University of Texas at Austin

Department of Electrical and Computer Engineering

EE 381K: Multidimensional Digital Signal Processing

Objective Image/Video Quality Measurement — A Literature Survey

Zhou Wang

Abstract

With the advent of various video compression standards and a proliferation of digital video coding products that are just beginning to appear in the marketplace, it has become increasingly important to devise image/video quality assessment algorithms that will standardize the assessment of compressed digital image/video quality. The subjective assessment of Mean Opinion Score (MOS) is very tedious, expensive and cannot be conducted in real time. It is also very difficult to be embedded into a practical video processing system because it cannot be implemented automatedly. In the last two decades, there have been a lot of attempts to provide objective measures for image/video quality. A recent trend is to incorporating Human Vision System (HVS) features into the quality metrics to make the new measurements more consistent with human visual perception. This literature survey is to give a general description on the various considerations on the development and implementation of image/video quality assessment systems.

I. INTRODUCTION

With the advent of various video compression standards and a proliferation of digital video coding products that are just beginning to appear in the marketplace, it has become increasingly important for the telecommunication, computer and media communities to devise image/video quality assessment algorithms that will standardize the objective assessment of compressed digital video quality to be utilized in multimedia, CDs, DVDs, HDTV, web-based video services, digital telephony, etc. The subjective measurement Mean Opinion Score (MOS) is a widely used method on the assessment of image/video quality, but it has two obvious disadvantages. First, it is very tedious and expensive, thus cannot be conducted in real time. Second, it is very difficult to be embedded into a practical video processing system because it is impossible to be implemented automatedly. Instead, an objective image/video quality metric can provide a quality value for a given image/video automatedly in a relatively short period of time. This is very important for real world applications.

In the last two decades, a lot of objective metrics have been proposed [2, 4-6, 9-24] to assess image/video quality. The easiest way to give a quality value is to use some simple statistics features on the numerical errors between the distorted image and a reference image. The most widely adopted statistics feature is the Mean Squared Error (MSE). However, MSE and its variants do not correlate well with subjective quality measures because human perception of image/video distortions and artifacts is unaccounted for. MSE is also not good because the residual image is not uncorrelated additive noise. It contains components of the original image. A detailed discussion on MSE is given by Griod [1].

A major emphasis in recent research has been given to a deeper analysis of the Human Visual System (HVS) [2] features. There are a lot of HVS characteristics [3] that may influence the human visual perception on image/video quality. Although HVS is too complex to fully understand with present psychophysical means, the incorporation of even a simplified model into objective measures reportedly leads to a better correlation with the response of the human observers [2]. Many algorithms have successfully employed HVS models [2, 4, 5, 6, 10-24].

Another important factor for the development of an image/video quality metric is the flexibility for practical implementations. Some of the metrics consider only some special types of distortions [4] or special image/video coding methods [5, 6]. As for color images, finding a good color space where each color channels can be considered independently is desired [3, 7, 8]. The implementation for a practical video quality assessment metric is difficult because of the computational complexity. In this case, speed is one of the major considerations.

II. SIMPLE STATISTICS ERROR METRICS

In [2], a number of simple statistics metrics on numerical errors are compared for gray scale image compression. These metrics include average difference, maximum difference, absolute error, MSE, peak MSE, Laplacian MSE, histogram, Hosaka plot (A graphic quality measure. The area and shape of the plot gives information about the type and amount of degradation.), etc. It is shown that although some numerical measures correlate well with the observers' response for a given compression technique, they are not reliable for an evaluation across different techniques.

The major advantage of the simple statistics error metrics is their simplicity. They can be very conveniently adapted by an image/video processing system. However, the lacking of considering HVS features make them not good for perceptual image/video distortion. It is shown in [2] that small improvement can be obtained by combing only a very simple HVS model.

III. HVS FEATURE BASED ALGORITHMS

A. HVS Features

Various HVS features are correlated with perceptual image/video quality [3]. Among them, the most commonly used are luminance contrast sensitivity, frequency contrast sensitivity and masking effects.

The human eye is sensitive to luminance rather than the absolute luminance value. According to Weber's law [9], if the luminance of a test stimulus is just noticeable from the surrounding luminance, then the ratio of just noticeable luminance difference to stimulus' luminance, known as Weber fraction is approximately constant [10]. In practice, due to the present of ambient illumination surrounding the display, the noise in very dark areas tends to be less visible than that occurring in regions of higher luminance. Therefore, as the background luminance is low, the Weber fraction increases as the background luminance decreases [11]. On the other hand, if the background luminance is high, the Weber fraction remains constant as the background luminance is increases.

