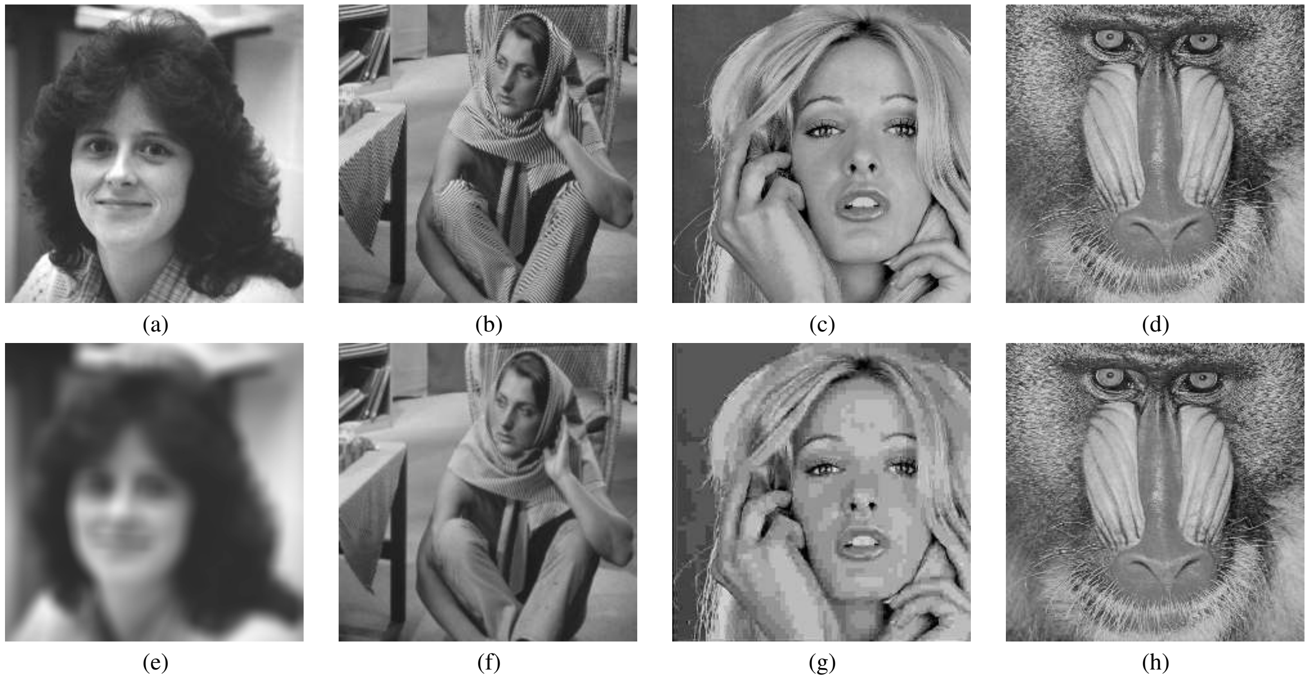

Fig. 2. (a), (b), (c) and (d): original test images “Woman”, “Barbara”, “Tiffany” and “Mandrill”. 512×512, 8bits/pixel; (e) Blurred “Woman” image, MSE = 200, Q = 0.3483; (f) Blurred “Barbara” image, MSE = 200, Q = 0.6594; (g) JPEG compressed “Tiffany” image, MSE = 165, Q = 0.3709; (h) JPEG compressed “Mandrill” image, MSE = 163, Q = 0.7959.

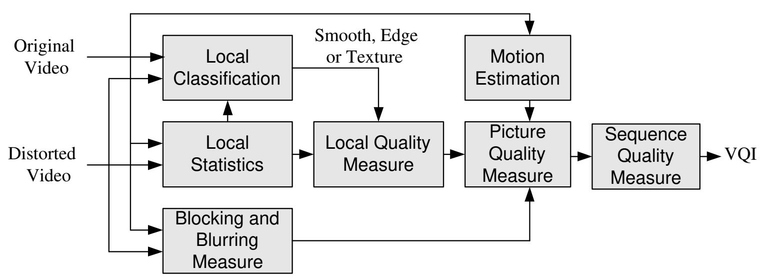

Fig. 3. Proposed video quality assessment system.

[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.

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

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

[5] 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.

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

[7] 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.

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

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

[10] Z. Wang and L. Lu, “Objective image and video quality assessment,” *Technical Report: IBM T. J. Watson Research Center*, Aug. 2001.

[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.

