VISUAL ARTIFACTS INTERFERENCE UNDERSTANDING AND MODELING (VARIUM)

Mylène C.Q. Farias[a], Ingrid Heyinderickx[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 the preliminary results obtained 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 are interested in understanding the characteristics of relevant digital artifacts in video, their interactions, and their relationship with content. We aim at designing an objective metric for overall video quality that takes into account specific spatial and temporal artifacts, their mutual impact and importance for a broad range of video content. In particular, in this paper, we report the results of our first experiment, which had the goal of studying a typical temporal artifact, "packet loss", by measuring its visibility and annoyance.

1. INTRODUCTION

Digital transformation of images and video offers many advantages over the existing analog methods. The advantages of digital visual material, however, do not come without some disadvantages as well. The quality of digital content may decrease when impairments are introduced during capture, transmission, storage, and/or display, as well as by any signal processing algorithm that may be applied along the way (e.g., compression, etc.). Impairments are defined as visible defects (flaws) and can be decomposed into a set of perceptual features called *artifacts* [1, 2, 3].

Spatial artifacts are characterized by the presence of degradations that vary (mainly) within the spatial domain. Examples of spatial artifacts include blockiness, blurriness, ringing, noisiness, mosaic patterns, etc. *Temporal* artifacts are degradations that vary across the temporal domain. Examples include packet-loss artifacts, motion compensation mismatches, mosquito effects, ghosting, smearing, jerkiness, and so on.

The most accurate way to determine the quality of a video is by measuring it using psychophysical experiments with human subjects [2]. Unfortunately, psychophysical experiments 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.* Objective metrics are algorithms that are used to (1) predict visual quality as perceived by human observers, (2) compare the performance of video processing system, and (3) optimize algorithms and parameters settings for a video processing system.

Quality metrics with best performances are the ones that analyze visible differences between a test and a reference signal, taking into account aspects of the human visual system (HVS) [4-5]. However, these metrics are often computationally expensive and hardly applicable in real-time contexts. One possible alternative is to use a feature extraction approach, which looks for higher-level features of the content that are considered relevant to quality (e.g., sharpness or blurriness, contrast, fluidity, artifacts, etc.). Popular types of feature extraction 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 characteristics of the artifacts. Indeed, most metrics exploit knowledge on the perceptual annoyance the occurrence of a specific artifact causes to the final user. A second disadvantage is that the artifact metrics need to be combined to obtain an overall quality estimate [6]. In fact, due to technological limitations, co-occurrence of different artifacts is highly likely in digital media at the moment of delivery. For example, packet loss artifacts can appear in videos already bearing blocky artifacts (from compression), to which also blurriness is added as a consequence of e.g. transcoding. The effect on perceived quality of the combination of artifacts can hardly be predicted by the linear combination of the annoyance estimated for the single artifacts. Masking and other interaction effects can occur, making the prediction strategy more complex and very dependent on the artifacts involved. Unfortunately, little work has been done on studying and characterizing artifacts and on combining them to an overall metric, as pointed out in [6]. Some research was done on studying the visibility, annoyance, and interaction of blockiness, ringing, noisiness, and blurriness and on relating them to the spatial content, as discussed in [7-8].

In this paper, we describe the project entitled "Visual ARtifacts Interference Understanding and Modeling (VARIUM)", which is currently being developed at University of Brasília (UnB) and Delft University of Technology (TUD). In this project, we are interested in further extending the first attempts of understanding the characteristics of relevant digital artifacts and their relationship with content. Our final goal is to develop an objective metric for overall video quality that takes into account specific spatial and temporal artifacts, their mutual impact and their mutual importance for a broad range of video content. In particular, in this paper we also report the results of our first experiment, which had the goal of studying a typical temporal artifact, "packet loss", by measuring its visibility and annoyance.

2. PROPOSED WORK

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 need to collect information on their perceptual impact when they interfere with each other. The most effective way to collect such information is to conduct a campaign of psychophysical tests [3] on impaired videos. At the core of the VARIUM project is therefore a series of subjective experiments aimed at gathering information about the *visibility*, *annoyance*, *description*, and *perceptual strengths* of artifacts, and at determining the *relative importance* of the artifacts to the overall perceived quality.

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, and it is dependent on the visibility and on the distorting signal strength. 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 the VARIUM project we will study the above mentioned perceptual characteristics for several artifacts (WP1 in fig. 1), first independently from one another (i.e., when not interfering with other artifacts, Experiment 1 or E1 in fig. 1) and then in combinations (E2 in fig. 1). 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).

