VISUAL ARTIFACTS INTERFERENCE UNDERSTANDING AND MODELING (VARIUM): A PROJECT OVERVIEW

Mylène C.Q. Farias[a], Ingrid Heynderickx[b,c], Bruno Machiavello[a], Judith A. Redi[b]*
[a] University of Brasília, Campus Universitário Darcy Ribeiro, Brasília - DF, Brazil, 70919-970;
[b] Delft University of Technology, Mekelweg 4, Delft, The Netherlands, 2628 CD;
[c] Philips Research Laboratories, Prof. Holstlaan 4, Eindhoven, The Netherlands, 5656 AE

ABSTRACT

In this paper, we present our approach in the project entitled "Visual ARtifacts Interference Understanding and Modeling (VARIUM)", currently being developed at University of Brasília (UnB) and Delft University of Technology (TUD). In this project, we aim at designing an objective metric for overall quality of digital video that takes into account specific spatial and temporal artifacts, their impact and mutual interactions for a broad range of video content. Our approach relies on first understanding the impact of various digital artifacts is isolation, and then combining them to evaluate their interaction. As a first step, we here present some results on studying the visibility and annoyance of "packet loss" artifacts in isolation of other digital artifacts.

1. INTRODUCTION

In modern digital imaging systems, the quality of the visual content can undergo a drastic decrease due to impairments introduced during capture, transmission, storage, and/or display, as well as by any signal processing algorithm that may be applied to the content along the way (e.g., compression.). Impairments are defined as visible defects (flaws) and can be decomposed into a set of perceptual features called *artifacts* [1, 2, 3]. Being able to detect artifacts and improve the quality of the visual content prior to its delivery to the user is therefore crucial to ensure a good quality of experience. At the basis of such a quality control mechanism, is an (automated) visual quality assessment system.

The most accurate way to determine the quality of a video is by using psychophysical experiments with human subjects [2]. Unfortunately, these are very expensive, time-consuming and hard to incorporate into a design process or an automatic quality of service control. Therefore, there is a great need for o*bjective quality metrics*, i.e., algorithms that can predict visual quality as perceived by human observers.

Quality metrics that analyze visible differences between a test and a reference signal, taking into account aspects of the human visual system (HVS), usually have the best performance [4-5], but are often computationally expensive and therefore hardly applicable in real-time contexts. Alternatives to these metrics are *artifact metrics*, which estimate the strength of the most perceptually relevant artifacts. Artifact metrics have the advantage of being simple and not necessarily requiring the reference signal. Also, they can be useful for post-processing algorithms, providing information about which artifacts need to be mitigated.

One disadvantage of artifact metrics is that their design requires a good understanding of the perceptual characteristics of each artifact. A second disadvantage is that the artifact metrics need to be combined to an overall quality estimate [6]. The latter cannot be done by simply linearly combining the estimated strength of each artifact, since masking among artifacts or other mutual interaction effects may occur. Both disadvantages are hampered by technological limitations in digital media at the point of delivery, namely the co-occurrence of different artifacts. For example, temporal artifacts caused by transmission errors can appear in videos already containing blockiness and blurriness artifacts, as a consequence of compression.

To address these disadvantages we propose a two-step approach: (1) evaluate the annoyance and visibility of digital artifacts in isolation, what can be achieved by generating 'pure' artifacts from high-quality videos, and (2) evaluate masking and interaction effects of these (isolated) artifacts when presented in combination to better understand how these artifacts combine to produce overall annoyance. This approach was already used in [7-8] to study the visibility, annoyance, and interaction of blockiness, ringing, noisiness, and blurriness and their relation with spatial content. In this paper, we extend this work to include temporal artifacts.

The paper is divided as follows. First, we explain our approach and the related project activities in more detail. Then, we report preliminary results on the visibility and annoyance of a typical temporal artifact, "packet loss", and briefly compare our results to studies that have evaluated packet loss artifacts in highly compressed videos (i.e., containing a mixture of digital artifacts).

