Eye Trackers in Quality Evaluation of Compressed Video

Submitted 20th June 2011; accepted 21st September 2011

Anna Ostaszewska-Liżewska, Rafał Kłoda, Sabina Żebrowska-Łucyk, Maciej Liżewski

1. Introduction

In case of video, because of the size of the source files, the use of lossy compression is almost a rule. The lossy compression is the process of discarding the part of information, that is the least important for the human perception [3]. Nevertheless the side effect of this process may occur as compression artifacts, which decrease perceived quality of the video. The final effect of compression depends on many factors such as the coding algorithm itself or its parameters used, the character of the scene (i.e. static or dynamic, amount of spatial information [4, 5], etc).

In fact, it is impossible to predict the quality of the output video, thus the only way to control it is measurement of the final results. There are two concepts of compressed video quality evaluation. The first is to use the human audience and to conduct the normalized experiment, where subjects give scores to a test material [6, 7, 8, 9]. The second idea is to automate the whole process with the use of the mathematical model of human perception [10, 11, 12, 13]. The model is created on the base of results of experiment with human audience [14, 15], so it is crucially important to make this kind of measurement as reliable and accurate as possible. The main problems in this area are scores given by unreliable viewers and the lack of a tool that would enable for linking the score with time and space of the test material. Both problems can be solved by the use of an eye tracker. Hence there is a need to obtain results for the same test material both from subjective method and an eye tracker and to develop a dedicated tool that would make such results possible to analyze.

2. Single Stimulus Continuous Quality Evaluation (SSCQE)

2.1. Methodology

Single Stimulus Continuous Quality Evaluation (SSCQE) is one of the subjective methods of compressed video quality assessment that were specified in recommendations [1, 2] developed by International Telecommunication Union (ITU). SSCQE method was specially designed to correspond to an actual home viewing situation: a series of video sequences is presented once to a viewer and a reference is not available. Subjects evaluate the instantaneous quality in real time using a slider with a continuous scale. The ratings are sampled with the frequency of 2 Hz.

The biggest advantage of this method is the relatively large amount of data – it records each temporal variation of the quality perceived.

2.2. Experiment design and carrying out

The SSCQE experiment was carried out with the use of 10-minute test material built of four 15-seconds sequences: 'bbc3', 'cact', 'mobl' and 'susi' (Fig. 1) coded in MPEG-2 standard with 10 levels of bitrate, ranging from 1 to 5 Mbps. The length of Group of Pictures (GOP) was 13 and it included two B-frames. Soundtrack was not included.

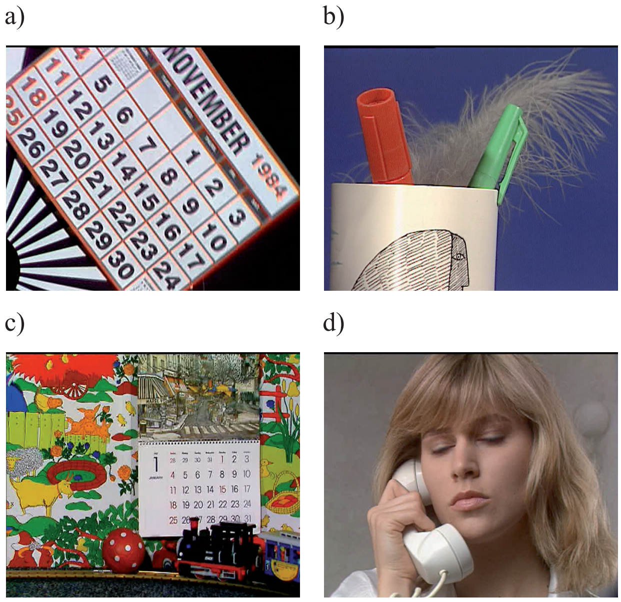

Fig. 1. Source sequences used in the test material: a) 'bbc3', b) 'cact', c) 'mobl', d) 'susi'.

Test equipment used in subjective assessments consisted of a 20" professional-grade monitor (SONY PVM-20M4E) and a professional DVD player (Pioneer DVD-V7300D). A slider, which the subjects used to evaluate the sequences, was a stand-alone hardware device with a hundred-level scale attached. SSCQE ratings were entering the PC directly through the NI 6013 card. Incoming data was synchronized with the timecode.

