A Testbed for the Evaluation of MPEG Video Transmission on ATM Networks

Christian J. van den Branden Lambrecht
Signal Processing Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
vdb@lts.de.epfl.ch
http://ltswww.epfl.ch/~vdb/

Andrea Basso
Telecommunications Laboratory
Swiss Federal Institute of Technology
CH-1015 Lausanne, Switzerland
basso@tcom.epfl.ch
http://tcomwww.epfl.ch/

ABSTRACT

Most of the new broadcasting and multimedia applications intensively rely on networked video. The current trend for distributing digital video on broadband ISDN networks is towards the adaptation of MPEG streams on ATM networks. End-to-end testing of such communication systems is very important and requires robust testing methodologies that are capable of evaluating both coding and transmission errors. This paper proposes a complete architecture for doing so. The system is entirely automatic, relies on synthetic test patterns and estimates the subjective quality of video coding and network transmission.

1 INTRODUCTION

Video transmission systems are currently in a state of transition from a completely analog system to a digital system. The digital system which will be extensively deployed will incorporate source coders and networked video on broadband integrated services digital networks (B-ISDN). The source coding will be using the MPEG [1] compression standards and B-ISDN will be asynchronous transfer mode (ATM) networks. In video delivery systems, end-to-end testing is essential. For analog video systems, the methodology for end-to-end testing is well established and consists of various signals that are inserted at the appropriate places in the analog video stream. The insertion of the test signals can be separate from the main program video and does not affect the performance of the program video. For digital video systems, this approach is not feasible since insertion of the test patterns can affect the quality of the program video as well.

This paper presents an architecture and methodology for end-to-end testing of MPEG video delivery systems on ATM networks. The system relies on a synthetic test pattern generator (TPG) that creates specific test patterns for MPEG based systems and on a perceptual video quality metric. The test pattern description is in function form; thus a replica of the test pattern can be easily generated at the decoder side to evaluate the performance of MPEG video at the decoder. By a careful choice of the test patterns, the proposed testing methodology facilitates independent evaluation of each artifact introduced by the coding process. The computational quality metric, termed moving pictures quality metric (MPQM), is based on a spatio-temporal model of vision so as to predict video quality as perceived by human observers. The paper is divided as follows: Section 2 presents the architecture of the testbed, namely the test pattern generator, the computational metric and the setup of the ATM network. The experiments carried out to validate the approach are described in Sec. 3 and Section 4 concludes the paper.

2 ARCHITECTURE

The general architecture of the testbed is depicted in Fig. 1. It features the TPG (divided into a emitter-receiver tandem), an MPEG-2 encoder/decoder and an ATM network. Its building blocks are now described.

2.1 The Test Pattern Generator

The TPG, a complete description of which can be found in [7], consists of two modules, an emitter and a receiver. The emitter consists solely of a pattern generator. The receiver contains the quality evaluation module and a pattern generator which is identical to the one at the emitter. Thus, the testing device has access to the non-coded and decoded versions of the test sequence from which quantitative measurements regarding image fidelity can be made at the receiver.

Since both pattern generators must generate the same sequences concurrently in order to perform quality estimation, a synchronization procedure has been designed and synchronization information is inserted into the test patterns. Furthermore, the pattern generator in the receiver must know which pattern was generated by the emitter. Pattern related information is also included in the test pattern and can be extracted by the receiver module from the output of the decoder.

The test methodology is based on a library of synthetic test patterns. Synthetic test patterns offer several advantages over natural scenes:

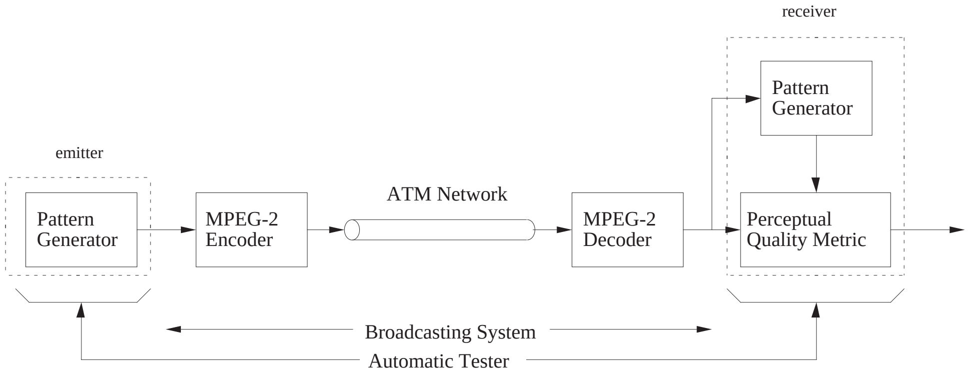

Figure 1: Architecture of the proposed testing system.