2. VARIUM APPROACH

Figure 1 shows an overview of the structure of the project, which is divided in work packages (WP). As a starting point for designing an objective metric that is robust to the co-presence of multiple artifacts, we will perform a series of psychophysical tests [3] to collect information on the perceptual characteristics of artifacts when presented alone or in combination. At the core of the VARIUM project (WP1) is therefore a series of subjective experiments aimed at gathering information about the *visibility*, *annoyance* and *description* (E1), and determining their *perceptual strengths* (E2).

The *visibility* of an artifact refers to whether the artifact is noticed within the content. It is defined according to a visibility threshold, which corresponds to the distorting signal strength that allows 50% of the observers to notice the artifact. The *annoyance* of an artifact is a measure of the degradation of the visual content when the distorting signal strength is above threshold. Observers can also be trained to recognize specific artifacts when combined (*description*) and estimate their *perceptual strength*. This way, we can have an idea of how the visibility and annoyance are being affected by the video content and the presence of other artifacts.

In a second stage, to connect visibility, annoyance, description and perceptual strength information to the overall appearance of the combined artifacts, we will measure overall quality scores of videos impaired with different combinations of artifacts at different strengths (WP2). An adaptation of Keelan's quality ruler [2] will be used for quantifying overall video quality. While performing all the above experiments, we will also evaluate the impact of (combined) artifacts on viewing behavior and visual attention (WP3). Such information has been shown to be highly relevant in visual quality assessment [9]. As a consequence, we will record eye-movements throughout the planned experiments. The collected information will eventually form the basis for the design of an effective video quality metric that is robust to combined artifacts (WP4).

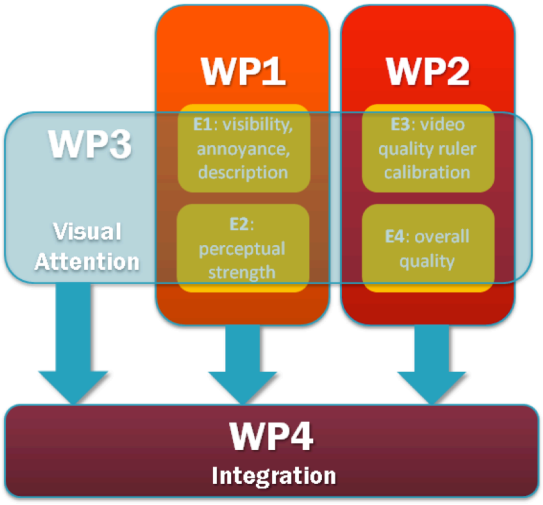

Fig. 1. Schematic representation of the planned work and division of the tasks.

3. VISIBILTY AND ANNOYANCE OF PACKET LOSS ARTIFACTS

In video transmission over IP networks, the network variability and the lack of service guarantees represent a big challenge. Transmission errors may occur due to network congestion and path loss. Typical impairments caused by these errors are packet loss, jitter, and delays. Among these, packet loss is probably the most annoying artifact. As the name suggests, packet loss artifacts are caused by a complete loss of the packet being transmitted.

Typically, for block-based video compression schemes (e.g. MPEG-1/2/4, H-261/2/3/4), consecutive macroblocks in a frame are transmitted as a slice in a single network packet. Therefore, the loss of network packets results in a loss of macroblocks. Because the compression process removes a lot of spatial and temporal redundancies from the original video, and because of the use of motion-compensated temporal prediction, a single loss of a packet can affect many subsequent frames. Therefore, packet loss artifacts are visually characterized by the presence of rectangular areas distributed over the video frames, whose content differs from the surrounding areas.

The visibility and annoyance of packet-loss artifacts depend heavily on how the video stream has been coded, how it has been mapped into flows and packetized, and what type of error concealment algorithm is being used. In the literature, there is a considerable amount of work on the visibility of the packet-loss, as summarized by Boulos *et al.* [10]. Some literature also investigated the effect of scene characteristics on the visibility of packet loss [11]. Some studies have attempted to address the visibility and annoyance of packet loss artifact [12, 13].