Contrast sensitivity is also varies with spatial frequency, which leads to the concept of Contrast Sensitivity Function (CSF) [3]. It has always been well known that the eye is much more sensitive to lower spatial frequency than to higher ones. The property has been widely exploited to design television sets and cameras. In fact, CSF is a multivariate function of the spatial frequency, the temporal frequency, the orientation, the viewing distance and the color direction.

Consider two different 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 masking effect. The masking effect is so complicated that no single theoretical formulation has been able to justify various forms of masking [10]. Nevertheless, some simplified forms will still be useful for the design of image/video quality metrics and the improvement of image coding efficiency.

B. Algorithms using HVS features

One of the first attempts to use vision science concepts is given by Mannos and Sakrison in their famous paper [12] in 1974, where they use a CSF to weight the importance of different frequency components in an image. Nill [5] incorporate CSF to the use of DCT to improve image coding efficiency. Saghri et al. [6] refined the model and took into accounts the display device calibration, viewing distance and image resolution. A well-known method is by Dally [13]. His model is made of three parts: an amplitude non-linearity (accounts for luminance sensitivity mentioned above), a CSF and a hierarchy of detection mechanisms.

The Just-Noticeable-Distortion (JND) [14,15] is a very important concept in the literature. The ideal JND provides each signal being represented with a threshold level of error visibility, below which reconstruction errors are rendered imperceptible. The JND concept is adopted by a number of algorithms [10, 16, 17], including the Sarnoff Visual Discrimination Model (VDM) proposed by Lubin [17].

Many investigations on masking effect have been conducted to determine the luminance difference threshold close to the luminance edge [4, 18, 19]. Watson consider luminance masking and contrast masking separately in DCT domain [20,21].

The Gabor function is optimal for space-frequency localization. As a result, a number of multichannel visual models based on Gabor function have been proposed [22, 23, 24]. By using Gabor decomposition, the image can be represented by a set of subband images characterized by frequency and orientation. The CSF and masking effect then can be considered in each channel separately. An alternative decomposition method is to use a set of cosine log filters [25], which sum to one and are symmetry on a logarithmic scale in the frequency domain.

IV. IMPLEMENTATION

Most of the algorithms in the literature are designed for gray scale still images. Some of the metrics consider only some special types of distortions such as the blocking artifacts [4]. Some others are designed mainly for DCT based block coding methods [5, 6]. These choices make the algorithms simplified for real world implementations.

To generalize the algorithms for color image quality evaluation, the direct way is to apply the same model with different parameters to the three (RGB) color channels respectively and then weigh and combine the errors in different channels together. However, this direct way is not good in the sense that the RGB channels are correlated with each other. In [3, 7, 8], Wandell and Poirson suggested a new color space, namely opponent color space, where the principal coordinates are perceptually orthogonal. Their three coordinates of the opponent color space correspond to luminance (B/W), red-green (R/G), and blue-yellow (B/Y), respectively. The change of color space (from RGB, YUV or YCrCb into Opp space) makes it more reasonable to deal with each pathway separately. C. J. van den Branden Lambrecht adopted this new color space in his color image quality metric [23].

Very a few works have been done for the quality assessment of video sequences [24]. One of the major reasons is due to the computational complexity. For example, in a certain image quality assessment algorithm, if we need 2 minutes to get a quality value for a gray scale still image, then when we apply it directly to a color video sequence with 3 color channels and 30 frames per second, the overall compuation time will be 180 minutes for just 1 second of video! It is obviously unacceptable for real world applications. Therefore, a very fast video quality assessment system is needed. To make it practical, only a fast quality assessment algorithm is not enough. The information redundancy between color channels and frames should also be employed.

V. CONCLUSIONS

In this literature survey, we give a brief overview on the current state-of-the-art image/video quality assessment algorithms. It is shown that in order to design a good quality metric, incorporation of HVS features is necessary. Future work should not only refine the current algorithms to get better quality measurements, but also provide fast implementations for the algorithms. The latter is especially important for the real world application of video quality assessment.

REFERENCE

[1] B. Girod, “What’s wrong with Mean-Squared Error,” *Digital Images and Human Vision*, A. B. Watson Ed., Chapter 15, pp. 207-220, the MIT press, 1993.