- Synthetic patterns generated algorithmically are resolution-independent and hence can be generated for any temporal and spatial resolution of the video frames.
- Synthetic patterns require much less memory than complex natural scenes.
- Algorithms can be designed to generate customizable patterns, which provide more latitude in testing the features of a particular coder.
- Test patterns are entirely deterministic, which makes quality evaluation easier.

Each pattern that the TPG generates is meant to test the coder for a particular artifact. The variety of tests that can be run is made of luminance and chrominance rendition, edge rendition, blocking effect, isotropy, abrupt scene change, text rendition, texture rendition, time aliasing, motion rendition and buffer control (for constant bitrate coder).

2.2 The Video Quality Metric

Several studies have shown that a correct estimation of subjective quality has to incorporate some modeling of human vision [5,9]. A spatio-temporal model of human vision has been developed for the assessment of video coding quality [6,8]. The model is based on the following properties of human vision:

- The responses of the neurons of the primary visual cortex (called area V1) are band-limited. The human visual system has a collection of mechanisms or detectors (called channels) that mediate perception. A channel is characterized by a localization in spatial frequency, spatial orientation and temporal frequency. Such channels are simulated by a three-dimensional filter bank.
- In a first approximation, those channels can be considered to be independent. Perception can thus be predicted channel by channel without interaction.
- Perception in a channel is characterized by two phenomena: contrast sensitivity and masking. Human sensitivity to contrast is known to be a function of frequency (spatial and temporal) as well as orientation. This leads to the concept of *contrast sensitivity function*, which specifies the threshold of detection for a stimulus as a function of frequency. Masking accounts for inter-stimuli interferences. It is known that the presence of a background stimulus will modify the detection of a foreground stimulus. Masking thus corresponds to a modification of the detection threshold according to the local contrast of the background.

The working model described in [6] incorporates the above described considerations of the HVS. The filter bank used in the model decomposes the data according to 5 spatial frequency bands, 4 orientation bands and 2 temporal frequency bands. It has been especially parameterized for the framework of video coding by means of psychophysical experiments.

Such a model predicts the response of the neurons in area V1 of the cortex and thus the *perceived* distortion. This is done by decomposing the data in perceptual channels and predicting perceived stimuli using contrast sensitivity and masking. Thereafter, a distortion measure is computed, accounting for the higher levels of cognition of the brain. At this stage, the metric also accounts for the focus of attention and is computed over blocks of the sequence. Such blocks are three-dimensional and their dimensions are chosen as follows: the temporal dimension is chosen to account for persistence of the images on the retina. The spatial dimension corresponds to the focus of attention, i.e. the size is computed so that a block covers two degrees of visual angle, which is the dimension of the fovea. The distortion measure is computed for each block by pooling the error over the channels. Basically, the magnitude of the channels output are combined by Minkowski summation with a higher exponent to weight the higher distortions more. The actual computation of the distortion M_E for a given block is computed according to Eq. 1:

$$M_E = \left(\frac{1}{N_c} \sum_{c=1}^{N_c} \left(\frac{1}{N_q} \sum_{t=1}^{N_t} \sum_{x=1}^{N_x} \sum_{y=1}^{N_y} |e[x,y,t,c]| \right)^{\beta} \right)^{\frac{1}{\beta}}, \quad (1)$$

where $e[x,y,t,c]$ is the masked error signal at position (x,y) and time t in the current block and in the channel c; N_x, N_y and N_t are the horizontal and vertical dimensions of the blocks; N_c is the number of channels and $N_q = N_x N_y N_t$. The exponent of the Minkowski summation is β and has a value of 4.

In this application, the error measure M_E is further mapped onto a quality scale from 1 to 5 according to the following function, relating the error measure to the quality index MPQM:

$$\text{MPQM} = \frac{5}{1 + N M_E},$$

where N ensures a mapping between 1 and 5. This free parameter has been estimated on the basis of the vision model [8] and has a value of $N = 0.623$.