In the experiment 45 typical end-users (mostly university students) participated in the experiment. Each of them was screened for normal visual acuity or corrective glasses and normal color vision (per Ishihara test).

2.3. Results and temporal inconsistency of scores

The raw data were processed according to the ITU recommendations [2] to cumulative probability plot (Fig. 2) and used for bitrate optimization.

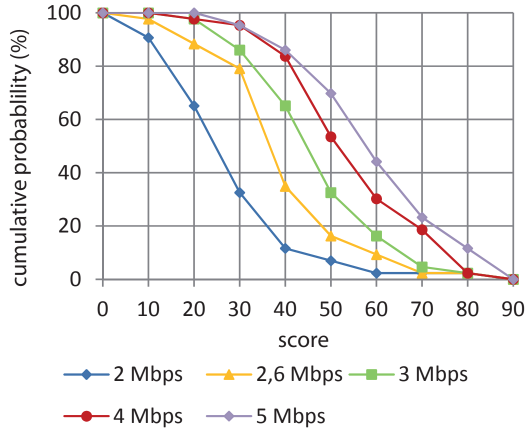

Fig. 2. Cumulative probability plot, the 'mobl' sequence.

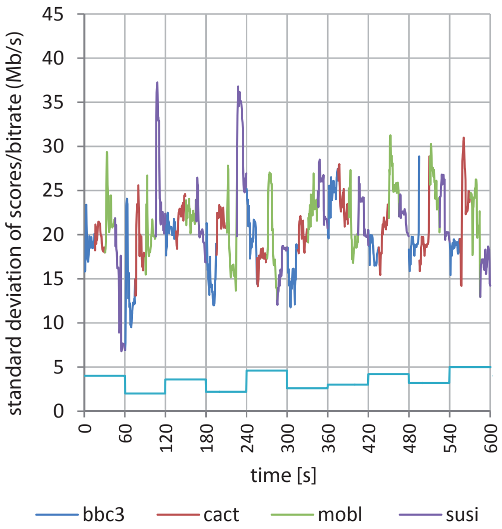

Fig. 3. Standard deviation of scores given by the audience for consecutive samples.

In general the final results were in accordance to expectations. Hence the analysis of the standard deviation of the scores given by the whole audience for each sample revealed peaks for the 'susi' sequence coded with low bitrate (2 – 3 Mbps, Fig. 3). For this particular scene, coded with bitrate less than 3 Mbps, compression artifacts were clearly visible to an experienced viewer, but they were appearing only in the 7-th second of the sequence and mainly in the area of hair. The same time the area of the face was almost free of any distortion (Fig. 4).

Fig. 4. Compression artifacts in 7th second of the 'susi' sequence coded with the bitrate of 2 Mbps, blockiness effect appears mainly in the area of hair.

Therefore a question was raised: is the high difference between individual scores caused by the fact that subjects observe the whole picture, but some of them don't perceive compression artifacts or some of viewers prefer to look just in the face of the actress and their scores relate only to that certain part of the picture.
In order to find the answer, the experiment with eye tracker was designed.

3. The experiment with eye tracker

3.1. Tobii T60 test equipment and Tobii Studio™ statistical analysis tool

Eye tracker is a device that enables for measuring eye positions and eye movement. For the purpose of the planned experiment, Tobii T60 Eye Tracker was

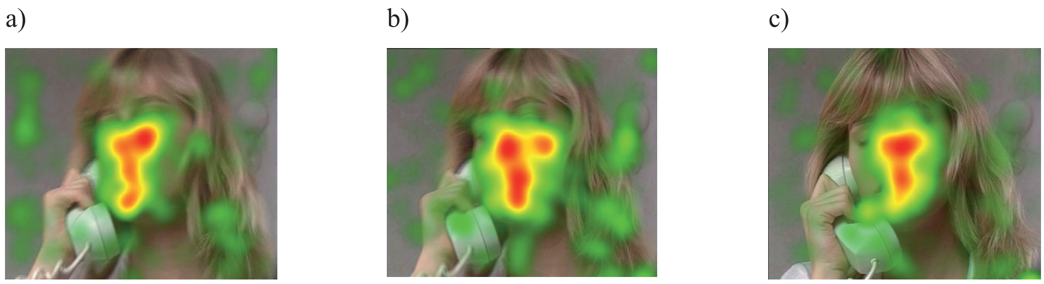

Fig. 5. Heat maps for the 'susi' sequence: a) 2 Mb/s, b) 3,6 Mb/s, c) 5 Mb/s.