[2] A. M. Eskicioglu and P. S. Fisher, “Image quality measures and their performance,” *IEEE Transactions on Communications*, vol. 43, no. 12, pp. 2959-2965, Dec. 1995.

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

[4] S. A. Karunasekera and N. G. Kingsbury, “A distortion measure for blocking artifacts in images based on human visual sensitivity,” *IEEE Transactions on Image Processing*, vol. 4, no. 6, pp. 713-724, June 1995.

[5] N. B. Nill, “A visual model weighted cosine transform for image compression and quality assessment,” *IEEE Transactions on Communications*, vol. 33, no. 6, pp. 551-557, June 1985.

[6] J. A. Saghri, “Image quality measure based on a human visual system model,” *Optical Engineering*, vol. 28, no. 7, pp. 813-818, July 1989.

[7] A. B. Poirson, and B. A. Wandell, "Appearance of colored patterns: pattern-color separability," *Journal of Optical Society of America A*, vol. 10, no. 12, pp. 2458-2470, Dec. 1993.

[8] A. B. Poirson, and B. A. Wandell, "Pattern-color separable pathways predict sensitivity to simple colored patterns," *Vision Research*, vol. 36, no. 4, pp. 515-526, 1996.

[9] E. C. Carterette and M. P. Friedman, Eds., *Handbook of Perception*, vol. 5, New York: Academic, 1975.

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

[11] A. N. Netravali and B. G. Haskell, *Digital Pictures: Representation and Compression*, New York: Plenum, 1988.

[12] J. L. Mannos, and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images," *IEEE Transactions on Information Theory*, vol. 20, no.4, pp. 525-536, 1974.

[13] S. Daly, "The visible differences predictor: an algorithm for the assessment of image fidelity," *Digital Images and Human Vision*, A. B. Watson Ed., Chapter 14, pp. 179-206, the MIT press, 1993.

[14] N. Jayant, "Signal compression: technology targets and research directions," *IEEE J. Select. Areas Communications*, vol. 10, pp. 314-323, June 1992.

[15] N. Jayant, J. Johnston, and R. Safranek, "Signal compression based on model of human perception," *Proceedings of the IEEE*, vol. 81, pp. 1385-1422, Oct. 1993.

[16] R. J. Safranek and J. D. Johnston, "A perceptually tuned subband image coder with image dependent quantization and post-quantizatioin data comperssion," *IEEE*

International Conference on Acoustics, Speech and Signal Processing, vol. 3, pp. 1945-1948, 1989.

[17] J. Lubin, “A visual discrimination model for imaging system design and evaluation,” *Vision Models for Target Detection and Recognition*, E. Peli ed., Chapter 10, pp. 245-283, World Scientific Publishing Co. Pte. Ltd., 1995.

[18] B. Girod, “Psychovisual aspects of image communication,” *Signal Processing*, vol. 28, no. 3, pp. 239-251, Sept. 1992.

[19] B. Girod, H. Almer, L. Bengsson, B. Christonsson, and P. Weiss, “A subjective evaluation of noise shaping quantization for adaptive intra interframe DPCM coding of color television signals,” *IEEE Transactions on Communications*, vol. 36, no. 3, pp. 332-346, Mar. 1988.

[20] A. B. Watson, “DCT quantization matrices visually optimize for individual images,” *Human Vision, Visual Processing and Digital Display IV*, Proc. SPIE, vol. 1913, pp. 202-216, 1993.

[21] A. M. Mayache, T. Eude, and H. Cherifi, “A comparison of image quality models and metrics based on human visual sensitivity,” *IEEE International Conference on Image Processing*,vol. 3, pp. 409-413, Oct. 1998.

[22] P. C. Teo and D. J. Heeger, “Perceptual image distortion,” *IEEE International Conference on Image Processing*, vol. 2, pp. 982-986, Austin, Nov. 1994.

[23] C. J. van den Branden Lambrecht, “Color moving pictures quality metric,” *IEEE International Conference on Image Processing*, vol. 1, pp. 885-888, 1996.

[24] Par Lindh, and C. J. van den Branden Lambrecht, “Efficient Spatio-Temporal Decomposition for Perceptual Processing of Video Sequences,” *IEEE International Conference on Image Processing*, vol. 3, pp.331-334, 1996.

[25] E. Peli, “Contrast in complex images,” *Journal of Optical Society of America*, vol. 7, no. 10, pp.2032-2040, Oct. 1990.