LARGE-SCALE VIDEO COMPRESSION RESEARCH WORK ON QUALITY OF EXPERIENCE (QOE) EVALUATION

Mikolaj Leszczuk, A. Glowacz, J. Derkacz, P. Romaniak, B. Szczerba

LARGE-SCALE VIDEO COMPRESSION RESEARCH WORK ON QUALITY OF EXPERIENCE (QOE) EVALUATION

M. Leszczuk, A. Głowacz, J. Derkacz, A. Dziech

AGH University
Department of Telecommunications
Kraków, Poland

P. Romaniak, B. Szczerba

AGH University
ACC CYFRONET
Kraków, Poland

ABSTRACT

The large variety of video data sources means variability not only in terms of included content, but also in terms of quality. The paper describes a comprehensive evaluation experiment on perceived video quality. Consequently, in summary, $19,200,000$ video frames will be processed. Given the scale of the experiment, it is set up on a computer cluster in order to accelerate the calculations significantly. This work on Quality of Experience (QoE) is synchronized with that conducted by the Video Quality Experts Group (VQEG). The resulting data will be available to the community as they may be useful in other projects as well.

Index Terms— Video, compression, QoE, MOS, H.264

I. INTRODUCTION AND MOTIVATION

The large variety of multimedia data sources means variability not only in terms of included content, but also in terms of quality. This applies in particular to video sequences, where quality is extremely variable, e.g. recordings currently made with an amateur camera at one extreme and recordings of specialized medical tests for diagnostic purposes at the other. Therefore, quality assessment provides a challenge, especially considering the number of places in the transmission chain where distortions can be introduced, namely (as shown in the Figure 1):

1) Original video acquisition,
2) Encoder and streaming server,
3) Network,
4) User equipment.

In terms of mechanisms delivering high-quality digital content, a common feature of existing systems is the need to ensure acceptable quality of streaming video sequences, regardless of the load on the transmission medium, the type of access network, or the end users equipment. The term ”acceptable quality” is not clearly defined, and mostly depends on the scenario under consideration. Video streaming systems incapable of providing acceptable quality ceases to be attractive to potential users, and as such they are not in general use. Due to the above requirements, an important element of the research is to develop a system to ensure adequate quality of video sequences. This system should include metrics for quality assessment and quality optimization mechanisms that use information provided by the metrics.

The Video Quality Experts Group (VQEG) has been studying quality assessment for several years [1]. For example, one of the previous large-scale evaluations from VQEG, “Multimedia Test-Plan”, consisted of 43 subjective experiments with 160 sequences each.

II. THE SYNAT PROJECT

One of the aims of the SYNAT project research is to develop methods of delivering high-quality multimedia content. Implementation of a variety of methods for segmentation of digital objects, transcoding, and adaptation of multimedia content is proposed. This will include metrics and algorithms for assessing the quality of video sequences. It is also proposed that additional information on the quality of digital objects for the materials that meet the criteria will be stored.

III. NOVELTY

The typical approach to learning about video quality involves conducting subjective experiments. They are time- and resources-consuming; it is also necessary to collect appropriate video content (see Figure 2 for example test sequences).

While 24 subjects (testers) are enough to test a small set of parameters, it is often necessary to use up to 100 testers for more parameters. The test, using the widely accepted methodology ITU’s ACR-HR (Absolute Category Rating with Hidden Reference, ITU-T P.910) [3], involves screening video sequences and collecting of quality marks, one by one (see Figure 3).

However, it is worth considering whether these experiments are really necessary.

The main innovation of the proposed solution is a comprehensive, large-scale evaluation of the perceived quality of video, and optimization based on information provided by the proposed metrics. The complexity of the evaluation is related to the fact that it uses distortions typical of the video acquisition process and those caused by lossy compression

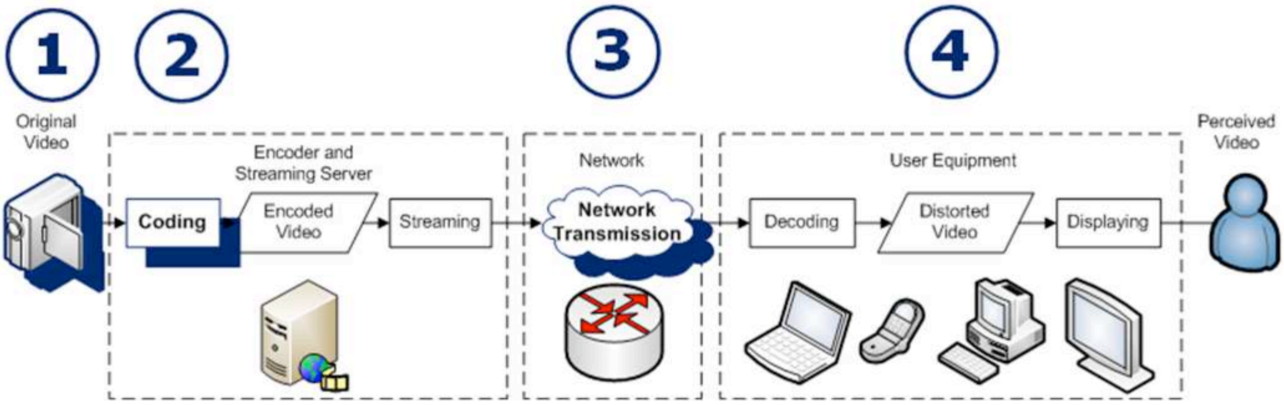

Fig. 1. Transmission chain.