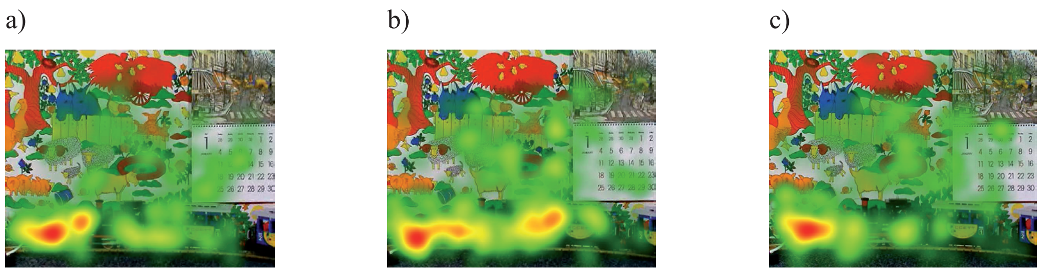

Fig. 6. Heat maps for the 'mobl' sequence: a) 2 Mb/s, b) 3,6 Mb/s, c) 5 Mb/s.

used. It is integrated in a 17-inch TFT monitor and allows for a large degree of head movement, providing a distraction-free test environment that ensures natural behavior, which is especially important in the case of SSCQE method. Eye tracker system was equipped with Tobii Studio™ - a dedicated software, which provides a platform for stimuli presentation, recording, observation, visualization and analysis of eye tracking.

3.2. The experiment design and carrying out

The test material was built on the basis of the material used in SSCQE experiment – three levels of bitrate were chosen: 2, 3 and 5 Mbps. Therefore the whole test material was only 3 minutes long.

Each test session started with eye tracker calibration for the individual subject.

16 viewers (the group similar to the group in SSCQE experiment) participated in the experiment.

3.2. Results obtained in Tobii Studio™

The first step of the data analysis was to prepare heat maps and gaze plots with the use of Tobii Studio™, for each of four sequences coded on each level of bitrate.

Heat maps were based on the summary of gaze time data from all recordings, for the whole sequence. In case of 'susi' it is clear, that for most time the girl's face was observed (Fig. 5). Although the green color appears in the area of hair also, but it is necessary to examine, if observers were looking at it when the artifacts occurred. In case of other sequences, there were also parts of the picture that turned to be more interesting to subjects (Fig. 6.), but at the same time those areas were prone to compression artifacts occurrence. Additionally comparison of heat maps for each level of bitrate proved, that the quality does not have a strong effect on the way that viewers scan the picture.

Gaze plots, which display gaze points, fixations, and scan paths superimposed over the whole of the sequence also seem to be resistant to bitrate changes. Another interesting discovery was the fact that some observers used to have episodes of distraction and they were taking a look at surroundings (Fig. 7, the scan paths which exceed the area of the monitor).

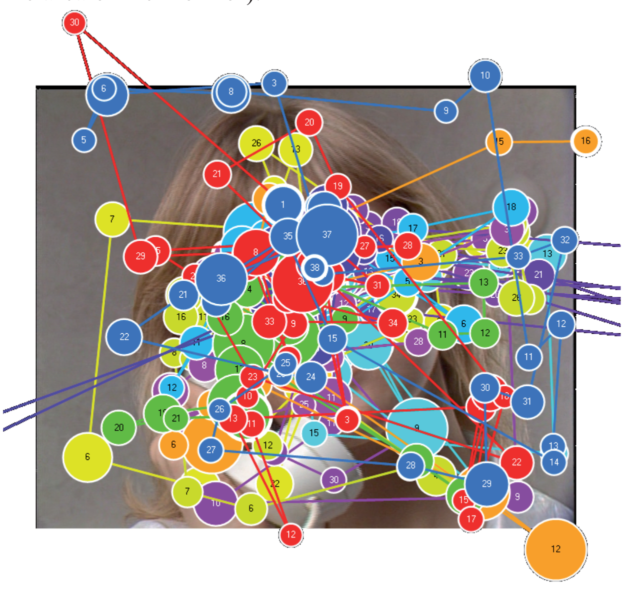

Fig. 7. Gaze plots for the 'susi' sequence, 2 Mbps.

