A Novel Method for Pre-Encoding Video Quality Prediction

Harilaos Koumaras, Thomas Pliakas, Anastasios Kourtis

Abstract— **This paper presents a novel method for fast and quantified estimation of the video quality for H.264 video content. Taking into account the instant video quality variation due to the video dynamics within a given H.264 encoded content, this paper introduces the Mean PQoS as a function of the video encoding rate and the picture resolution, and exploits it as a metric for objective video quality assessment. Based on the proposed metric, the paper presents a novel method for pre-encoding video quality prediction.**

Index Terms—**Video Quality, MPEG, SSIM**

I. INTRODUCTION

CURRENTLY, the evaluation of the PQoS is a matter of objective and subjective evaluation procedures, each time taking place after the encoding process (post-encoding evaluation).

The subjective test methods, which have mainly been proposed by International Telecommunications Union (ITU) and Video Quality Experts Group (VQEG), involve an audience of people, who watch a video sequence and score its quality as perceived by them, under specific and controlled watching conditions. Afterwards, the statistical analysis of the collected data is used for the evaluation of the perceived quality. The Mean Opinion Score (MOS) is regarded as the most reliable method of quality measurement and has been applied on the most known subjective techniques.

Subjective picture/audio quality evaluation processes require large amount of human resources, establishing it as a time-consuming process (e.g. large audiences evaluating video/audio sequences). Objective evaluation methods, on the other hand, can provide video quality evaluation results faster, but require large amount of machine resources and sophisticated apparatus configurations. Towards this, objective evaluation methods are based and make use of multiple metrics, which are related to the content's artifacts (i.e. tilling, blurriness, error blocks, etc.) resulting during an encoding process.

The majority of the proposed objective methods in the literature require the undistorted source video sequence as a reference entity in the quality evaluation process. Due to this reason are characterized as Full Reference (FR) Methods [1-3]. However it has been reported that these complicated FR methods do not provide more accurate results than the simple mathematical error frameworks (such as PSNR). Due to this some new full reference metrics that are based on the video structural distortion, and not on error measurement, have been proposed [1].

On the other hand, the fact that these methods require the original video signal as reference deprives their use in commercial video service applications, where the initial undistorted clips are not accessible. Moreover, even if the reference clip is available, then synchronization predicaments between the undistorted and the distorted signal (which may have experienced frame loss) make the implementation of the FR Methods difficult and impractical.

Due to these reasons, the recent research has been focused on developing methods that can evaluate the PQoS level based on metrics, which use only some extracted structural features from the original signal (Reduced Reference Methods) [4] or do not require any reference video signal (No Reference Methods). The NR methods can be classified into two classes: The NR-pixel based and the NR-bitstream based. The first methods must initially decode the bit stream and estimate the video quality at the pixel/visual layer [5], while the second ones can evaluate the perceived quality by receiving as input the compressed bit stream, without requiring any decoding.

However, neither these methods are appropriate for commercial implementations by a content provider, who must prepare encoded video signals at various quality levels. All, the existing methods require numerous encodings and post-assessments in order to define the bit rate that corresponds to a given quality level, for given content and encoding parameters (i.e. spatial and temporal resolution). Thus, the issue of developing methods for predicting the video quality at a pre-encoding stage [7-9] is still open and under research.

This paper presents a method for video quality prediction of MPEG-based sequences. The rest of the paper is organized as follows: Next section presents a theoretical approach of video quality vs. bit rate curves. Section III demonstrates the proposed video quality prediction algorithm. Finally, Section IV discusses the limitations of the proposed method, while Section V concludes the paper.

II. MODELING VIDEO QUALITY VS. BIT RATE

In digital video encoding the Block Discrete Cosine

Transformation (BDCT) is exploited, since it exhibits very good energy compaction and de-correlation properties.

The DCT operates on a X block of NxN image samples or residual values after prediction and creates Y, which is a NxN block of coefficients. The action of the DCT can be described in terms of a transform matrix A. The forward DCT is given by:

$$Y = AXA^T \quad (1)$$

where X is a matrix of samples, Y is a matrix of coefficients and A is an N X N transform matrix. The elements of A are:

$$A_{ij} = C_i \cos \frac{(2j+1)i\pi}{2N} \quad (2)$$

$$C_i = \begin{cases} \sqrt{1/N}, i=0 \\ \sqrt{2/N}, i>0 \end{cases} \quad (3)$$

Therefore the DCT can be written as:

$$Y_{xy} = C_x C_y \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} X_{ij} \cos \frac{(2j+1)y\pi}{2N} \cos \frac{(2i+1)x\pi}{2N} \quad (4)$$

The advantage of the DCT transformation is that it is possible to reconstruct quite satisfactorily the original image, applying the reverse DCT on a subset of the DCT coefficients, without taking under consideration the rest coefficients with insignificant magnitudes (see figure 1).
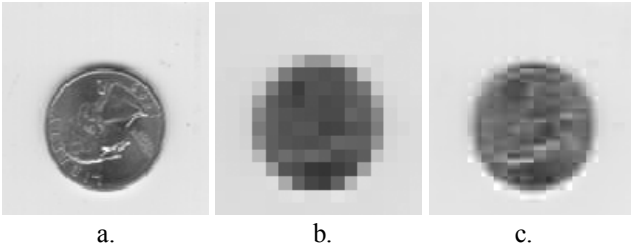

| a. | b. | c. |

Fig.1. Figure a is the source image, while b is reconstructed using only 1 DCT coefficient. Image c uses 4 coefficients out of the total 64 DCT Coefficients for each block (i.e. 8x8).

