Efficient Models for Objective Video Quality Assessment

Radim JAVŮREK

Dept. of Radio Electronics, Czech Technical University, Technická 2, 166 27 Praha, Czech Republic

javurer@fel.cvut.cz

Abstract. *Two models for objective assessment of compressed video quality and results from subjective and objective tests of several low bit-rate video coders are introduced in this paper. First model is based on mathematical measures that describe perception of video distortion by human eye. Second model for quality evaluation is based on human visual system (HVS) characteristics. Obtained quality values from both models have been compared with subjective results. Both models are computationally efficient and produce results that are correlated with subjective results.*

Keywords

Video quality evaluation, video compression, objective assessment, HVS.

1. Introduction

Video compression schemes produce artefacts whose visibility strongly depends on actual image content. When first compression methods were introduced it has become very important to devise quality assessment algorithms. For video quality assessment objective or subjective methods are used.

The subjective measurement Mean Opinion Score (MOS) is a widely used method on the assessment of image or video quality, but it has several obvious disadvantages. It is very tedious, expensive and impossible to be executed automatically. Instead, an objective image or video quality metric can provide a quality value for a given image or video automatically in a relatively short time. This is very important for real world applications.

All types of objective video quality assessments methods are based on measurement of differences between the original video and received/degraded video sequence in some way. The easiest objective quality measures are some simple statistics features on the numerical error between the reference and the distorted image. Widely used statistics are Mean Squared Error (MSE) and Peak Signal to Noise Ratio (PSNR). However, MSE and PSNR do not correlate well with subjective quality measures because human of perception distortions and artefacts is unaccounted for.

Systems for objective quality measurement can use measures that correlate with perceptual distortion or can use Human Visual System (HVS) features. Widely used models are Tektronix/Sarnoff, NASA – DVQ Digital Video Quality Model and EPFL – PDM Perceptual Distortion Metric.

This paper introduces two objective video quality assessment models for quality assessment of compressed video and results from subjective and objective tests of low-bitrate video coders.

2. Models for Quality Assessment

2.1 Model Based on Mathematical Measures

First system for video quality evaluation uses mathematical measures that correlate with perception of distortion (Fig. 1). These mathematical characteristics were selected by ITS in Colorado [3] from the quality features and parameters suggested by American National Standard Institute, ANSI. Final quality metric is linear combination of three quality impairment measures. Impairment measures are based upon two quantities, namely spatial information and temporal information. The spatial information feature is based on the Sobel filter. The frame is filtered at defined time with Sobel operator. The standard deviation over pixels is then computed. This operation is repeated for each frame of the video sequence. The temporal information feature is based upon the difference image, which is composed of the differences between pixel values at the same location in space but at different time or frames. The temporal information feature is defined as the standard deviation of difference images over the horizontal and vertical spatial dimensions.

Three measures are defined which are to be linearly combined to get the final quality measure. The constants for linear combination were found by applying least squares error criteria to a training set that was composed of 10 sequences. First measurement is a measure of spatial distortion and is given by the relative change in SI between the original and degraded video. It describes blurring and false edges.

Second and third measurements are both measures of temporal distortion. Second measure describes effect of temporally localized motion in the degraded video that were no in the original video and is non-zero only when degraded sequence has lost motion energy with respect to the original sequence.

Third measurement selects the frame that has the largest added motion. It should be frame with maximal jerky motion or with the worst uncorrelated block errors.

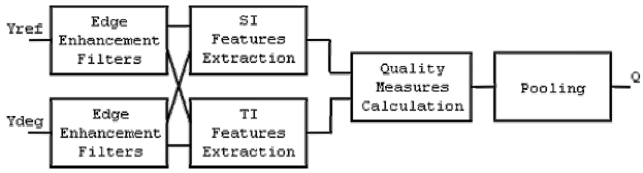

Fig. 1. Model based on mathematical measures.