Presented gaze plots and heat maps are a static summary of the whole data recorded for the time of sequence duration. For the purpose of data analysis in space and time, Tobii Studio™ offers animated versions of graphs. They can be used for rough assessment of the number of gaze points in a specified area of the picture (area of interest, AOI). However there is still no tool which would enable for exact calculation.

For this reason an independent application was developed, which in assumption would make possible the areas of interest move and change its shape according to the content of the video test sequence.

3.4. *Common Sense for Tobii* and the results obtained

The application called Common Sense for Tobii [16] consists of two modules. The first of them written in C++ as a plugin for VirtualDub video editing software, on the basis of the test sequence and the data from Tobii, generates the video in the original resolution of the Tobii monitor (to get the same view as during the experiment) with the animation of gaze points. There is a possibility of choosing viewers and, if necessary, not choosing any of them to get the view without any gaze points. Second part is written in C# and VideoLab library as Windows application.

Gaze points are animated – their growing diameter reflects the time of fixation. The color is matched with the particular observer. The video generated by the first module and the original data from Tobii are the input to the second module, which enables for defining animated areas of interest. Areas of interest are rectangles set in keyframes. The user controls their attributes: position, size, rotation and visibility (the areas can disappear and reappear). Tweening of areas between keyframes is used. There is a possibility of selecting the more sophisticated shape of the area of interest just by grouping several rectangles which are still controlled independently. Grouping can be nested as a tree with regions as leaves and grouping nodes.

The final result of using the application is the video with animated gazepoints and the statistics: the total number of all gazepoints and the number of the gazepoints per each area and group of areas of interest (with the information about the percentage).

CS for Tobii enabled for finding the reason for high standard deviation of scores given for 'susi' coded with low bitrate. In the first step of analysis, four independent, animated areas of interest were assigned: face, hand, hair and background. Thanks to use of 17 keyframes, the areas were changing shape and dimensions in order to adjust to object movement and perspective. The obtained statistics for three levels of bitrate are very similar (difference does not exceed 5% of the final result). The Fig. 8 presents the average of the percentage of gaze points following to each of defined, animated area of interest. For the whole duration the face was observed by 65 % of the viewers, and the hair (prone to compression artifacts) only by 11 %. In the second step, analysis concerned the aforementioned 7th second of 'susi' coded with bitrate of 2 Mbps. Two areas were defined: the face and the piece of hair that was seriously affected by blockiness effect. It appeared that even though the half of the picture for one second turned into dynamic, visible blocks, only

4% of people were puzzled by this sudden change and gazed at it. The same time 44 % were observing the fast moving face and the rest were scanning the background. This proves that only some observers are able to score the actual quality of the whole material. Most of them fail to take into account even very severe distortions, which is the reason for significant differences between scores given by individual subjects.

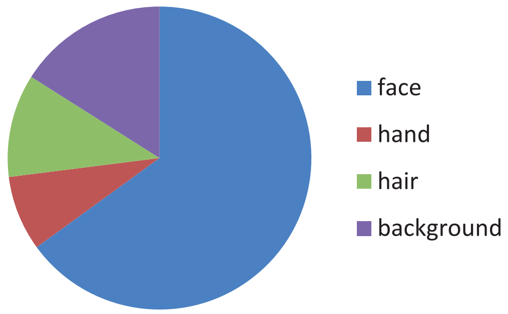

Fig. 8. The number of gaze for the 'susi' sequence, 2 Mbps.

4. Conclusion

To sum up, the experimental station for conducting subjective quality evaluation should be integrated with the eye tracker. That would enable for rejecting scores given at the time of the episode of lack of attention, when the test material was not observed. This kind of data filtration would make the final result more reliable.