In this paper, we report the results of our first experiment (E1), which is part of WP1. This experiment has the goal of studying a typical temporal artifact, "packet loss", by measuring its visibility and annoyance.

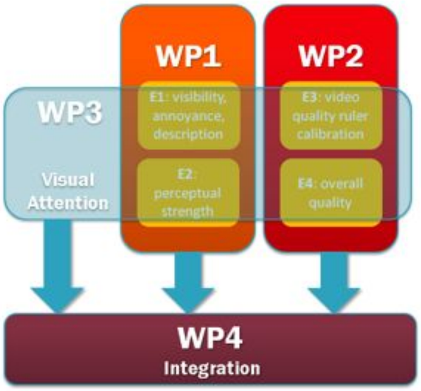

Fig. 1. Schematic representation of the planned work and division of the tasks.

3. ANNOYANCE AND VISIBILITY OF PACKET LOSS ARTIFACTS

In video transmission over IP networks, the network variability and the lack of service guarantees represent a big challenge. Video packets typically traverse a number of links to get to its destination. Losses (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, as a consequence of transmission errors.

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, every packet is important for the video reconstruction. Moreoever, 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 contents differ from the surrounding areas.

The visibility and annoyance of packet-loss impairments 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 this section, we present the results of a study conducted to investigate the perceptual properties of such artifacts.

3.1. Experimental Methodology

We used seven high-definition videos with spatial resolution of 1920x720 and temporal resolution of 50 frames per second (fps). The videos were all ten seconds long and were chosen with the goal of generating a diverse content, as it can be seen by examining the first frames of the originals depicted in Fig. 2. Figure 3 shows spatial and temporal perceptual measures for all videos.

To generate test sequences with several levels of packet loss artifacts, we used the reference H.264 codec. To avoid inserting additional artifacts (such as ringing, blurriness, and blockiness), we compressed the original videos with high bitrates 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 durations. We used M = 4, 8, and 12 frame intervals. 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 4 (percentages) x 3 (frame intervals) = 12 stimuli, generating a total of 13 (12 stimuli + original) x 7 (originals) = 91 test sequences.

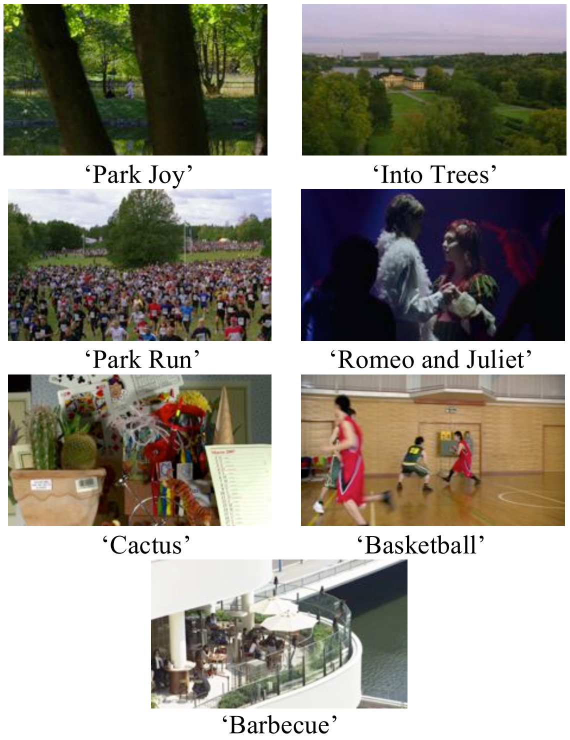
‘Park Joy’ ‘Into Trees’
‘Park Run’ ‘Romeo and Juliet’
‘Cactus’ ‘Basketball’
‘Barbecue’

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

The experiment was run with one subject at a time using a PC computer and a Samsung LCD monitor of 23 inches (Sync Master XL2370HD). The dynamic contrast of the monitor was turned off and the contrast was set at 100 and the brightness at 50. The software **Presentation®** from **Neurobehavioral Systems Inc.** was used to run the experiment and record the subjects’ data. 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 height. We used a chin rest to guarantee that the distance between the subject’s eyes

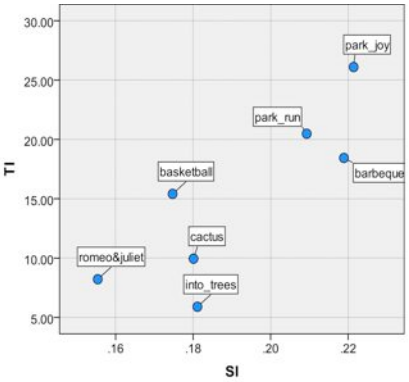

Fig. 3. Temporal and spatial characteristics of the videos included in the experiment

