An Experimental Approach of Video Quality Level Dependence on Video Content Dynamics

Harilaos Koumaras
Business College of Athens (BCA)
NCSR Demokritos
Athens, Greece
+30 210 650 3107
koumaras@iit.demokritos.gr

Julien Arnaud, Daniel Negru
CNRS – LaBRI Lab.,
University of Bordeaux
Talence, France
+33 5 40 00 3797
{arnaud, negru}@labri.fr

Anastasios Kourtis
NCSR Demokritos
Inst. of Informatics and Telecom.
Aghia Paraskevi, Greece
+30 210 650 3166
kourtis@iit.demokritos.gr

ABSTRACT

This paper deals with the notion of user satisfaction relative to the consumption of modern encoded video applications and services. Due to the process of encoding/compression of a video signal, respective quality degradation takes place, which in turn introduces the need for quality assessment methods and procedures. The objective of this paper is to research the impact of the spatiotemporal dynamics of the video content on the deduced perceptual quality. More specifically it is presented how the spatiotemporal activity affects i) the highest quality level that each video can reach, ii) the video quality acceptance threshold such as the lowest quality level and iii) the video quality vs. bit rate pattern.

General Terms

Algorithms, Human Factors.

Keywords

PQoS, Video quality, video dynamics, MPEG-4, H.264

1. INTRODUCTION

Current modern technology has made very popular the wide production, distribution and consumption of video data over the Internet and mobile communication networks. Although the capacity of the various access and core networks has today reached levels that may leave the opportunity for over provisioning, the use of encoding techniques for the compression of video streams remains a necessity in order to reduce the high multimedia data volume in datacenters. Thus, the evaluation of the respective quality degradation introduced by the compression process still remains as an active research topic.

The existing literature of video quality assessment techniques focuses on models and techniques evaluating and assessing the perceptual level of an already encoded and/or served video service. Currently the evaluation of the video quality is a matter of subjective and objective procedures both applied on the encoded signal. The subjective test methods, mainly proposed by International Telecommunications Union (ITU) and Video Quality Experts Group (VQEG), involve an audience watching a video sequence and scoring its quality as perceived by the participants. This evaluation is controlled under specific watching conditions. On the other hand, objective evaluation methods provide faster quality assessment, exploiting multiple metrics that use mathematical models to quantify the perceptual impact of the encoding artifacts (e.g. tilling, blurriness, error blocks, etc) on the video quality level. Nevertheless, the majority of the objective methods require the undistorted video source as a reference entity in the quality evaluation process. Due to this requirement, they are characterized as Full Reference (FR) Methods [1-3]. The recent research is focused also on developing methods that can evaluate the video quality level based on metrics, which use only some extracted structural features from the original signal (Reduced Reference Methods) [4-8] or do not require any reference video signal (No Reference Methods).

Thus, the aim of the current methods is the quantification of the user experience in terms of satisfaction. However, from a service provider aspect, which is interested to provide its contents free of charge, there is a need in term of more efficient bandwidth management for specifying i) the threshold up to which the user considers the quality of the encoded service as acceptable or below which considers it as unacceptable ii) the maximum perceived quality level that each video content can reach upon encoding and iii) the pattern of the video quality level vs. the encoding bit rate (which will provide to the user the capacity to offer a video at various quality levels). Apart from the various encoding parameters that play significant role in the deduced perceived quality level (e.g. bit rate, spatial and temporal resolution), the dynamics of the content (i.e. spatial and temporal activity of the content) are critical for the final perceptual outcome. Although a lot of research is focused on developing techniques and methods estimating the video quality of a compressed/encoded video signal, the impact of the video spatiotemporal dynamics on the video quality after encoding is not well addressed by the research community and hence explains the motivation of our work.

The main contribution of this paper is an experimental approach of the spatiotemporal content dynamics impacting i) the video quality acceptance threshold (i.e. the perceptual quality level below a certain quality which the user considers as unacceptable),

Mobimedia'09, September 7-9, 2009, London, UK.

ii) the highest achievable video quality level and iii) the pattern of video quality vs. encoding bit rate.

More specifically, this paper presents a study on the perceptual quality of the spatiotemporal dynamics of the content in correlation with the encoding bit rate. We consider that the other encoding parameters (e.g. spatial and temporal resolution, encoding scheme, GOP pattern etc.) remain constant. Towards this, we provide results, depicting the actual perceived efficiency for various activity levels. We consider not only the engineering effectiveness such as simple error-based metrics is considered but also as videos are actually perceived by the human visual system through a respective objective assessment metric.