2.3 ATM setup

In networked multimedia real time applications and particularly in networked video, *glitching* refers to perturbations due to the lack or damage of the video stream [3]. The relevant parameters of a glitch are its spatial extent, its temporal extent and its rate.

The spatial extent of a glitch is the part of the picture that is incorrectly displayed. Its occurrence depends on many factors. An important one is the method used to code the picture. As an example, in the context of MPEG coding, a glitch occurring on an intra coded frame (I frame) is the origin of distortions in the subsequent frames due to the predictive coding method of MPEG. This also defines the temporal extent of a glitch as the number of frames onto which the glitch propagates. In the context of this work, the goal of the ATM network simulation is to reproduce the essential visual effects of glitching.

The ATM setup that has been used for the simulations has been designed so as to simulate the most significative ATM network errors. Attention has been focused on two parameters: the mean loss rate and the congestion at the switch level.

The MPEG-2 video elementary streams are mapped onto AAL-5 packets (ATM adaptation layer). The compressed video data is placed in the protocol data unit (PDU) as soon as it arrives regardless of the macroblock boundaries. The chosen PDU size is 376 bytes. The losses occur in bursts and are uniformly distributed in time. Typical simulated bursts range from 2 to 100 packets. Simulations showed that this setup permits to generate a large variety of glitches.

Any AAL-5 damaged packet is considered lost and thus discarded. This choice can be considered as a worst case compared to the current trend in ATM implementations where, when a cyclic redundancy check (CRC) of mismatched PDU size is observed, the decoder can be notified of the error and the corrupted PDU can be passed to the application with an error indication. The motivation of our choice is related to the fact that a basic MPEG-2 decoder, without any recovering or concealment capabilities, has been considered.

3 SIMULATIONS

3.1 Video Material

The synthetic sequence that has been used in this experiment has been constructed as follows. It features a total of 69 frames. The first 5 frames are complex textures uncorrelated in time. This has the effect to quickly fill the output buffer of a constant bit rate (CBR) coder. The following 64 frames are the blocking effect test. It is made of a a series of squares whose dimensions decrease with time. Each square is filled with a slowly varying function of luminance that will favor the appearance of the blocking effect due to the use of the block DCT. As the complexity of the sequence increases, the importance of the blocking effect increases as well.

The sequence has been encoded at bitrates of 3, 6, 9 and 15 Mbit/sec. using a software simulator of the test model 5 of MPEG-2 operating in main profile, main level. The software simulation has been released by the MPEG software simulation group [4]. Search windows of 15 pixels for P frames and 7 pixels for B frames have been used. The output buffer has been set to its maximum size.

3.2 Network

The compressed video streams have been fed into the simulator of the ATM link and transmission errors have been simulated. Loss rates of 10^{-5}, 10^{-4} and 10^{-3} have been used in the experiment, as they are typical values for such setups.

3.3 Results

Simulations results are now presented. Due to computational complexity and memory management, the MPQM could only be applied to 32 frames of the sequences. It has been chosen to always use the first 32 frames of the video stream. Figure 2 presents the results of the quality assessment by MPQM for the various streams. The quality rating on a scale from 1 to 5 is presented as a function of the bitrate for the considered loss rates. The solid line is the quality assessment when no loss is introduced. The curve is very similar to those obtained in [2, 8]. Quality increases more or less linearly with the bitrate in the low to medium range of bitrates and starts to saturate at higher bitrates. The other curves, corresponding to corrupted streams, have been rated lower in quality by MPQM, as they exhibit more distortion. Their appearance is more erratic tough and quality can even drop as the bitrate increases (especially for the streams having a loss rate of 10^{-4} in this experiment). This behavior is due to the fact that quality is directly dependent on the glitch extent (more than any other factor, namely the loss rate). The streams compressed at 9 and 15 Mbit/sec., for example, have been rated very low in quality. The distribution of errors for these sequences showed that large bursts appeared on the first two I frames. Such errors propagate over the whole group of pictures (GOP), yielding a sequence that looks much worse than the stream compressed at 6 Mbit/sec.