Fig. 3. Methodology: ITU's ACR-HR (Absolute Category Rating with Hidden Reference, ITU-T P.910 [3].

Fig. 2. Example test sequences [2].

for streaming. Through subjective testing and application of advanced statistical methods for the analysis of data received, it will be possible to create models mapping measured quality parameters onto values for the resultant video quality, expressed in terms understandable to the user (e.g. 5-point scale of the Mean Opinion Score, MOS). Another important issue is to optimize the quality of the information provided by the metric, written in the form of metadata, or computed on demand. Optimizing the quality will take into account not only the network parameters but also the characteristics of video sequences, which is an important innovation.

IV. EXPERIMENT IMPLEMENTATION

The work on QoE algorithms in the SYNAT project was synchronized with that conducted by the VQEG, in particular the project of the Joint Efforts Group (JEG) – Hybrid group [4]. The objective of the JEG-Hybrid group is to assess the quality of video sequences based on the bit stream and decoded frames and motion, that is, information that is readily available on the DVD player/receiver. The first assessment of the JEG-Hybrid group is limited to video encoding based on the H.264 standard and compatible with packet transmission.

Recent activity in the SYNAT project, in parallel with the JEG-Hybrid group, resulted in the selection of sets of Source Video Sequences (SRCs) and of parameter sets – Hypothetical Reference Circuits (HRCs).

IV-A. Collections of Source Reference Channels/Circuits (SRCs)

In this experiment, it was assumed that there are no restrictions imposed on the overall quality of the SRC source video sequences. However, the SRC content, which is of lower quality than MOS value 4 (on the 1–5 scale), needs to be treated separately. For example, for measurements in the Full Reference model, if a reliable prediction of subjective assessment is required, this type of content cannot be used.

This is a preliminary estimate. Certain video sequences can be repeated in some collections, which will require further verification. Many sequences originate from the database of the Consumer Digital Video Library (CDVL) [5].

IV-B. Hypothetical Reference Circuit (HRC) Parameter Sets

The set of HRC parameters defines the parameter set for a video encoder which can be used for compression. Each of the proposed HRCs differs from the others in at least one of the parameters under consideration. The entire range of HRCs should cover all probable video compression scenarios, designed for digital storage in digital libraries. Each SRC should be distorted (encoded) using each of the HRCs, which will result in a number of distorted Processed Video Sequences (PVSs) equal to SRC $\times$ HRC sequences.

The video compression parameters under consideration are as follows:

- Bit-rate/quantization factor, with at least 5+5 choices
- Different Group of Pictures (GoP) sizes and structures:
  - 2 choices of a number of I frames per second
  - 2 choices of a number of B/P frames per second
  - Hierarchical coding – of 2 horizontal variants
- Structure of slices – of at least 2 different lengths
- Number of frames per second – original and halved
- Resolution – original and halved

IV-C. Experiment Scale

The effort required to generate this database is significant. In order to highlight this, the number of bit streams that will be created is calculated as follows. First of all, approximately 100 SRCs need to be selected. Each one needs to be:

- 10 seconds long,
- With 30 FPS, and
- With resolution ranging from SD ($720 \times 480/576$) to HD (1920×1080).

HRCs need to be selected next, currently giving a total of 640 combinations. Consequently, in summary, $100 \times 10 \times 30 \times 640 = 19,200,000$ video frames will be processed:

- Each video frame needs to be encoded (1 s is required to encode a single video frame using JM Reference Software using a single CPU core),
- For each video frame, video quality metrics need to be calculated,
- For each video frame, video content characteristics need to be calculated.

Peak Signal-to-Noise Ratio (PSNR) is probably the best known metric. It is a data metric, meaning that it looks at the fidelity of the signal without considering its actual content [6]. Data metrics have been an extremely popular research topic in the past, and they are widely used in image and video quality assessment. However, they always operate on the whole frame and do not consider any other important factors that can strongly affect the perceived quality (such as human visual system, HVS, characteristics). Figure 4 shows

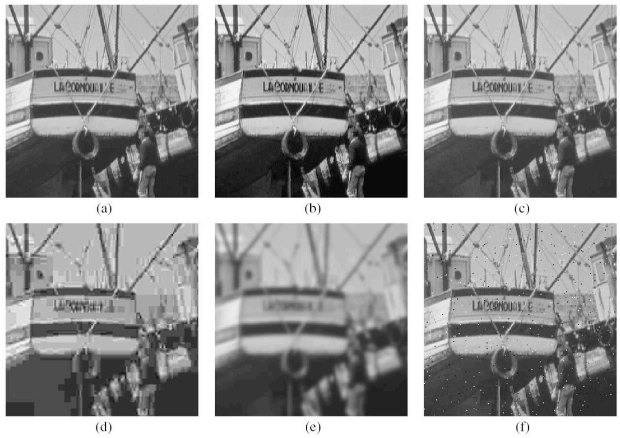

Fig. 4. PSNR — is it good enough?