In this framework this paper uses reference video clips, which are representative of different spatial and temporal activity levels, covering by this way all the range of the spatiotemporal scale. Afterwards, for each clip the relative PQoS vs. Bit rate curve for MPEG-4 encoding is drawn, showing how the differentiation in the content affects the deduced video quality.

The rest of this paper is organized as follows: Section 2 presents a two-dimensional approach on classifying the content dynamics of the video signals. In Section 3, we present two objective metrics for classifying a video sequence according to its spatiotemporal. Section 4 presents the spatiotemporal characteristics of the test signals that have been used in this paper. The relationship of the video quality to the spatial and temporal level of the video content is discussed in Section 5. Finally, Section 6 concludes this paper discussing the perspectives of the current research outcomes.

2. Spatiotemporal Content Plane: A two-dimensional classification of the content dynamics

The content of each video clip may differ substantially depending on its dynamics (i.e. the spatial complexity and/or the temporal activity of the depicted visual signal). The quantification of this diversity is of high interest to the video coding experts, because the spatiotemporal content dynamics of a video signal specify and determine the efficiency of a coding procedure.

From the perceptual aspect, the quality of a video sequence is dependent on the spatiotemporal dynamics of the content. More specifically, it is known from the fundamental principles of the video coding theory that action clips with high dynamic content are perceived as degraded in comparison to the sequences with slow-moving clips, subject to identical encoding procedures.

Thus the classification of the various video signals according to their spatiotemporal characteristics will provide to the video research community the ability to quantify the perceptual impact of the various content dynamics on the perceptual efficiency of the modern encoding standards.

Towards this classification, in [9] it is proposed a spatiotemporal plane, where each video signal (subject to short duration and homogeneous content) is depicted as Cartesian point in the spatiotemporal plane, where the horizontal axis refers to the spatial component of its content dynamics and the vertical axis refers to the temporal ones. The respective plane is depicted on Figure 1.

Therefore, according to this approach, each video clip can be classified to four categories depending on its content dynamics, namely:

- Low Spatial Activity – Low Temporal Activity (upper left)
- High Spatial Activity – Low Temporal Activity (upper right)
- Low Spatial Activity – High Temporal Activity (lower left)
- High Spatial Activity – High Temporal Activity (lower right)

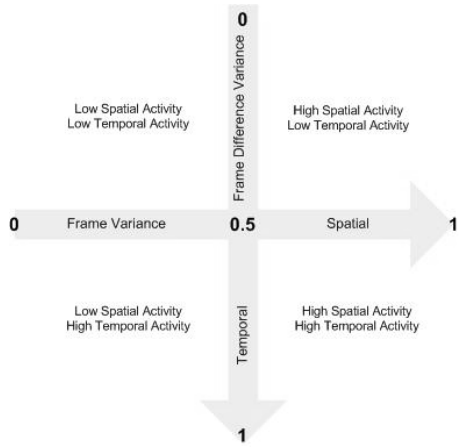

Figure 1: The Spatiotemporal grid used for classifying a video sequence according to its content dynamics

The accuracy of the proposed spatiotemporal content plane is subject to the duration of the video signal and the homogeneity of the content. For short duration and homogeneous content video clips, the classification is representative and efficient. However, for video clips of longer duration and heterogeneous content, their spatiotemporal classification is becoming difficult.

3. Objective Metrics for the Spatiotemporal Classification of Video Content

We propose to use two discrete metrics, one for the spatial component and one for the temporal one in order to cover the spatiotemporal plane and the needs of this paper.

The averaged frame variance is proposed for the spatial component of the video signal. This objective metric permits the quantification of the spatial dynamics of a video signal short in duration and homogeneous. Considering that a frame y is composed of N pixels x_i, then the variance of a frame is defined in equation 1:

Eq1: $$\sigma^2{}_{frame_y} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2$$

Derived from equation 1, equation 2 presents the averaged frame variance for the whole video duration. K represents the number of frames in the video.

Eq2: $$\frac{1}{K}\sum_{k=1}^{K}\sigma^2_{frame_y} = \frac{1}{K}\frac{1}{N}\sum_{k=1}^{K}\sum_{i=1}^{N}(x_{k,i} - \bar{x}_k)^2$$

The averaged variance of the successive *y* frame luminance difference is proposed as a metric for the quantification of the temporal dynamics of a video sequence. Considering that a frame contains *N* pixels x_i and K the number of frames in the video, then the averaged frame difference of the successive frame pairs is defined in equation 3.

