A Human Visual System-Based Objective Video Distortion Measurement System

Zhou Wang and Alan C. Bovik
Laboratory for Image and Video Engineering (LIVE), Department of Electrical and Computer Engineering,
The University of Texas at Austin, Austin TX-78703, USA.
Email: {zwang, bovik}@ece.utexas.edu

Abstract-- **Recently, it has become increasingly important to devise video quality/distortion measurement algorithms that help to evaluate, compare and improve video coding techniques, and hence to provide effective digital quality video services. This paper introduces a practical model for objective visual distortion measurement of compressed videos. The model is established by making use of both spatial and temporal human visual system (HVS) features, which include spatial frequency sensitivity, luminance masking, texture masking, temporal frequency sensitivity, and short-term memory effect. A 'Video Compare' software is developed to demonstrate the system.**

Key Words-- **Human Visual System, Video Distortion Measurement.**

I. INTRODUCTION

In recent years, there have emerged many successful digital video coding techniques as well as a proliferation of digital video coding products. These techniques and products address a broad range of applications, having different quality and bandwidth requirements. As a result, it has become increasingly important to devise video quality/distortion measurements that help to evaluate, compare and improve the video coding techniques and products, and hence to provide effective digital quality video services. Currently, the only widely accepted methods are the subjective measurement of mean opinion score (MOS) and the objective measurements of mean square error (MSE). MOS is very tedious, expensive and slow. In addition, it is very difficult to be embedded into a practical video processing system because of its impossibility of automatic implementation. Instead, MSE is more widely used because it has no ambiguity, it is simple and fast to calculate, and it is mathematically easy to deal with. However, the MSE is a poor measure [1] of the perceived image and video distortion because it does not consider the human visual perception of image/video distortions.

In the last three decades, there have been many efforts to design objective image distortion models by taking advantage of the human visual system (HVS) features [1-8]. Only a few models are proposed for video distortion measurements [9-11]. This paper introduces a practical objective visual distortion measurement model for digital video compression. The primary purpose of our work is to evaluate the video coding algorithms for the compression and transmission of high quality digital video data over the OC3-ATM networks, where a bandwidth of about 110~130Mbps is allowed. However, this model can also be used for the evaluation of lower quality videos such as the H.263, MPEG-1, and MPEG-2 compressed videos. Our visual distortion assessment model combines the most significant HVS features, which include spatial frequency sensitivity, luminance masking, texture masking, temporal frequency sensitivity, and short-term memory effect. Moreover, it has a simple and clear structure, and is easy for software implementation.

II. HVS-BASED VIDEO DISTORTION MEASUREMENT

A. HVS Features

Various HVS features are correlated with perceptual image/video quality [12]. We choose the most significant ones among them.

First, the error sensitivity of human eyes is a function of spatial frequency. Basically, the function can be viewed as a band pass filter with a frequency response reaching the highest value at about 4 cycles per degree of visual angle and decreasing very fast with increasing spatial frequency. We designed a twodimensional filter to model this feature. Its frequency response is shown in Fig. 1.

Second, the HVS's sensitivity to the variations in luminance depends on the local mean luminance. This is called "light adaptation" or "luminance masking". The luminance masking function is non-linear. A model developed in [3] is shown in Fig. 2. We adopt this model in our system.

Third, consider two stimuli in the same image, the presence and the features of one stimulus will influence the way the other one is perceived. This is what we called the texture masking effect. Our system considers the error between the original and the test images as the target and the original and test images as the maskers. The texture masking effect is determined by local frequency distribution and texture direction. An example is given in Fig. 3, where the same amount of random noise is added to the areas with different frequency distribution backgrounds. It can be observed that the noise added to flat (low frequency) background is much more visible than that added to texture (high frequency) background.

Fourth, the visual error sensitivity is also a function of temporal frequency. In general, the transfer function is a band-pass filter with the highest response at about 8Hz. A digital filter is designed to implement this. The frequency response of the filter is shown in Fig. 4.

Furthermore, short-term memory effect is also considered. Because of the short-term memory effect, the influence of a strong stimulus will last for a short time. Fig. 5 is an illustration of this effect. Basically, the short-term memory effect is a kind of smoothing effect on the distortion measure of the frames. An interesting feature is that people are more likely to remember the bad quality frames than the good ones. Insert a bad frame in a consecutive sequence of high quality frames, people can easily notice that bad frame and it seems to them that there exist several bad frames instead of only one. On the contrary, if a high quality frame is inserted into a consecutive sequence of bad frames, it only has very little effect on the overall visual quality of the video sequence.