In [12], the authors show that the annoyance of packet loss artifacts is correlated with their length (propagation throughout frames) and with the severity of the losses (PSNR), whereas their visibility seems not to be related to the length of the loss itself, but rather to the overall degradation of the video. However, these results are based on the subjective evaluation of degraded versions of a single video content and both visibility and annoyance are

Fig. 2. Screenshots of the first frame of the sequences included in Experiment 1 (E1).

not analyzed in relation to the spatial and temporal characteristics of the video. Furthermore, a loss is considered visible if it generates a drop in visual quality; whereas it might be argued that a loss might be visible and yet not generate annoyance (and quality loss as a consequence). The study in [13] relates the visibility of packet loss artifacts to the percentages of slices lost, the type of frames where the loss happened (I, B or P), the duration of the loss and the amount of motion in the video. Unfortunately, the analysis is not extended to the annoyance of visible artifacts.

In most studies, packet loss artifacts are generated by varying parameters of compression algorithms (codec type, bitrate, etc.) and digital transmissions (loss rate, channel model, etc.). As a consequence, the generated videos contain compression artifacts (e.g., blockiness, blurriness, and ringing) besides the packet loss artifacts. Therefore, these procedures cannot be used in this project to study the perceptual characteristics of artifacts. In the following, we report the preliminary results of an experiment aimed at relating visibility and annoyance of packet loss artifacts to the temporal and spatial properties [14] of different video contents.

3.1. Experimental Methodology

We used seven high-definition (1920x720, 50 fps) of ten seconds that corresponded to a diverse content, as depicted in Fig. 2. Figure 3 shows Spatial (SI) and Temporal (TI) perceptual information measures (computed as per [14]) for all videos. To avoid inserting additional artifacts, we compressed the original videos at a high bitrate and used the H.264 standard error concealment algorithm, generating videos with Peak Signal to Noise Ratio (PSNR) well above 70dB. We also varied the frame intervals (M) between I-frames with the goal of having artifacts with different time duration. We used M = 4, 8, and 12 frame intervals, no P frames, and 8 slices per frame. Then, we randomly deleted packets from the coded video bitstream, varying the percentage of deleted packets from 0.5% to 9%. For each original, we had 91 test sequences.

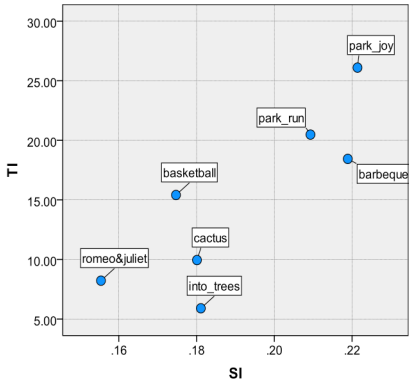

Fig. 1. Temporal and spatial characteristics of the videos included in the experiment

The experiment tested one subject at a time using a PC computer and a Samsung LCD monitor of 23 inches. The room where the experiment was performed had illumination conditions compliant to ITU-T Recommendation BT.500-8 [3]. The subject was seated straight ahead of the monitor, centered at or slightly below eye height for most subjects. The distance between the subject's eyes and the video monitor was 3 times the monitor screen's height. A chin rest was used to guarantee a constant distance between the subject's eyes and the monitor.

Fifteen subjects from Delft University of Technology participated in the experiment. They were considered naïve to most kinds of digital video defects and the associated terminology. They were asked to wear glasses or contact lenses if they needed them to watch TV. A test session was broken into five stages. In the first stage, the subject was verbally given instructions. In the second stage, we showed examples of original and highly impaired videos to establish the range of annoyance used in the experiment. In the third stage, the subject carried out practice trials to allow the responses to stabilize. The fourth stage was the main experiment. At the last stage, we asked the subject for qualitative descriptions of the impairments.

The main experiment was performed with the set of test sequences presented in random order. After each test sequence was played the subject was asked "Did you see a defect or an impairment?", prompting for a 'yes' or 'no' answer (detection task). Then participants were asked to perform the annoyance task consisting of giving a numerical judgment of how annoying the detected impairment was. Any defect as annoying as the worst impairment shown in the second stage of the experiment should be given '100', half as annoying '50', ten percent as annoying '10', etc.

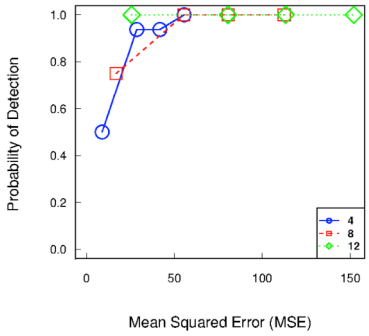

Fig. 3. Probability of Detection for video 'Park Joy'.

