Objective Measurement of Perceived QoS for Homogeneous MPEG-4 Video Content

Harilaos Koumaras
University of Athens, Greece

Drakoulis Martakos
National and Kapodistrian University of Athens, Greece

Anastasios Kourtis
Institute of Informatics and Telecommunications NCSR Demokritos, Greece

INTRODUCTION

Multimedia applications over 3G and 4G (third and fourth generation) networks will be based on digital encoding techniques (e.g., MPEG-4) that achieve high compression ratios by exploiting the spatial and temporal redundancy in video sequences. However, digital encoding causes image artifacts, which result in perceived-quality degradation. Due to the fact that the parameters with strong influence on the video quality are normally those set at the encoder (most importantly, the bit rate and resolution), the issue of user satisfaction in correlation with the encoding parameters has been raised (MPEG Test, 1999).

One of the 3G-4G visions is to provide audiovisual (AV) content at different qualities and price levels. There are many approaches to this issue, one being the perceived quality of service (PQoS) concept. The evaluation of the PQoS for multimedia and audiovisual content will provide a user with a range of potential choices, covering the possibilities of low, medium, or high quality levels. Moreover the PQoS evaluation gives the service provider and network operator the capability to minimize the storage and network resources by allocating only the resources that are sufficient to maintain a specific level of user satisfaction.

This paper presents an objective PQoS evaluation method for MPEG-4-video-encoded sources based on a single metric experimentally derived from the spatial and temporal (S-T) activity level within a given MPEG-4 video.

Toward this, a quality meter tool was used (Lauterjung, 1998), providing objective PQoS results for each frame within a video clip. The graphical representation of these results vs. time demonstrated the instant PQoS of each frame within the video clip, besides indicating the mean PQoS (MPQoS) of the entire video (for the whole clip duration). The results of these experiments were used to draw up experimental curves of the MPQoS as a function of the encoding parameters (i.e., bit rate). The same procedure was applied for a set of homogeneous video sequences, each one representing a specific S-T activity level.

Furthermore, this paper shows that the experimental MPQoS vs. bit-rate curves can be successfully approximated by a group of exponential functions, which confines the QoS characteristics of each individual video test sequence to three parameters. Showing the interconnection of these parameters, it is deduced that the experimental measurement of just one of them, for a given short video clip, is sufficient for the determination of the other two. Thus, the MPQoS is exploited as a criterion for preencoding decisions concerning the encoding parameters that satisfy a certain PQoS in respect to a given S-T activity level of a video signal.

BACKGROUND

Over the last years, emphasis has been put on developing methods and techniques for evaluating the perceived quality of video content. These methods are mainly categorized into two classes: the subjective and objective ones.

The subjective test methods involve an audience of people who watch a video sequence and score its quality as perceived by them under specific and controlled watching conditions. 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: the single-stimulus continue quality evaluation (SSCQE) and the double-stimulus continue quality evaluation (DSCQE) (Alpert & Contin, 1997; ITU-R, 1996; Pereira & Alpert, 1997). However the MOS method is inconvenient due to the fact that the preparation and execution of subjective tests is costly and time consuming.

For this reason, a lot of effort has recently been focused on developing cheaper, faster, and easier applicable objective evaluation methods. These techniques successfully emulate the subjective quality-assessment results based on criteria and metrics that can be measured objectively. The objective methods are classified according to the availability of the original video signal, which is considered to be of high quality.

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, and due to this, they are characterized as full-reference methods (Tan & Ghanbari, 2000; Wolf & Pinson, 1999). These methods are based on an error-sensitivity framework with the most widely used metrics: the peak-signal-to-noise ratio (PSNR) and the mean square error (MSE).

$$\mathrm{PSNR} = 10\log_{10}\frac{L^2}{\mathrm{MSE}} \tag{1}$$

where L denotes the dynamic range of pixel values (equal to 255 for 8 bits/pixel monotonic signal).

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(x_i - y_i)^2 \tag{2}$$

where N denotes the number of pixels, and x_i/y_i is the i^{th} pixel in the original and distorted signal.

However, these overused metrics have seriously been criticized as not providing reliable measurements of the perceived quality (Wang, Bovik, & Lu, 2002). For this reason, a lot of effort has been focused on developing assessment methods that emulate characteristics of the human visual system (HVS) (Bradley, 1999; Daly, 1992; Lai & Kuo, 2000; Watson, Hu, & McGowan, 2001) using contrast-sensitivity functions (CSFs), channel decomposition, error normalization, weighting, and Minkowski error pooling for combining the error measurements into a single perceived-quality estimation. An analytical description of the framework that these methods use can be found in Wang, Sheikh, and Bovik (2003).

However, it has been reported (VQEG, 2000; Wang et al., 2002) that these complicated methods do not provide more reliable results than the simple mathematical measures (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 (Wang, Bovik, Sheikh, & Simoncelli, 2004; Wang, Lu, & Bovik, 2004).

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, the synchronization predicaments between the undistorted and the distorted signal (which may have experienced frame loss) make the implementation of the full-reference methods difficult.

Due to these reasons, recent research has been focused on developing methods that can evaluate the PQoS based on metrics that use only some extracted features from the original signal (reduced-reference methods) (Guawan & Ghanbari, 2003) or do not require any reference video signal (no-reference methods) (Lauterjung, 1998; Lu, Wang, Bovik, & Kouloheris, 2002).

A software implementation that is representative of this nonreference objective evaluation class is the quality meter software (QMS) that was used for the needs of this paper (Lauterjung, 1998). The QMS tool measures objectively the instant PQoS level (on a scale from 1 to 100) of digital video clips. The metrics used by the QMS are vectors, which contain information about the averaged luminance differences of adjacent pixel pairs that are located across and on both sides of adjacent DCT-block (8x8