Eq3: $$\frac{1}{K-1}\sum_{k=2}^{K}\frac{1}{N}\sum_{i=1}^{N}(x_{k,i} - x_{k-1,i})$$

Therefore, the averaged variance for the overall duration of the test signal is defined in equation 4.

Eq4: $$\frac{1}{K-1}\sum_{k=2}^{K}\left(\frac{1}{N}\sum_{i=1}^{N}(x_{k,i} - x_{k-1,i}) - \frac{1}{K-1}\sum_{k=2}^{K}\frac{1}{N}\sum_{i=1}^{N}(x_{k,i} - x_{k-1,i})\right)^2$$

The scale in both axes refers to the normalized measurements (considering a scale from 0 up to 1) of the spatial and temporal component, according to the aforementioned metrics. The normalization procedure applied in this paper, sets the test signal with the highest spatiotemporal content to the lower right quarter and specifically to the Cartesian (Spatial, Temporal) values (0.75, 0.75). This hypothesis, without any loss of generality, allows to our classification grid the possibility to consider also test signals that may have higher spatiotemporal content in comparison to the tested ones.

4. Classification of the Test Signals to the Spatiotemporal Content Plane

For the needs of this paper five short reference sequences are used. These sequences are depicted in table 1. Applying the described spatial and temporal metrics on the reference signals of Table 1, their classification on the proposed spatiotemporal grid is depicted on Figure 2.

According to Figure 2, it can be observed that the spatiotemporal dynamics of the selected reference signals are distributed to all the four quarters of the spatiotemporal grid, indicating their diverse nature of the content dynamics. Moreover, the validity of the proposed metrics is certified by these experimental results, showing that they provide adequate differentiation among the dynamics of the signals under test.

Based on the experimental results of Figure 2 and Table 1, it can be observed that the selected video signals are representatives of the whole range of the spatiotemporal activity range of the content dynamics and the spatiotemporal content plane.

In the next Section, we discuss the spatiotemporal content dynamics impact on i) the video quality acceptance threshold (i.e. the perceptual quality level below which the user considers that an encoded video is of unacceptable quality), ii) the highest achievable video quality level and iii) the pattern of video quality vs. encoding bit rate.

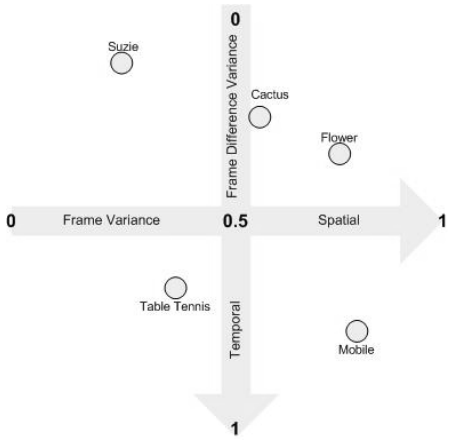

Figure 2: The Spatiotemporal classification of the test signals.

Suzie	
Cactus	
Flower Garden	
Table Tennis	
Mobile & Calendar	

Table 1: The five reference test signals

5. Spatiotemporal Activity and Video Quality

This section focuses on the impact of the spatiotemporal activity of the content on the video quality. The encoding bit rate needs to be adjusted according to this impact in order to provide a satisfying video quality to the end-user. It must be noted that the used sequences in this paper are reference signals with limited duration and therefore with practically homogeneous content (i.e. constant spatial and temporal activity level). The study with longer videos is out of the scope of this paper.

Each test video clip of Table 1, is encoded from its original uncompressed format to ISO MPEG-4 Visual Simple Profile format, at different constant bit rates (spanning a range from 50kbps to 1.5Mbps for CIF (Common Intermediate Format) with key-frame period equal to 100 frames in both cases). For each corresponding bit rate, a different ISO MPEG-4 compliant file is created. The frame rate is set at 25 frames per second (fps) for the whole encoding process.

Each ISO MPEG-4 video clip is then used as input in a no-reference objective quality measurement tool [10]. From the resulting quality per frame measurements, the average quality for the whole clip is calculated.

5.1 The impact of content dynamics on the video quality vs. bit rate pattern

This experimental procedure is repeated for each tested video clip and the respective curves representing the video quality vs. the encoding bit rate is depicted in Figure 3. The curves are following a general exponential pattern and present a significant leeway between the various spatiotemporal dynamics.