2.2 Model Based on HVS Characteristics

Second model for video quality evaluation is shown on Fig. 2. It simulates some stages of human visual system HVS. The inputs for this model are test and reference video sequences. The first step is decomposition of sequence into different spatial channels. This step is implemented by using of discrete cosine transform DCT. The result is converted to local contrast LC. It is the ratio of DCT coefficients to DC amplitude for the corresponding block. Next two blocks are connected with contrast sensitivity function CSF. First block is temporal filtering which implements temporal part of CSF. In next block DCT coefficients are converted to just-noticeable difference (jnd) by dividing each DCT coefficient by its respective visual threshold. At the next stage two sequences are subtracted and mean and maximum distortion is calculated. Mean and maximum distortions are weighted and pooled to implement masking operation. The result is quality of video sequence.

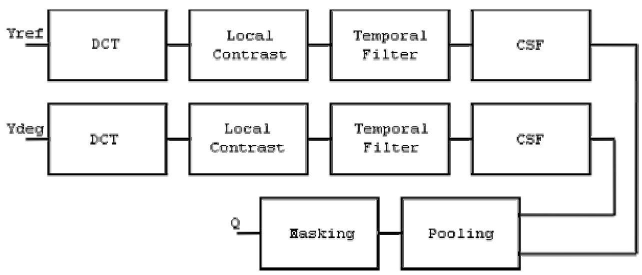

Fig. 2. Model based on HVS characteristics.

3. Models Evaluation

For quality tests of the models we have used five test sequences from the set of EBU test sequences. These sequences are in ITU-R 601 format and are 10 seconds in duration. We have tested 6 widely used compression methods (DivX, XviD, Quick Time, Windows Media, MP2 Tsunami and VP6). For each tested compression method we used 5 compression levels at 4096 kbit/s, 2048 kbit/s, 1024 kbit/s, 512 kbit/s and 256 kbit/s and we obtained set of 150 degraded sequences. As a reference values subjective rating of 14 observers were used. Subjective assessment was based on ITU-R Recommendation BT.500. We used Double Stimulus Continuous Quality Scale (DSCQS) method. In this method each trial consists of a pair of stimuli, the reference and the test sequence. The stimuli are each presented twice in trial and with randomly chosen order. The subjects rate each stimulus on a continuous quality scale.

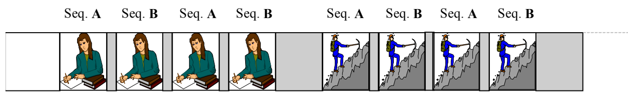

Fig. 3. Test sequence sample.

Performance of the objective models was evaluated with respect to two aspects of their ability to estimate subjective assessment of video quality according the VQEG [2]. These attributes are accuracy and consistency and are defined as follows:

3.1 Prediction Accuracy

Prediction accuracy is the ability of the metric to predict subjective rating with minimum average error and can be determined by means of Pearson linear correlation coefficient, which is defined as:

$$r_p = \frac{\sum (X - \mu_X)(Y - \mu_Y)}{N\sigma_X\sigma_Y} \tag{1}$$

where X and Y are data vectors, μ_X, μ_Y and σ_X, σ_Y are means resp. standard deviations of respective data vectors.

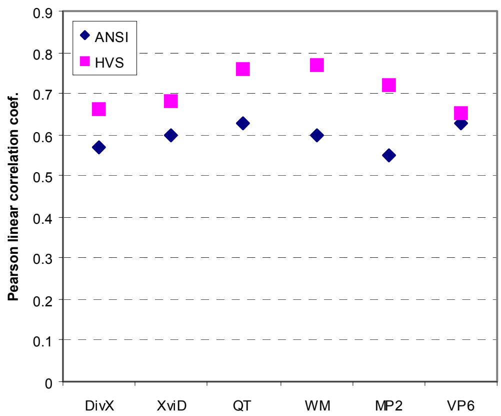

Fig. 4. Pearson's correlation coefficient between subjective and objective results for 6 video coders describing prediction accuracy.