Thus, with cost of some video quality degradation, the original frame can be satisfactorily reconstructed with a reduced number of coefficient values.

A latest perceived quality metric, which provides very reliable assessment of the video quality is the SSIM metric [1], [3]. The SSIM is a FR metric for measuring the structural similarity between two image/video sequences, exploiting the general principle that the main function of the human visual system is the extraction of structural information from the viewing field. The concept of averaging the SSIM for the whole video duration can be exploited for deriving a single perceived quality measurement, which is more practical, especially for the service providers.

Fig. 2. The five used reference clips in the evaluation process

For the experimental needs of this paper, each reference clip of Fig.2 was transcoded from its original uncompressed format to ISO H.264 Baseline Profile, at different VBR bit rates. For each corresponding bit rate, a different H.264 compliant file with CIF (Common Interface Format) resolution (352x288) was created. The frame rate was set at 25 frames per second (fps) for the transcoding process in all test videos.
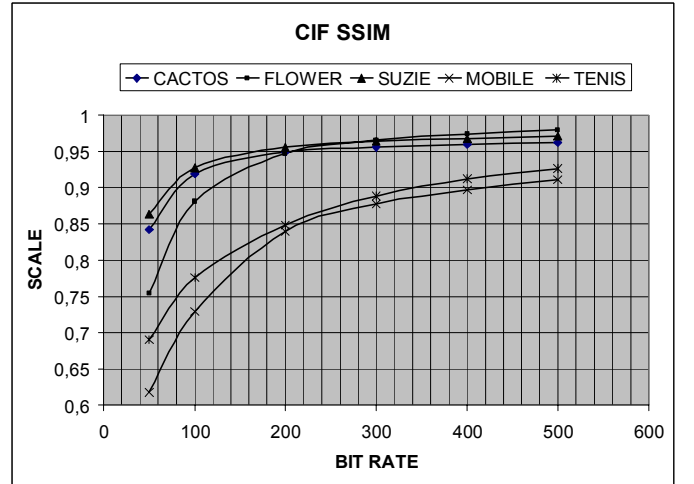

Fig. 3. The <PQoS>$_{SSIM}$ vs. bit rate (kbps) curves for various test signals

Each H.264 video clip was then used as input in the SSIM estimation algorithm. From the resulting SSIM vs. time graph, the average <PQoS> value of each clip was calculated. This experimental procedure was repeated for each video clip in CIF resolution. The results of these experiments are depicted in Figure 3.

Thus, any <PQoS>$_{SSIM}$ vs. bit rate curve can be successfully described by a logarithmic function of the general form

$$< PQoS >_{SSIM} = C_1 \ln(BitRate) + C_2 \quad (5)$$

where C_1 and C_2 are constants strongly related to the spatial and temporal activity level of the content. Table I depicts the corresponding logarithmic functions for the test signals of Figure 2 along with their R^2 factor, which denotes the fitting efficiency of the theoretical graph to the experimental one.

TABLE I
FITTING PARAMETERS AND R^2 FOR THE TEST SIGNALS

Test Signal	Logarithmic Function	R^2 factor
Cactos	$0.0490\ln(x)+0.6719$	0.8593
Mobile&Calendar	$0.1295\ln(x)+0.1274$	0.9759
Flower Garden	$0.0947\ln(x)+0.4163$	0.9979
Table Tennis	$0.1033\ln(x)+0.2940$	0.9938
Suzie	$0.0443\ln(x)+0.7075$	0.8901

Based on the aforementioned analysis, we can describe the derived $<PQoS>_{SSIM}$ vs. bit rate curve of each test signal with N total frames, which is encoded at bit rate n from $BitRate_{min}$ to $BitRate_{max}$ as a set C, where each element F_n is a triplet, consisting the $<PQoS>_{SSIM}$ of the specific bit rate the constant C_1 and C_2, which is derived by the analytical logarithmic expression of Table I:

$$C_{S-T} \triangleq \{m : (\frac{1}{N}\sum_{i=1}^{N} SSIM(f_i), C_1, C_2)_n = F_n, n \in [BitRate_{min}, BitRate_{max}]\} \ (6)$$

where the following notation has been followed:

-SSIM( ) is the function that calculates the perceived quality of each frame according to the SSIM metric

-N the total number of frames f_i that consist the movie m

Thus, deriving the sets C_n for various contents, ranging from static to very high spatial and temporal ones, we define a reference hyper-set RS, containing various C_n sets for specific spatiotemporal levels can be deduced:

$$RS = \{C_{S-T_{Low}},...,C_{S-T_{High}}\} \tag{7}$$

A hypothetical service provider, should retain a database of this hyper set RS, which will contain a great range of sets C_n, coming from reference video clips with various content dynamics. The longer the database, the more accurate will be the proposed method of the next Section for predicting pre-encodingly the video quality of a MPEG-based sequence at a specific bit rate.