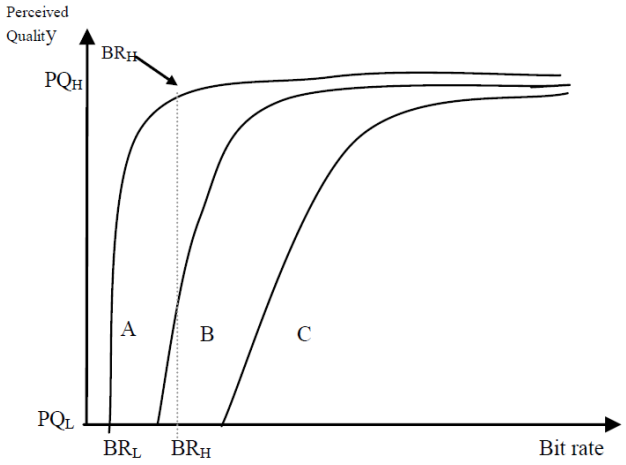

Figure 3: Impact of dynamics on the video quality vs. bit rate curves

More specifically, it can be observed that curve A represents video clip with low temporal and spatial dynamics, i.e. video content with "poor" movements and low picture complexity such as a talk show scene. Curve C represents video clip with high dynamics, such as a football match. Curve B represents an intermediate case. Practically, it can be observed that in low bitrates curve A reaches a higher perceptual level compared to curve B depicting a sequence with higher spatiotemporal content. On the other hand, the curve C) requires higher bit rate in order to reach a satisfactory PQoS level.

Nevertheless, curve(C) reaches its maximum PQoS value more smoothly than in the low activity case.

Moreover, each curve -and therefore each video clip- can be characterized by: (a) a low bit rate (BRL), which corresponds to the lower value of the accepted PQoS (PQL) by the audience, (b) the high bit rate (BRH), which corresponds to the minimum value of the bit rate for which the PQoS reaches its maximum PQH value (see BRH for curve (A) in figure 3) and (c) the mean inclination of the curve, which can be defined as ME = (PQH – PQL) / (BRH – BRL). From the curves of Figure 1, it can be deduced that video clips with low dynamics have lower BRL and higher ME than clips with high dynamics.

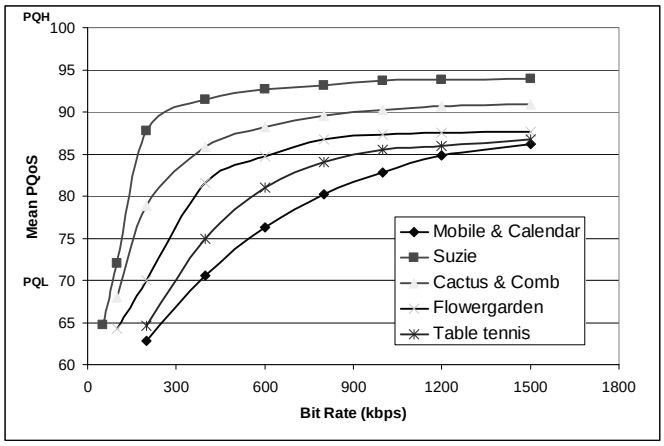

Figure 4: The Video Quality vs. Bit Rate curves

Following the general pattern in Figure 3, the respective experimental data for the reference signals that have been tested are depicted in Figure 4. As it can be observed, the impact of the spatiotemporal activity on the content is depicted very clear. It also shows two more important outcomes:

i) For video signals with low spatiotemporal activity, a saturation point appears, above which the perceptual enhancement is negligible even for very high encoding bit rates.

ii) As the spatiotemporal activity of the content becomes higher, the respective perceptual saturation point (i.e. the highest perceptual quality level) becomes lower, which practically means that video of high dynamics never reach a very high perceptual level.

Based on these observations, the next sub section examines in more details the impact of the content dynamics on the perceptual saturation point (i.e. the highest perceptual quality level).

5.2 The impact of content dynamics on the highest perceptual quality level

Focusing more on the impact of the spatiotemporal content dynamics on the perceptual saturation point (i.e. the highest perceptual quality level that each video signal can achieve), it can be observed directly from both Figures 3 and 4 that video signals with relatively low spatiotemporal content achieve higher perceptual levels than video signals that contain content of high dynamics.

In this framework, Figure 5 depicts the experimental results for the test signals of this paper, concerning the highest perceptual quality level (PQH) for both CIF and QCIF spatial resolution.

It can be observed that for both CIF and QCIF spatial resolution, the impact of the spatiotemporal activity is significant making especially the signals of low content dynamics less demanding in terms of encoding bit rate for a certain perceived threshold.