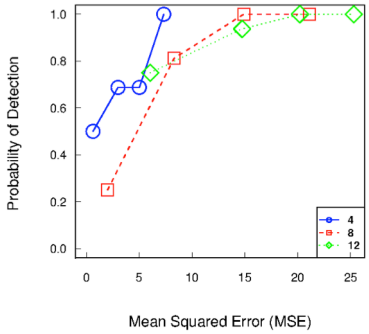

Fig. 4. Probability of Detection for video 'Park Run'.

4. EXPERIMENTAL RESULTS

To estimate the visibility of the packet loss artifact, we calculated the probability of detection by dividing the number of subjects that detected the artifact by the total number of subjects. Figures 4 and 5 show the probability of detection for two sample test sequences, i.e., 'Park Joy' and 'Park Run'. The *x* axis in the graphs corresponds to the Mean Squared Error (MSE) between the original and the impaired video, while the *y* axis corresponds to the Probability of Detection. The different curves in the graphs correspond to different values of M (4, 8 or 12).

For the videos 'Into Trees' and 'Barbecue', all values of the probability of detection were equal to '1', i.e. *every* subject of the pool was able to detect the artifact in all test sequences. These two videos had camera movements and large smooth light areas (e.g., sky areas in 'Into Trees' and concrete areas in 'Barbecue' as shown in Fig. 2), what might have made the artifacts in these scenes easier to detect. The videos 'Park Joy' (see Fig. 4), 'Cactus', and 'Basketball' had probabilities of detection curves that increased very fast with the MSE. This means that artifacts in these videos were also relatively easy to detect. The videos 'Romeo and Juliet' and 'Park Run' (see Fig. 5), on the other hand, had a less steep slope for the probability of detection, and so, in these contents, the artifacts were harder to detect. The video 'Romeo and Juliet' is a relatively dark video with a clear focus of attention (i.e., the couple in the middle of the scene). The video 'Park Run' has a lot of spatial details (i.e., the crowd) and temporal activity and not a lot of camera movement. Note that the steepness of the slope in the probability of detection is not straightforwardly related to the video characteristics SI and TI in Figure 3. This result is somewhat in contrast with that of [13], where the authors found that low-motion content (low TI) was more capable to conceal packet losses.

To get insight in the results of the annoyance task, the Mean Annoyance Value (MAV) was calculated by averaging the annoyance score over all observers for each test video. Figures 6-8 show the MAV as a function of MSE for the videos 'Joy Park', 'Park Run', and 'Barbecue'. Notice that, as expected, the higher the MSE, the higher the MAV. Again, the graphs show three curves, corresponding to the three different frame intervals (i.e., M = 4, 8 1n 12). As expected, the larger the value of M, the higher the value of MAV, consistently with what found in [12, 13]. For some of the videos ('Barbecue', and 'Romeo and Juliet) the MAV curves for M = 8 and 12 are very similar (see Fig. 8), i.e. subjects did not notice a difference in quality between artifacts appearing with different time intervals. Notice also that, the video 'Barbecue', which had probability of detection equal to '1', had annoyance scores higher than the annoyance scores given to other videos (i.e., compare Fig. 8 with Figs. 6 and 7).

5. CONCLUSIONS

In this paper, we presented our approach in the project entitled VARIUM, which has the goal of understanding the characteristics of relevant digital artifacts, their interactions, and their relationship with content. In particular, in this paper we reported the first results on the visibility and annoyance of a typical temporal artifact, "packet loss", and showed its (non-trivial) interactions with spatial and temporal characteristics in the video.

10. REFERENCES

[1] M. Yuen and H. R. Wu, "A survey of hybrid MC/DPCM/DCT video coding distortions," Signal Processing, vol. 70, pp. 247 - 278, October 1998.

[2] B. Keelan, "Handbook of image quality: characterization and prediction," Marcel Dekker, Inc., New York, 2002

[3] Internat. Telecom. Union, "ITU-T Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures," 1998.