Besides it would be useful to link up the score with the part of the picture that the observer was looking at, as the quality of the frame hardly ever is uniform. This would provide researchers working on human visual system with interesting data both on human perception and the usability of certain sequence for quality evaluation of compressed video.

AUTHORS

Anna Ostaszewska-Liżewska* – Institute of Metrology and Biomedical Engineering, Warsaw University of Technology, Warsaw, Poland. anna.ostaszewska@gmail.com

Rafał Kłoda – Institute of Metrology and Biomedical Engineering, Warsaw University of Technology and Industrial Research Institute for Automation and Measurements PIAP, Warsaw, Poland. E-mail: kloda@mchtr.pw.edu.pl

Sabina Żebrowska-Łucyk – Institute of Metrology and Biomedical Engineering, Warsaw University of Technology, Warsaw, Poland. E-mail: szl@mchtr.pw.edu.pl

Maciej Liżewski – 3e Internet Software House, CIO, ul. Podbipięty 51, 02-732 Warsaw, Poland. E-mail: maciej.lizewski@3e.pl

References

[1] ITU-T Recommendation P.911, Subjective audiovisual quality assessment methods for multimedia applications,1996.

[2] ITU-R BT.500-12 Methodology for the subjective assessment of the quality of television pictures, 2009.

[3] K. Sayood, *Introduction to Data Compression*, Morgan Kaufmann, 3rd ed., 2005.

[4] A. Ostaszewska, R. Kłoda, "Quantifying the amount of spatial and temporal information in video test sequences", *Recent Advances in Mechatronics*, Springer Verlag, Berlin Heidelberg 2007, pp. 11–15.

[5] R. Aldridge, J. Davidoff, M. Ghanbari, D. Hands, D. Pearson, "Measurement of scene-dependent quality variations in digitally coded television pictures", *IEE Proceedings – Visual Image Signal Process*, vol. 142, no. 3, June 1995.

[6] Th. Alpert, J.P. Evain, "Subjective quality evaluation – The SSCQE and DSCQE methodo-logies", *EBU Technical Review*, Spring 1997.

[7] F. Kozamernik, P. Sunna, E. Wyckens, D.I. Pettersen, "Subjective quality of Internet video codecs", *EBU Technical Review,* no. 1, 2005, pp. 1-22.

[8] M. Pinson, S. Wolf, "Comparing subjective video quality testing methodologies", *SPIE Video Communications and Image Processing Conference*, Lugano, Switzerland, Jul. 8-11, 2003.

[9] R. Aldridge, D. Hands, D. Pearson, N.K. Lodge, "Continuous quality assessment of digitally-coded television pictures", *IEE Proceedings online,* no. 1998, 116-123.

[10] M.P. Eckert, A.P. Bradley, "Perceptual quality metrics applied to still image compression", *Signal Processing*, 1998, pp. 177-200.

[11] P.N. Gardiner, M. Ghanbari, D.E. Pearson, K.T. Tan, "Development of a perceptual distortion meter for digital video", *IEE Conference Publication Proceedings of the 1997 International Broadcasting Convention*, vol. 1, no. 447, Amsterdam, 1997, pp. 493–497.

[12] D.J. Heeger, P.C. Teo, "A model of perceptual image fidelity", *Proceedings of the International Conference on Image Processing*, Washington, DC, 23rd–26th October 1995, pp. 343–345.

[13] M.A. Masry, S.S. Hemami, "CVQE: A metric for continuous video quality evaluation at low bit rates", *Proceedings of the SIGCHI conference on Human factors in computing systems*, Vienna, Austria, 2004, pp. 535–542.

[14] D.S. Hands, M.D. Brotherton, A. Bourret, D. Bayart, "Subjective quality assessment for objective quality model development", *Electronic Letters*, 2005, vol.41, no. 7, pp. 408-409.

[15] J. Lubin, "Human vision system model for objective picture quality measurements", *IEE Conference Publication Proceedings of the 1997 International Broadcasting Convention*, vol. 447, Amsterdam, 1997, pp. 498–503.

[16] http://www.common-sense.com.pl