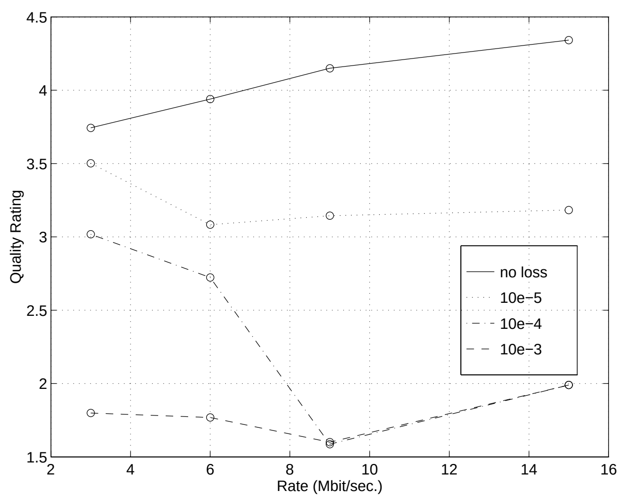

Figure 2: Quality assessment by MPQM for the synthetic sequences as a function of the bitrate and the loss rate.

4 CONCLUSION

An architecture for the end-to-end testing of digital video transmission system has been presented. The video delivery system that has been considered transmits MPEG compressed video streams over ATM networks. The proposed testing architecture relies on two key aspects: a test pattern generator and a video quality metric. The TPG creates synthetic patterns that favor the apparition of typical coding artifacts. Its structure permits a convenient end-to-end testing of transmission systems as quality assessment can be performed at the receiver end with no need to transmit the uncompressed video material. The quality metric is based on a spatio-temporal model of the human visual system and predicts the quality of video sequences as perceived by human observers.

The simulations that have been presented show that the system is suited for the targeted application as it permits to measure both coding distortion and transmission errors. It has also been observed that the quality metric is able to assess quality of streams that are corrupted by the two types of distortion, i.e. coding and transmission.

References

[1] ISO/IEC JTc1/SC29/WG11 13818, "Generic Coding if Moving Pictures and Associated Audio; Part 2: Video", Technical report, International Standardization Organization/International Electrotechnical Commission, Geneva, Switzerland, 1995.

[2] Andrea Basso, İsmail Dalgıç, Fouad A. Tobagi, and Christian J. van den Branden Lambrecht, "Study of MPEG-2 Coding Performance based on a Perceptual Quality Metric", *in Proceedings of the Picture Coding Symposium*, Melbourne, Australia, March 13-15 1996, available on `http://ltswww.epfl.ch/~vdb`.

[3] İsmail Dalgıç and Fouad A. Tobagi, "Glitchs as a Measure of Video Quality Degradation Caused by Packet Loss", *in Proceedings of the 7th International Workshop on Packet Video*, pp. 201–206, March 18-19 1996.

[4] Chadd Fogg, "mpeg2encode/mpeg2decode version 1.1", available via anonymous ftp at netcom.com, June 1994, MPEG Software Simulation Group.

[5] J. L. Mannos and D. J. Sakrison, "The effects of a visual fidelity criterion on the encoding of images", *IEEE Transactions on Information Theory*, Vol. IT-20, No. 4, pp. 525–536, 1974.

[6] Christian J. van den Branden Lambrecht, "A Working Spatio-Temporal Model of the Human Visual System for Image Restoration and Quality Assessment Applications", *in Proceedings of the International Conference on Acoustics, Speech, and Signal Processing*, Atlanta, GA, May 7-10 1996, available on `http://ltswww.epfl.ch/~vdb`.

[7] Christian J. van den Branden Lambrecht, Vasudev Bhaskaran, Al Kovalick, and Murat Kunt, "Automatically Assessing MPEG Coding Fidelity", *IEEE Design and Test Magazine*, Vol. 12, No. 4, pp. 28–33, winter 1995.

[8] Christian J. van den Branden Lambrecht and Olivier Verscheure, "Perceptual Quality Measure using a Spatio-Temporal Model of the Human Visual System", *in Proceedings of the IS&T Symposium on Electronic Imaging: Science and Technology*, San Jose, CA, January 28 - February 2 1996. IS&T/SPIE, available on `http://ltswww.epfl.ch/~vdb`.

[9] Andrew B. Watson, "Efficiency of an Image Code Based on Human Vision", *Journal of the Optical Society of America*, Vol. 4, No. 12, pp. 2401–2417, December 1987.