3.2 Prediction Consistency

Prediction consistency of a model prediction can be measured by number of outliers. An outlier is defined as a data point for which the prediction error is greater than a certain threshold. In our case the threshold is twice the standard deviation σ of the subjective rating differences for this data point.

$$\left|x_i - y_i\right| > 2\sigma_{sr} \ . \qquad (2)$$

The outlier ratio is defined as the number of outliers determined this fashion in relation to the number of data points

$$r_O = N_O / N \ . \qquad (3)$$

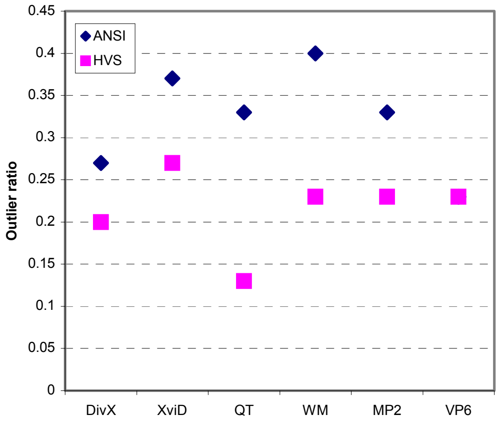

Fig. 5. Outlier ratio for 6 video coders describing prediction consistency.

4. Conclusion

Two models for objective video quality assessment have been presented in this paper. Both models and subjective Double Stimulus Continuous Quality Scale (DSCQS) method have been used for quality tests at our department. Obtained data is correlated and it is possible to use proposed objective models, which are faster and cheaper, instead subjective methods for quality tests.

Acknowledgements

This work has been conducted at the Department of Radioelectronics, Faculty of Electrical Engineering, Czech Technical University in Prague in the frame of the research project "Qualitative Aspects of Image Compression Methods in Multimedia Systems" and has been supported by the grant No. 102/02/0133 of the Grant Agency of Czech Republic.

A part of this research work has been partially supported by the GAČR grant No. 102/03/H109 "Methods, Structures and Components of Electronic Wireless Communication".

References

[1] WINKLER, S. A perceptual distortion metric for digital color video. In *Proc. SPIE*, vol. 3644, pp. 175-184, San Jose (CA), 1999.

[2] VQEG: Final report from the Video Quality Experts Group on the validation of objective models of video quality assessment, 2000.

[3] WEBSTER, A. A., JONES, C. T., PINSON, M.H., VORAN, S. D., WOLF, S. An objective video quality assessment system based on human perception. In *SPIE Human Vision, Visual Processing and digital display IV*, vol. 1913, pp. 15-26, San Jose (CA), February 1993.

[4] ITU-R Recommendation BT.500-11: Methodology for the subjective assessment of the quality of television pictures, International Telecommunication Union, Geneva, Switzerland, 2002.

[5] LAMBRECHT, C. J. VAN DEN BRANDEN. Perceptual Models and Architectures for Video Coding Applications. *Ph.D. thesis*, Ecole Polytechnique Federale de Lausanne, Switzerland, 1996.

[6] WANG, Z., LU, L., BOVIK, A.C. Video quality assessment based on structural distortion measurement, submitted to *Signal Processing: Image Communication, special issue on "Objective video quality metrics"*, 2003.

[7] WATSON, A.B., HU, J., McGOWAN, J. F. III. DVQ: A digital video quality metric based on human vision, *Journal of Electronic Imaging*, 2001, vol. 10, no. 1, pp. 20–29.

[8] DALY, S. The visible differences predictor: An algorithm for the assessment of image fidelity. In: A. B. Watson (ed.), *Digital Images and Human Vision*, pp. 179–206, MIT Press, 1993.

About Author...

Radim JAVŮREK was born in 1977 in Prague, Czech Republic. He received the Ing. (M.Sc.) degree in electrical engineering from the Faculty of Electrical Engineering, Czech Technical University in Prague (CTU FEE) in 2002. At present he is a Ph.D. student at the Department of Radioelectronics CTU FEE in Prague.