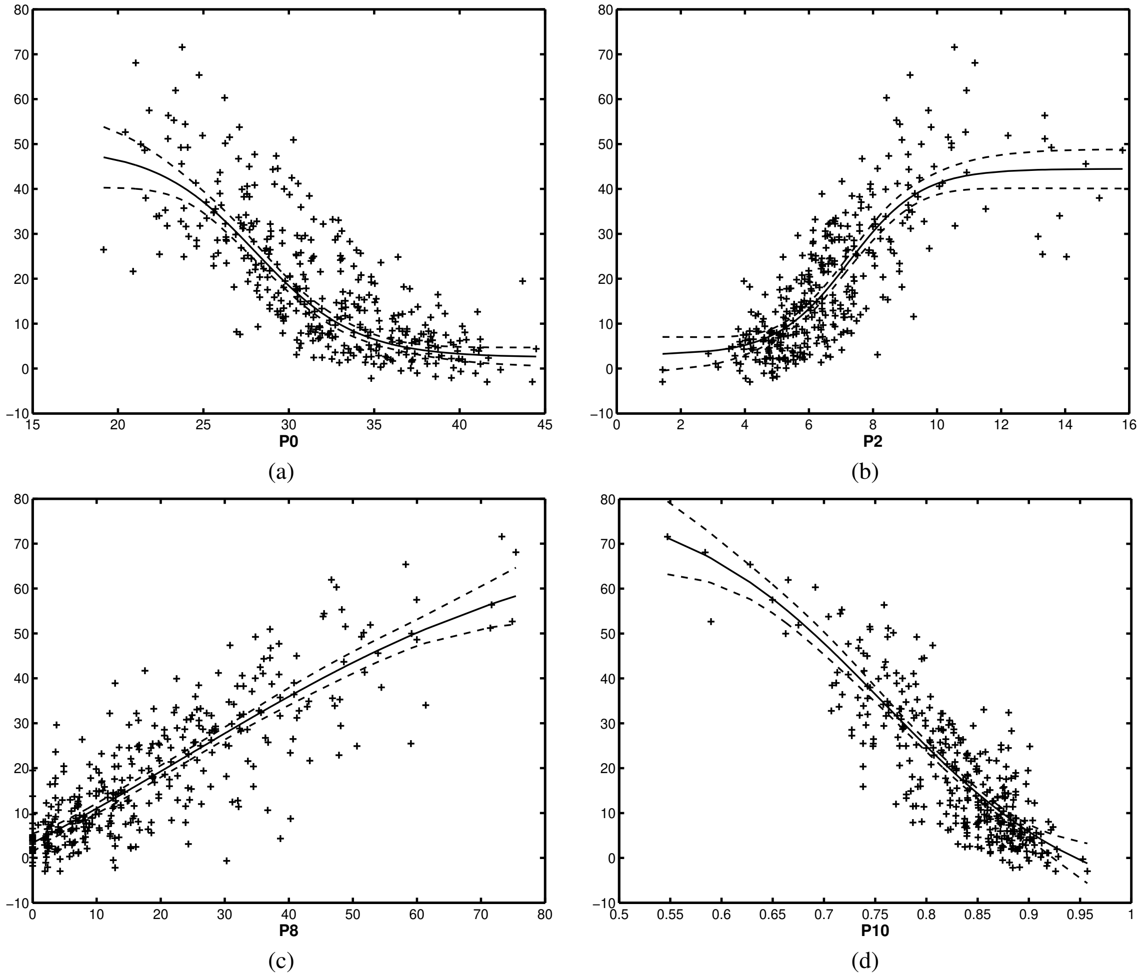

Fig. 4. Comparison on VQEG test data set. Vertical and horizontal axes are for subjective and objective measurements, respectively. Each sample point represents one test video sequence. (a) PSNR; (b) Sarnoff/Tektronix model; (c) Swisscom/KPN model; (d) Proposed method.

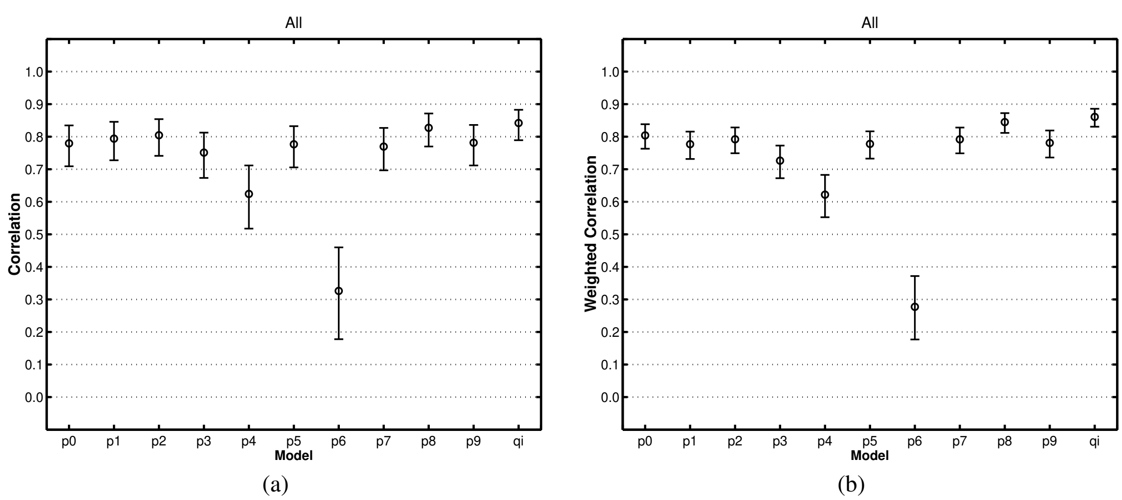

Fig. 5. Regression correlation comparisons. p0∼p9: Indices of the proponents in VQEG Phase I test [1]. qi: the proposed video quality index. The error bars represent 95% confidence intervals. (a) Regression correlation; (b) Variance-weighted regression correlation.