B. The Measurement System

The video quality measurement system is shown in Fig.6. First, we compare directly each of the original image frame with the corresponding test video frame. The numerical differences are computed as the initial error map. Various numerical distortion measures such as MSE, MAE (mean absolute error) and maximum absolute error can be calculated based on this initial error map. Second, the 2-D spatial frequency error sensitivity filter is applied to each frame. Third, the in-frame masker is evaluated with both the original and the test image frames. Both luminance masking and contrast masking effects at each point in the image are computed. The resulting masker map is then used to scale the filtered error map. Up to this point, we think of the video sequence as a collection of independent frames and only spatial HVS features are considered. By in-frame error pooling, the visual distortion of each frame is achieved. Such in-frame measurement is useful for very high quality video coding applications, where people require to keep every visual detail within the video data and may stop at any frame and try to find visual distortions. Next, the temporal frequency error sensitivity filter is applied, followed by a frame-by-frame error pooling. The result is a frame-varying distortion curve. The short-term memory effect is modeled as a smoothing filter on this curve. The smoothing is asymmetric. For each frame, we first classify it as a good frame or a bad frame by comparing its distortion value with its adjacent frames. The good frames are smoothed out by their adjacent frames while the bad frames do not change and degrade the quality of the next several frames. Finally, the overall visual distortion measure of the video sequence is given by the average value of the smoothed distortion curve.

A 'Video Compare' software, shown in Fig. 7, is implemented to demonstrate our measurement system. It is developed under Microsoft Windows NT/98 environment using Microsoft Visual C++. The software can run multiple YCrCb format video sequences simultaneously and can stop at any specified frame. This helps the users to do subjective measurement test. The software also shows numerical error maps, visual error maps, masker maps, numerical error distortion measurement results and histograms, visual error distortion measurement results and histograms, and the overall distortion value. In addition, it has many other functions, such as the enhancement of error maps and masker maps.

III. CONCLUSIONS

In this paper, we propose a practical HVS-based video distortion measurement model. The model has a simple structure and is easy for software implementation. A "Video Compare" software is developed to demonstrate the system. The system provides a useful tool for the evaluation and design of the video compression algorithms that will be employed in various video coding and transmission applications.

IV. REFERENCES

[1] B. Girod, "What's wrong with mean-squared error," *Digital Images and Human Vision*, A. B. Watson, ed., pp. 207-220, the MIT press, 1993.

[2] T. N. Pappas, and R. J. Safranek, "Perceptual criteria for image quality evaluation," in *Handbook of Image and Video Processing*, A. C. Bovik, ed., New York: Academic Press, May 2000.

[3] C. H. Chou and Y. C. Li, "A perceptually tuned subband image coder based on the measure of Just-Noticeable-Distortion profile," *IEEE Trans. Circuits & Systems for Video Technology*, vol.5, no. 6, pp. 467-476, Dec. 1995.

[4] S. Daly, "The visible difference predictor: An algorithm for the assessment of image fidelity," *Proceedings of the SPIE*, vol. 1616, pp. 2-15, 1992.

[5] J. Lubin, "A visual discrimination mode for image system design and evaluation," *Visual Models for Target Detection and Recognition*, E., Peli, Ed., World Scientific Publishers, Singapore, 1995.

[6] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in image based on human visual sensitivity," *IEEE Transactions on Image Processing*, vol. 4, pp. 713-724, June 1995.

[7] Z. Wang, A. C. Bovik, and B. L. Evans, "Blind measurement of blocking artifacts in images," *IEEE International Conference on Image Processing*, accepted for publication, Vancouver, Sep. 2000.

[8] A. M. Eskicioglu, and P. S. Fisher, "Image quality measures and their performance," *IEEE Trans. Commun.*, vol. 43, no. 12, pp. 2959-2965, Dec. 1995.

[9] C. J. van den Branden Lambrecht, "Perceptual quality measure using a spatio-temporal model of the human visual system," *Proceedings of the SPIE*, vol. 2668, pp. 450-461, San Jose, 1996.

[10] A. B. Watson, "Toward a perceptual video quality metric," *Human Vision and Electronic Imaging III*, *Proce. of the SPIE*, vol. 3299, pp. 139-147, Jan. 1998.