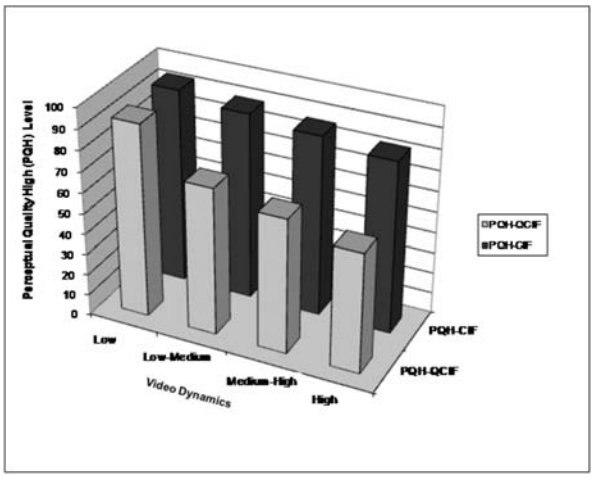

Figure 5: Impact of dynamics on the PQoS saturation point

5.3 The impact of content dynamics on the video quality acceptance threshold

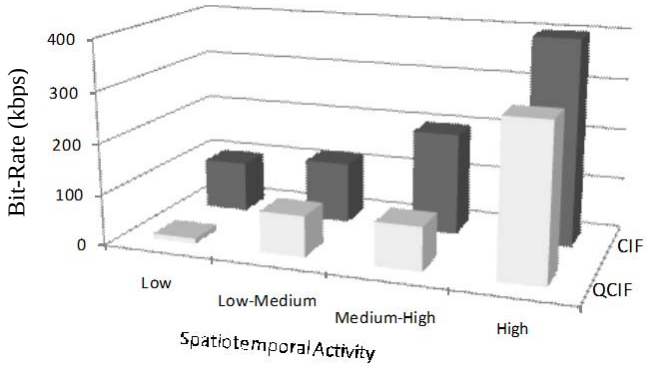

Figure 6: The impact of dynamics on the acceptance PQoS threshold

This sub section examines the impact of the spatiotemporal activity of the content on the perceptual acceptance threshold for the various test signals.

The respective results are depicted in Figure 6. The lowest acceptable perceptual level is fixed to 3.5 in the MOS scale. Based on these experimental results, it is shown that for both CIF and QCIF spatial resolution need higher bit rate in order to achieve the perceptual acceptance threshold when the spatiotemporal activity becomes more complex. Especially for the case of CIF, the demand in terms of bit rate becomes higher than for the case of QCIF.

6. Conclusions

This paper presents the impact of the video spatiotemporal dynamics on the deduced perceptual quality. More specifically it shows how the spatiotemporal activity affects i) the highest quality level that each video can reach, ii) the video quality acceptance threshold (i.e. the lowest quality level) and iii) the video quality vs. bit rate pattern. This paper proves that the spatiotemporal activity has a significant impact on the video quality of the encoded signal and can be used in streaming applications or IPTV services over heterogeneous devices. This work can have an impact on the way operators exploit their networks by maximizing the End-User PQoS while optimizing the network resources (bandwidth).

7. ACKNOWLEDGMENT

The work in this paper has been performed within the research framework of FP7 ICT-214751 ADAMANTIUM Project.

8. 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.

[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] M. Montenovo, A. Perot, M. Carli, P. Cicchetti, A. Neri, Objective evaluation of video services. Proc. of 2nd Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2006.

[6] S. S. Hemami, M. A. Masry, A scalable video quality metric and applications. Proc. of 1st Int. Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2005.

[7] O. A. Lotfallah, M. Reisslein, S. Panchanathan, A framework for advanced video traces: Evaluating visual quality for video transmission over lossy networks. (Article ID 42083) EURASIP Journal on Applied Signal Processing, 2006. 2006.

[8] Zhou Wang, Guixing Wu, Hamid R. Sheikh, Eero P. Simoncelli, En-Hui Yang and Alan C. Bovik, Quality-Aware Images. IEEE Transactions on Image Processing.

[9] N. Cranley and L. Murphy, "Incorporating User Perception in Adaptive Video Streaming Systems", in Digital Multimedia Perception and Design (Eds. G. Ghinea and S. Chen), published by Idea Group, Inc., May 2006. ISBN: 1-59140-860-1/1-59140-861-X

[10] J. Lauterjung, "Picture Quality Measurement", Proceedings of the International Broadcasting Convention (IBC), Amsterdam, 1998, pp. 413-417