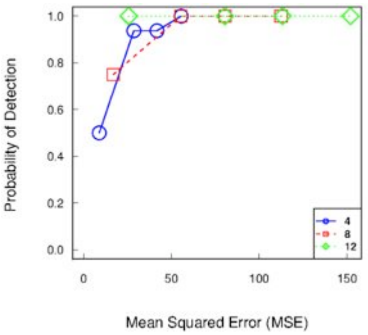

Fig. 4. Probability of Detection for video 'Park Joy'.

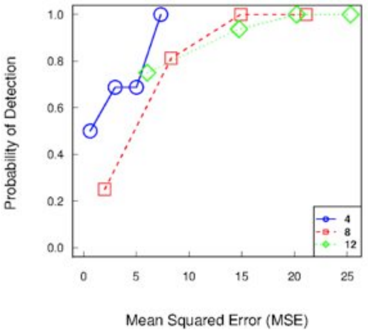

Fig. 5. Probability of Detection for video 'Park Run'.

and the monitor remained constant.

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. Subjects were asked to detect the impairment in the test sequence (detection task). 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. 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.

4. EXPERIMENTAL RESULTS

To estimate the visibility of the packet loss artifact, we first calculated the probability of detection of the artifact in the test sequences by dividing the number of subjects that detected it by the total number of subjects. In Figures 4 and 5, we show graphs of 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 (i.e., 4, 8 or 12).

For the videos 'Into Trees' and 'Barbecue', the values of the probability of detection were equal to '1' for all test cases, i.e. *every* subject of the pool was able to detect the artifact in all test sequences for these two originals. 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.

For the videos 'Romeo and Juliet' and 'Park Run' (see Fig. 5), on the other hand, the probability of detection curves had a less steep slope. This might indicate that, in these originals, the artifacts were harder to detect. The video 'Romeo and Juliet', for example, is a relatively dark video with a clear focus of attention (i.e., the couple in the middle of the scene). All of this makes it harder to spot the artifacts. In the case of the video 'Park Run', there are a lot of spatial details (i.e., the crowd) and temporal activity and not a lot of camera movement. Therefore, it is again *not* easy to spot the artifacts.

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. In Figures 6-8, we show the graphs of MAV 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.

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). This may indicate that there could be a correlation between visibility and annoyance.

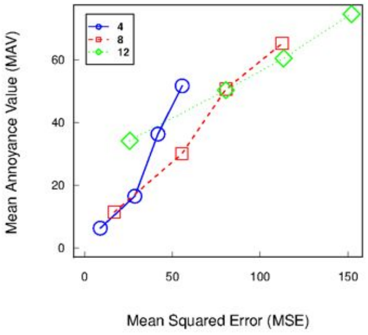

Fig. 6. MAV for video ‘Joy Park’.

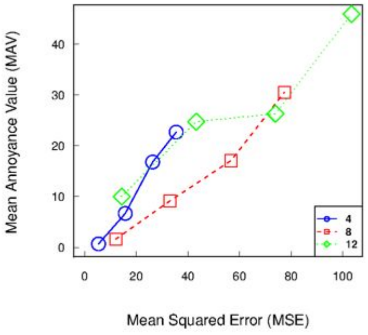

Fig. 7. MAV for video ‘Park Run’.

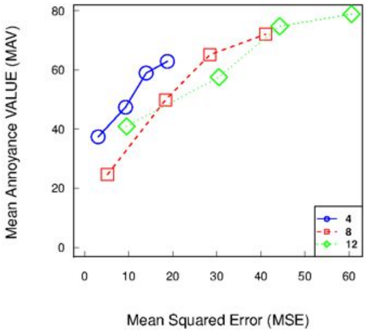

Fig. 8. MAV for video ‘Barbecue’

5. CONCLUSIONS

In this paper, we presented preliminary results obtained 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 results of our first experiment, which studied a typical temporal artifact, “packet loss”, by measuring its visibility and annoyance.

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] International Telecommunication 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?” International Journal of Multimedia Tools and Applications, Vol: 51 No: 2, February 2011, Page(s): 675-696

[6] J. Caviedes and J. Jung, "No-Reference Metric for a Video Quality Control Loop," Proc. 5th World Multiconference 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