[11] K. T. Tan, M. Ghanbari, and D. E. Pearson, "An objective measurement tool for MPEG video quality," *Signal Processing*, vol.70, no.3, pp. 279-294, Nov. 1998.

[12] B. A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., 1995.B. A. Wandell, *Foundations of Vision*, Sinauer Associates, Inc., 1995.

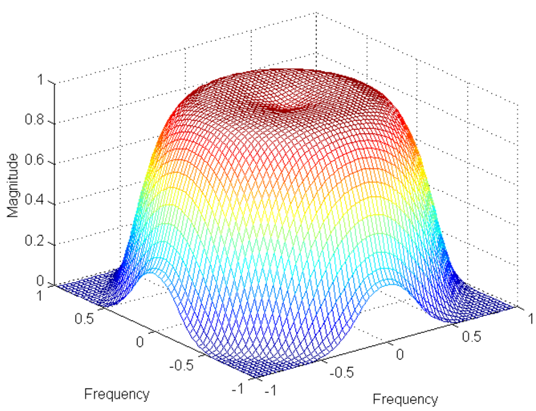

Fig. 1. Spatial frequency error sensitivity filter.

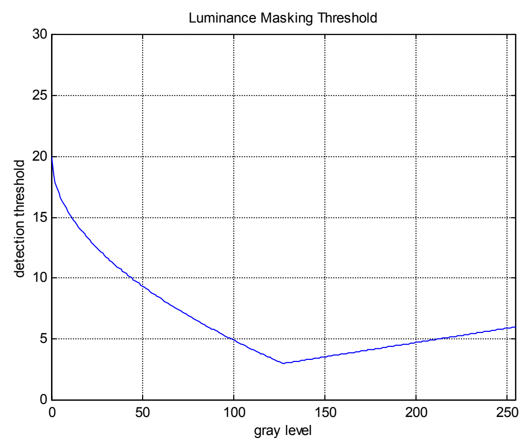

Fig. 2. Luminance masking threshold.

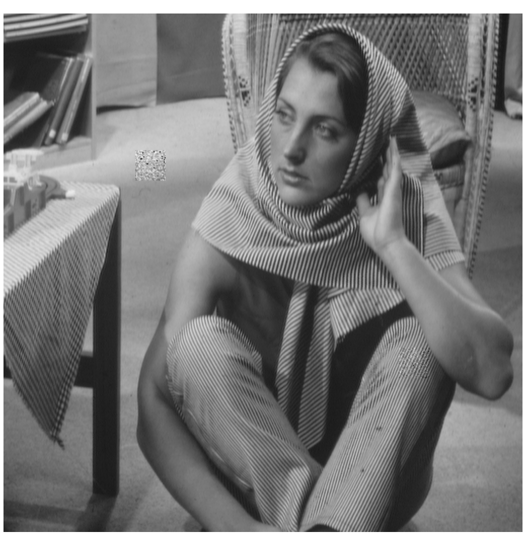

Fig. 3. Texture masking.

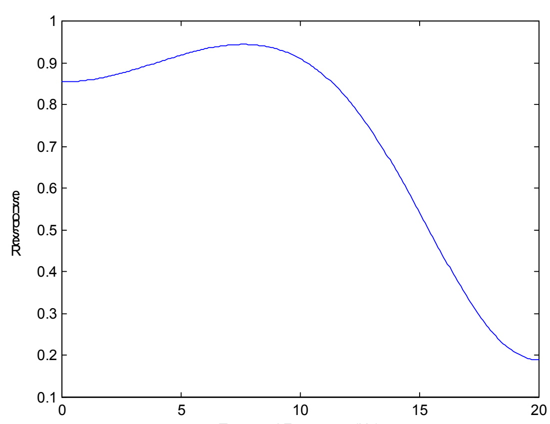

Fig. 4. Temporal frequency error sensitivity function.

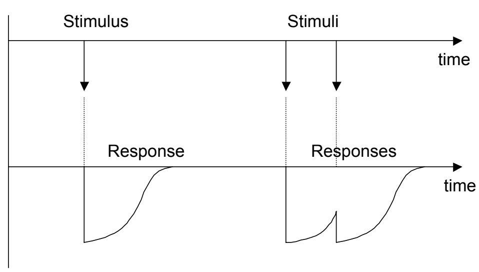

Fig. 5. Short-term memory effect.

