Survey of Objective Video Quality Measurements

Yubing Wang
wang_yubing@emc.com
EMC Corporation Hopkinton, MA 01748, USA

1. Introduction

My thesis models feedback-based error control schemes. Initially I proposed to use Peak-Signal-to-Noise-Ratio (PSNR) as video quality metric because of its simplicity. However, it is well-known that PSNR does not necessarily accurately model perceptual quality. This document surveys video quality measurement approaches proposed in the recent years. This survey provides a brief description of each video quality metric and the main video characteristics it measures. In particular, this survey compares each quality metric with PSNR to see any potential advantages of using that metric over PSNR for our analytical models. The video quality metrics in this survey include PSNR, VQM, MPQM, SSIM and NQM. This survey concludes with a comparison of these metrics in terms of computational complexity, correlation with subjective video quality measurement, and accessibility.

2. Peak-Signal-to-Noise-Ratio (PSNR)

PSNR is derived by setting the mean squared error (MSE) in relation to the maximum possible value of the luminance (for a typical 8-bit value this is $2^8-1 = 255$) as follows:

$$MSE = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left[(f(i,j) - F(i,j)\right]^2}{M \cdot N}$$

$$PSNR = 20 \cdot \log_{10}\left(\frac{255}{\sqrt{MSE}}\right)$$

Where f(i,j) is the original signal at pixel (i, j), F(i, j) is the reconstructed signal, and M x N is the picture size. The result is a single number in decibels, ranging from 30 to 40 for medium to high quality video.

Despite several objective video quality models have been developed in the past two decades, PSNR continues to be the most popular evaluation of the quality difference among pictures.

3. Video Quality Metric (VQM)

VQM [1] is developed by ITS[1] to provide an objective measurement for perceived video quality. It measures the perceptual effects of video impairments including blurring, jerky/unnatural motion, global noise, block distortion and color distortion, and combines them into a single metric. The testing results show VQM has a high

[1] The Institute for Telecommunication Science

correlation with subjective video quality assessment and has been adopted by ANSI as an objective video quality standard.

VQM takes the original video and the processed video as input and is computed as follows:

- Calibration
  This step calibrates the sampled video in preparation for feature extraction. It estimates and corrects the spatial and temporal shift as well as the contrast and brightness offset of the processed video sequence with respect to the original video sequence.
- Quality Features Extraction
  This step extracts a set of quality features that characterizes perceptual changes in the spatial, temporal, and chrominance properties from spatial-temporal sub-regions of video streams using a mathematical function.
- Quality Parameters Calculation
  This step computes a set of quality parameters that describe perceptual changes in video quality by comparing features extracted from the processed video with those extracted from the original video.
- VQM Calculation
  VQM is computed using a linear combination of parameters calculated from previous steps.

VQM can be computed using various models based on certain optimization criteria. These models include (1) Television (2) Videoconferencing (3) General (4) Developer (5) PSNR.

The general model uses a linear combination of seven parameters. Four parameters are based on features extracted from spatial gradients of the Y luminance component, two parameters are based on features extracted from the vector formed by the two chrominance components (CB, CR), and one parameter is based on contrast and absolute temporal information features, both extracted from the Y luminance component.

Figure 1 illustrates a summary of the test results from eleven experiments during 1992-1999 that were performed by Wolf and Pinson. The results show a high correlation coefficient of 0.95 between subjective tests and the VQM general model (VQMG).

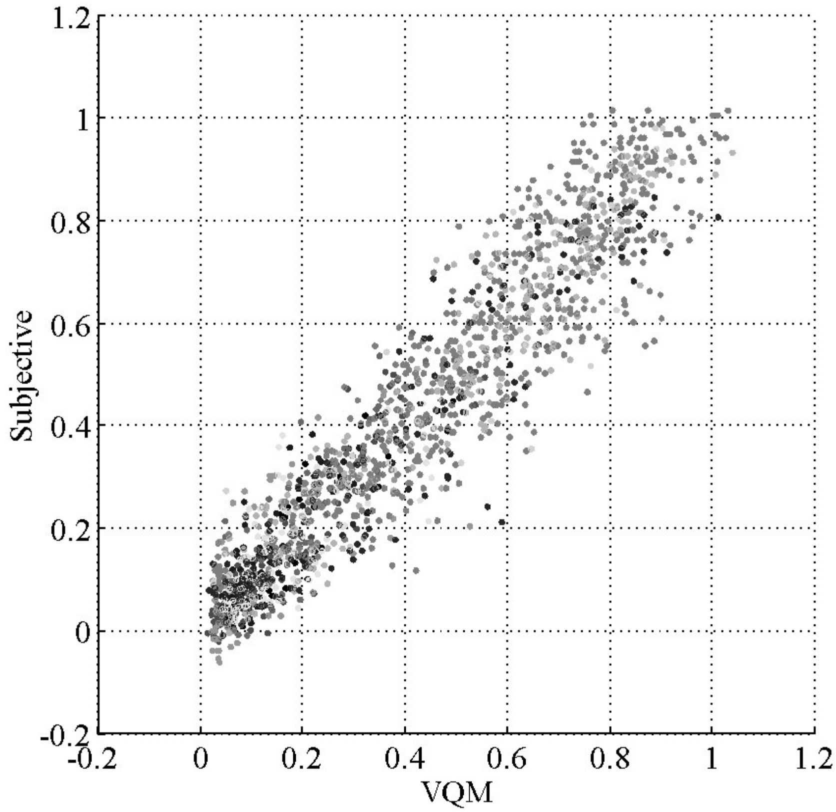

Figure 1. Clip subjective quality vs. clip VQM_G (Wolf et al, 2002).