[4] W. Lin, C.-C. Jay Kuo, Perceptual Visual Quality Metrics: A Survey. J. Vis. Commun. (2011).

[5] A.K. Moorthy and A.C. Bovik, "Visual Quality Assessment Algorithms : What Does the Future Hold?" Intern. Journal of Multimedia Tools and Applic., Vol: 51 No. 2, Feb. 2011, pp. 675-696 .
[6] J. Caviedes and J. Jung, "No-Reference Metric for a Video Quality Control Loop," Proc. 5th World Multiconf. on Systemics, cybernetics, and Informatics, July 2001, vol. 13, Part 2, pp. 290-5.
[7] Farias, Mylene C. Q., Foley, John M, Mitra, Sanjit Koumar; "Detectability and Annoyance of Synthetic Blocky, Blurry, Noisy, and Ringing Artifacts." IEEE Trans. on Signal Processing, v. 55, p. 2954-2964, 2007.
[8] M. S. Moore, J. M. Foley, and S. K. Mitra, "Defect visibility and content importance: Effects on perceived impairment," Image Communication, vol. 19, pp.185-203, Feb. 2004.
[9] U. Engelke, H. Kaprykowsky; H.-J. Zepernick and P. Ndjiki-Nya; "Visual Attention in Quality Assessment," IEEE Signal Processing Magazine, vol. 6, pp. 50-59, 2011
[10] F. Boulos, B. Parrein, P. Le Callet, D. Hands, "Perceptual Effects of Packet Loss on H.264/AVC Encoded Videos", Fourth International Workshop on Video Processing and Quality Metrics for Consumer Electronics (VPQM), 2009.
[11] Amy R. Reibman, David Poole, "Predicting packet-loss visibility using scene characteristics," Packet Video, pp. 308-317, 12-13 Nov. 2007.
[12] T. Liu, Y. Wang, J. Boyce, H. Yang; Z. Wu, "A Novel Video Quality Metric for Low Bit-Rate Video Considering Both Coding and Packet-Loss Artifacts," Selected Topics in Signal Processing, IEEE Journal of, vol.3, no.2, pp.280-293, April 2009
[13] N. Staelens, G. Van Wallendael, K. Crombecq, N. Vercammen, J. De Cock, B. Vermeulen, R. Van de Walle, T. Dhaene, P. Demeester, "No-Reference Bitstream-Based Visual Quality Impairment Detection for High Definition H.264/AVC Encoded Video Sequences," Broadcasting, IEEE Trans. on , vol.58, no.2, pp.187-199, June 2012
[14] A. Ostaszewska and R. Kloda, "Quantifying the amount of spatial and temporal information in video test sequences," in Recent Advances in Mechatronics, Springer, 2007, pp. 11–15.

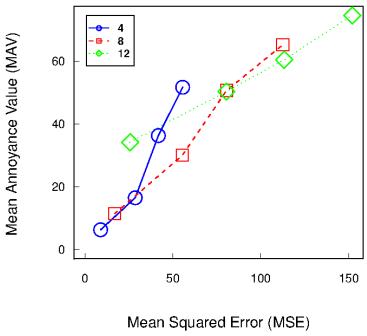

Fig. 5. MAV for video 'Joy Park'.

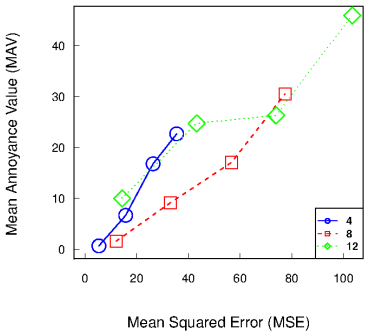

Fig. 6. MAV for video 'Park Run'.

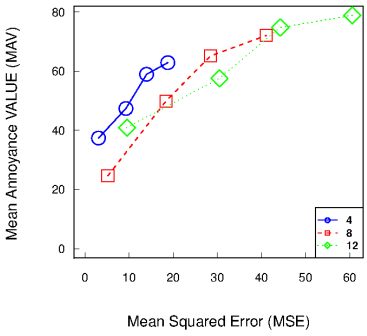

Fig. 7. MAV for video 'Barbecue'.