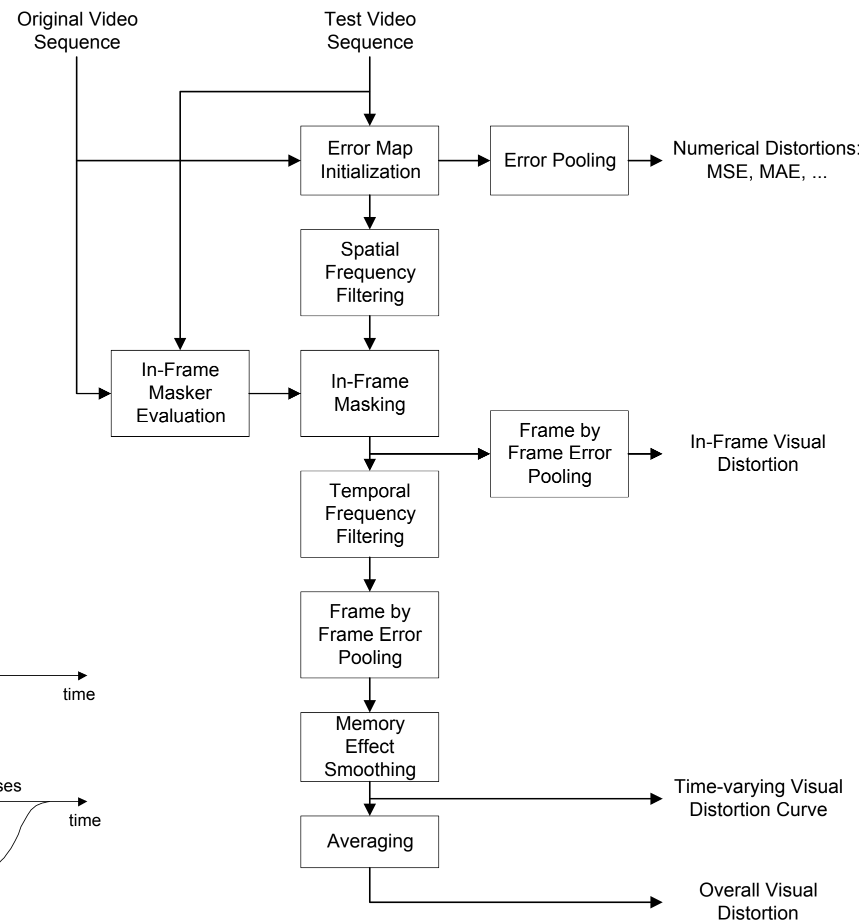

Fig. 6. The video distortion measurement system.

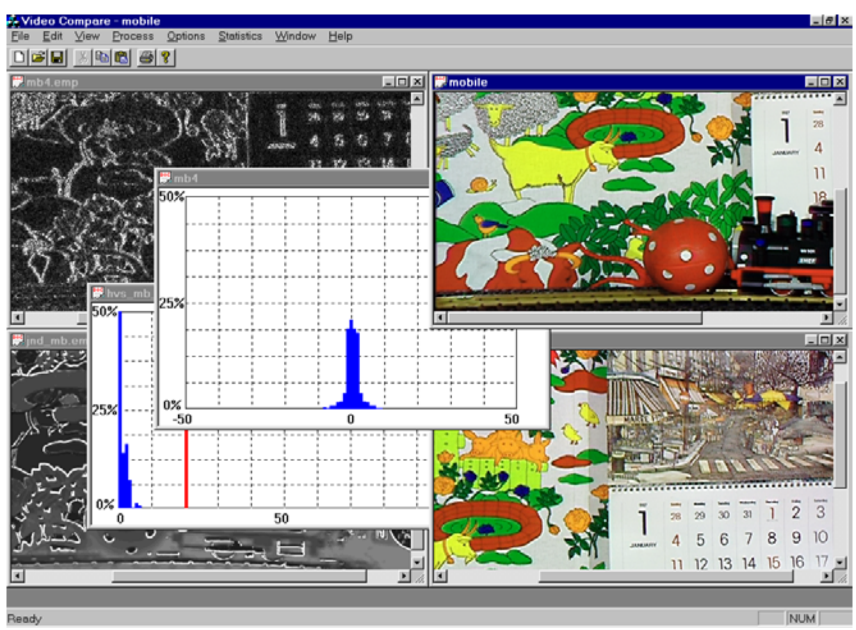

Fig. 7 The ‘Video Compare’ software.