III. PREDICTING VIDEO QUALITY AT SPECIFIC BIT RATE

Consider an unknown video clip, which is uncompressed and for which we want to predict its corresponding C_{S-T} set that better describes its perceived quality vs. bit rate curve before the actual encoding process takes place. Then, we define for all the sets C_{S-T} the Absolute Difference Value (ADV) between the first C_{S-T} triplet element (i.e. the $<PQoS>_{SSIM}$ at a specific encoding Bit Rate) and the experimental measurement of the average SSIM for the test signal at the same encoding bit rate n:

$$ADV = | F_{BitRate_i} : (\frac{1}{N}\sum_{i=1}^{N} SSIM(f_i)) - F'_{BitRate_i} : (\sum_{i=1}^{N} SSIM(f'_i)) | \ (8)$$

Due to the fact that the additive property is valid, it is concluded that when the ADV between the reference $F_{BitRate_i}$ and experimental $F_{BitRate_i}$ average SSIM is minimum for a specific set C_{S-T}, then this set contains the triplet element that minimizes the ADV, describes better the specific video.

Thus, we have successfully approximated the PQoS vs. Bit rate curve of the specific video with actual cost only one testing encoding at bit rate n.

IV. FUTURE WORK

This paper has examined the case of video quality assessment due to the encoding process, proposing a novel and theoretical framework for video quality prediction pre-encodingly. Regarding video quality in this paper, we consider that the defined encoding quality is not degraded during the transmission/streaming process and it is finally delivered to the end-user.

However, the case of the perceived quality degradation of the video quality due to network problems (e.g. packet loss) during the streaming process has not been examined, because it is out of the scope of this paper. The authors already work on developing a theoretical framework, which will map the network QoS sensitive parameters to video quality degradation.

By integrating this under-development framework with the current one, a service provider will be able to perform an end-to-end video quality prediction, based on the selected encoding parameters and the condition of the network status.

V. CONCLUSION

This paper has presented a novel method for fast and quantified estimation of the video quality for H.264 video content. Taking into account the instant video quality variation due to the video dynamics within a given H.264 encoded content, this paper introduced the Mean PQoS as a function of the video encoding rate and the picture resolution, exploiting it as a metric for objective video quality assessment. Based on the proposed metric, the paper presents a novel method for pre-encoding video quality prediction.

REFERENCES

[1] Wang, Z., H.R. Sheikh, and A.C. Bovik, Objective video quality assessment, in The Handbook of Video Databases: Design and Applications, B. Furht and O. Marqure, Editors. 2003, CRC Press. p. 1041-1078.

[2] VQEG. Final Report From the Video Quality Experts Group on the Validation of Objective Models of Video Quality Assessment. 2000 [cited; Available from: http://www.vqeg.org.

[3] Wang, Z., A.C. Bovik, and L. Lu. Why is image quality assessment so difficult? in IEEE International Conference on Acoustics, Speech, and Signal Processing. 2002.

[4] Gunawan, I.P. and M. Ghanbari. Reduced-Reference Picture Quality Estimation by Using Local Harmonic Amplitude Information. in London Communications Symposium 2003. 2003.

[5] Lu, L., et al. Full-reference video quality assessment considering structural distortion and no-reference quality evaluation of MPEG video. in IEEE International Conference on Multimedia. 2002.

[6] H. Koumaras, A. Kourtis, D. Martakos, "Evaluation of Video Quality Based on Objectively Estimated Metric", Journal of Communications and Networking, Korean Institute of Communications Sciences (KICS), Vol. 7, No.3, pp.235-242, Sep 2005 (available online http://citeseer.comp.nus.edu.sg/747634.html)

[7] H. Koumaras, A. Kourtis, D. Martakos, J. Lauterjung, "Quantified PQoS Assessment Based on Fast Estimation of the Spatial and Temporal Activity Level", Multimedia Tools and Applications, Springer Editions, accepted for publication.

[8] H. Koumaras, E. Pallis, G. Xilouris, A. Kourtis, D. Martakos, J. Lauterjung, "Pre-Encoding PQoS Assessment Method for Optimized Resource Utilization", 2nd Inter. Conference on Performance Modelling and Evaluation of Heterogeneous Networks, Het-NeTs04, Ilkley, United Kingdom, 2004.