4. Moving Pictures Quality Metric (MPQM)[2]

PSNR does not take the visual masking phenomenon into consideration. In other words, every single pixel error contributes to the decrease of the PSNR, even if this error is not perceived. This issue is addressed by means of incorporating some modeling of the Human Visual System. In particular, two key human perception phenomenon that have been intensively studied: contrast sensitivity and masking. The first phenomenon accounts for the fact that a signal is detected by the eye only if its contrast is greater that some threshold. The eye sensitivity varies as a function of spatial frequency, orientation and temporal frequency. The second phenomenon is related to the human vision response to the combination of several signals. A stimulus consists of two types of signals (foreground and background). The detection threshold of the foreground will be modified as a function of the contrast of the background.

MPQM is an objective quality metric for moving picture which incorporates two human vision characteristics as mentioned above. It first decomposes an original sequence and a distorted version of it into perceptual channels. A channel-based distortion measure is then computed, accounting for contrast sensitivity and masking. Finally, the data is pooled over all the channels to compute the quality rating which is then scaled from 1 to 5 (from bad to excellent).

MPQM does not take into consideration the chrominance and that is why the method Color MPQM (CMPQM) has been introduced. The first step is to convert the color components to RGB values that are linear with luminance. Then the RGB values are converted to coordinate values that correspond to luminance (B/W), red-green (R/G) and blue-yellow (B/Y) channels. Then each component of the original and error sequence is analyzed by a filter bank. The B/W is processed as the luminance but as R/G and B/Y are much less sensitive only nine spatial and one temporal filter is used for these signals. The rest of the calculation is the same as for MPQM.

MPQM represents the typical image quality assessment models based on the error sensitivity. The widely adopted assumption of these models is that the loss of perceptual quality is directly related to the visibility of the error signal. Most perceptual image quality assessment approaches proposed in the literature attempt to weight different aspects of the error signal according to their visibility, as determined by psychophysical measurements in humans or physiological measurements in animals [6]. The underlying principle of the error-sensitivity approach is that perceptual quality is best estimated by quantifying the visibility of errors. This is essentially accomplished by simulating the functional properties of early stages of the HVS, as characterized by both psychophysical and physiological experiments. Although this bottom-up approach to the problem has found nearly universal acceptance, it is important to recognize its limitations. In particular, the HVS is a complex and highly nonlinear system, but most models of early vision are based on linear or quasilinear operators that have been characterized using restricted and simplistic stimuli [6].

MPQM is known to give good correlation with subjective tests for some material but give bad results for other according to studies conducted by Mohammed [3].

5. Structural Similarity Index (SSIM)

A different approach for video quality assessment is presented by Zhou Wang [4]. This method differs from the previously described methods, which all are error based, by using the structural distortion measurement instead of the error. The idea behind this is that the human vision system is highly specialized in extracting structural information from the viewing field and it is not specialized in extracting the errors. Thus, a measurement on structural distortion should give a better correlation to the subjective impression.

Many different quality assessment methods can be developed from this assumption but Wang proposes a simple but effective index algorithm [5][6]. If you let $x = \{x_i \mid i = 1,2,\ldots,N\}$ be the original signal and $y = \{y_i \mid i = 1,2,\ldots,N\}$ be the distorted signal the structural similarity index can be calculated as:

$$SSIM = \frac{(2\overline{xy} + C_1)(2\sigma_{xy} + C_2)}{\left[(\bar{x})^2 + (\bar{y})^2 + C_1\right](\sigma_x^2 + \sigma_y^2 + C_2)}$$

In this equation $\bar{x}, \bar{y}, \sigma_x, \sigma_y, \sigma_{xy}$ are the estimates of the mean of *x*, mean of *y*, the variance of *x*, the variance of *y* and the covariance of *x* and *y*. C_1 and C_2 are constants. The value of *SSIM* is between -1 and 1 and gets the best value of 1 if *xi* = *yi* for all values of *i*. The quality index is applied to every image using a sliding window with 11 x11 circular-symmetric Gaussian weighting function for which the quality index is calculated and the total index of the image is the average of all the quality indexes of the image.

According to tests made by Wang the correlation between SSIM and subjective impression are good as shown in figure 2. More experiments are needed to improve and fully test the system.