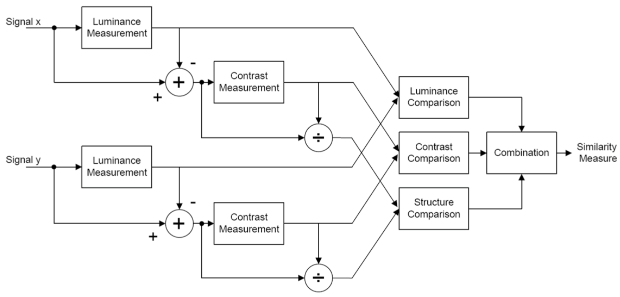

Fig. 5. SSIM.

distorted pictures with the same quality according to the PSNR metrics. However, it is clear that the perceived quality is very different. The conclusion is that data metrics have a relatively low performance and low feasibility.

Consequently, other metrics under consideration include the *Structural Similarity Index* (SSIM) [7] and the *Video Quality Metric* (VQM).

The SSIM is a top-down approach (see Figure 5) using a functional model of the HVS (a more detailed description of HVS can be found in [6]).

The VQM measures the perceived effects of video impairment including blurring, jerky/unnatural motion, global noise, block distortion and colour distortion, and combines them into a single metric [8].

IV-D. Calculation Time using Single Core

Calculation time using a single core has been estimated as approx. 889 days. This was based on the following assumptions:

Assumption 1: $19,200,000$ video frames.
Assumption 2: compression of a single video frame in 1 s.
Assumption 3: three video quality metrics to be calculated for all video frames.

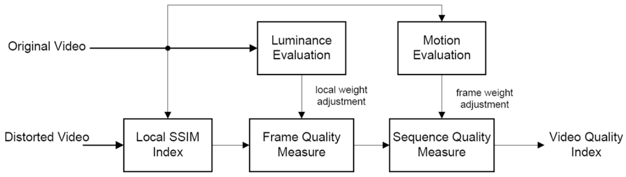

Fig. 6. VQM.

Assumption 4: calculation of a single video quality metric for one frame in approx. 1 s.

V. CONCLUSION

Given the above, the experiment is set up on a computer cluster in order to accelerate the calculations significantly. This will allow researchers to verify the encoding parameters of the resulting video sequences and determine the next steps of the experiments (in particular dates and resources to store PVS and uncompressed sequences). Calculations resulted in the creation of a very large collection of test sequences that can be used in the subjective and objective tests, which can be used to develop appropriate algorithms for video encoding. The resulting data will be available to the community as they may be useful in other projects as well.

ACKNOWLEDGEMENTS

Work financed by The National Centre for Research and Development (NCBiR) as part of project no. SP/I/1/77065/10.

VI. REFERENCES

[1] *The Video Quality Experts Group*, VQEG, http://www.vqeg.org/.

[2] M. Mu, P. Romaniak, A. Mauthe, M. Leszczuk, L. Janowski, and E. Cerqueira, "Framework for the integrated video quality assessment," *Multimedia Tools and Applications*, pp. 1–31, 10.1007/s11042-011-0946-3. [Online]. Available: http://dx.doi.org/10.1007/s11042-011-0946-3

[3] *ITU-T P.910, Subjective video quality assessment methods for multimedia applications*, International Telecommunication Union Recommendation, 1999. [Online]. Available: http://www.itu.int/rec/T-REC-P.910-200804-I

[4] M. Barkowsky, N. Staelens, and L. Janowski, *The JEG Hybrid Group*, VQEG, http://www.its.bldrdoc.gov/vqeg/project-pages/jeg/jeg.aspx.

[5] *The Consumer Digital Video Library*, CDVL, http://www.cdvl.org/.

[6] S. Winkler, *Digital Video Quality - Vision Models and Metrics*. John Wiley & Sons, Ltd, 2005.

[7] Z. Wang, L. Lu, and A. C. Bovik, "Video quality assessment based on structural distortion measurement," *Signal Processing: Image Communication*, vol. 19, no. 2, pp. 121–13, 2004.

[8] S. Wolf and M. H. Pinson, "Spatial-temporal distortion metrics for in-service quality monitoring of any digital video system," *in Proc. SPIE*, vol. 3845, pp. 266–277, 1999.