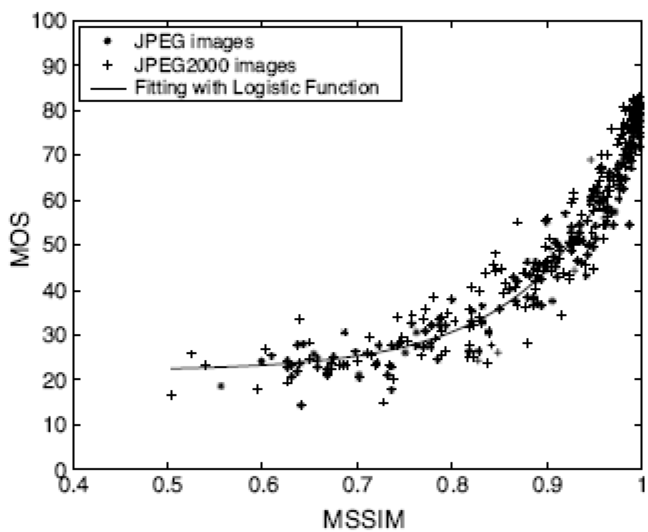

Figure 2. Correlation between SSIM and MOS subjective impression.

6. Noise Quality Measure (NQM) [7]

In this quality measurement metric, a degraded image is modeled as an original image that has been modeled as an original image that has been subjective to linear frequency distortion and additive noise injection. These two sources of degradation are considered independent and are decoupled into two quality measures: a distortion measure (DM) of the effect of frequency distortion, and a noise quality measure (NQM) of the effect of additive noise. The NQM takes into account : (1) variation in contrast sensitivity with distance, image dimensions; (2) variation in the local luminance mean; (3) contrast interaction between spatial frequencies; (4) contrast masking effects. The DM is computed in three steps. First, the frequency distortion in the degraded image is found. Second, the deviation of this frequency distortion from an all-pass response unity gain is computed. Finally, the deviation is weighted by a model of the frequency response of the human visual system.

A quality metric based on the two measures (NQM and DM) is yet to be defined.

7. Comparison

Quality Metric	Mathematical Complexity	Correlation with Subj. Methods	Accessibility
PSNR	Simple	Poor	Easy
MPQM	Complex	Varying	Not Available
VQM	Very Complex	Good	Not Available
SSIM	Complex	Fairly good	Available (MATLAB)
NQM	Complex	Unknown	Not Available

Quite a few alternative models have been proposed but none of them have been commonly accepted. PSNR is still being widely used in literature. A comparison of the objective metrics is presented in the table above. From this table it can be seen that if only one evaluation method is used the VQM or the SSIM give the most reliable result. However the computational complexities of these two methods make them difficult to apply to real-time applications (such as video conferencing). Moreover, the test results for SSIM are based on still images yet the performance of this metric on the video sequence remains unknown. The varying results on MPQM also make it difficult to recommend it for the purpose of my thesis. Furthermore, the previous published test results (e.g., [8], [9]) showed that the performance of most objective video quality models are statistically equivalent to root mean squared error [8] and PSNR [9]. This may explain why PSNR is still the most commonly used metric in literature. For this reason, I will still use PSNR as a baseline video quality metric for my models. However, I will extend our models to incorporate VQM. My hypothesis is that the relationships uncovered by the models using PSNR would be approximately equivalent to those revealed by the models using VQM.

Reference:

[1] M. Pinson and S. Wolf. "A New Standardized Method for Objectively Measuring Video Quality," IEEE Transactions on Broadcasting, VOL. 50, NO.3, pp. 312-322, September, 2004. [cited by 114]

[2] C. J. Branden Lambrecht and O. Verscheure. "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System," Proc. SPIE Vol. 2668, p. 450-461, March, 1996. [cited by 96]

[3] S. A. Mohamed. http://www.irisa.fr/armor/lesmembres/Mohamed/Thesis.pdf, 2003

[4] Z. Wang, L. Lu, A. C. Bovic. "Video quality assessment using structural distortion measurement," Signal Processing: Image Communication, special issue on "Objective video quality metrics", vol. 19, no. 2, pp. 121-132, February 2004. [cited by 17]

[5] Z. Wang and A. C. Bovic. A universal image quality index. IEEE Signal Processing Letters, vol. 9, pp. 81-84, March 2002. [cited by 106]

[6] Z. Wang, A. C. Bovik, H. R. Sheikh and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," IEEE Transactions on Image Processing, vol. 13, no. 4, pp. 600-612, Apr. 2004.

[7] D. Niranjan et al. “Image Quality Assessment Based on a Degradation Model,” IEEE Transaction on Image Processing, VOL. 9, NO. 4, April 2000. [cited by 53]

[8] J.-B. Martens and L. Meesters, “Image dissimilarity,” *Signal Processing*, vol. 70, pp. 155–176, Nov. 1998.

[9] VQEG. (2000, Mar.) Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. [Online] Available: http://www.vqeg.org/