REPORT: Analysis of the influence of different artifacts in the video content from three experiments.

Alexandre Fieno da Silva[1], Judith A. Redi[2], and Mylène C.Q. Farias[1]

[1]University of Brasilia, Campus Universitario Darcy Ribeiro, Brasilia - DF, Brazil, 70919-970
[2]Delft University of Technology, Mekelweg 4, Delft, The Netherlands, 2628 CD

In this project, we are interested in understanding the impact of a set of artifacts on the quality of video, their relationship with the content and to determine their impact on visual quality perception. The set of artifacts used in this study was chosen among those most perceptually relevant for digital video applications (e.g. blockiness, blurriness, packet-loss etc). We first plan to perform a series of psychophysical experiments on visual quality perception. These experiments will determine the impact on quality first of a set of spatial and temporal artifacts by themselves, and later of their combinations, taking into account different content with varying spatial and temporal characteristics.

Index Terms—**Quality of video, visual quality perception, blockiness, blurriness and packet loss.**

I. INTRODUCTION

ONE approach to design video quality metrics is to use artifact metrics. They have several advantages, for example simplicity and the ability to provide a good description of what is wrong (or right) with the video. But, in order to design good artifact metrics or any quality metric, it is necessary to have a good understanding of how artifacts affect quality. Therefore, it is important to study the characteristics of the most relevant artifacts present in digital video, what includes characteristics of annoyance, visibility, interactions among artifacts and content dependence. This way, we can use this data together with the data gathered previous to determine annoyance, visibility, and interaction for a diverse content.

In this experiment, we focus on 3 particular artifacts that can be considered among the most relevant for digital video: blurriness, blockiness, and packet loss. We present these artifacts at several strengths, in several combinations and for seven original videos. In summary, we are interested in understanding ”how” artifacts combine to determine annoyance and degradation in order to design better metrics.

A. *Previous Experiments*

1) *Experiment 1*

In Experiment 1 (*Exp.* 1), all videos were encoded using the $H.264AVC$ codec at a very high bitrate (approximately 120 Mbps), with 8 packets per frame. The coding process generated three sequences per video, with a Group of Pictures (GOP) size set to 4, 8 and 12, respectively. Packet loss artifacts were then generated by dropping data packets from the bitstream [2]. The packet loss ratios for all videos were 0.7%, 2.6%, 4.3% and 8.1%. In total, it was generated 7 videos x 3 GOP x 4 packet loss ratios = 84 sequences. The 7 pristine videos were added to the set, making a total of 13 different settings per original video and resulting in 91 test sequences [2].

2) *Experiment 2*

Experiment 2 (*Exp.* 2) is composed of two artifacts (blockiness and blurriness) considered the most salient present in digital videos [8]. This experiment follows the ITU-T Recommendation P.930 that suggests to introduce blockiness only in regions where these artifacts are more visible. Thus, blockiness artifacts were produced by using the difference between the average of each 8×8 block and the average of the 24×24 surrounding area (which has the current 8×8 block as a center) to make each block stand out. Recommendation P.930 suggests the generation of blurriness with the use of a low-pass filter [8]. Thus, blurriness artifacts were generated by applying a symmetric, two dimensional finite duration impulse response (FIR) low-pass filter to the digital image array. Different filters with varying cut-off frequencies could be used to control the amount of blurriness introduced. The blockiness and blurriness strengths for all videos were 0.4, 0.6 and 0.8. In total, it was generated 7 videos $\times$ 10 blockiness and blurriness strengths = 70 sequences. The 7 pristine videos were added to the set, resulting in 77 test sequences [4].

3) *Experiment 3*

In Experiment 3 (*Exp.* 3) has been considered all original videos of the previous experiments. Also, it is composed of the three artifacts has been used in previous experiments (i.e., blockiness, blurriness and packet loss). In Exp. 3 we are interested to understanding the impact on perceived quality of combinations of these artifacts and their relationship with the content. In total, we used 140 test sequences and, details are given in the sub-section II-B.

II. EXPERIMENTAL SETUP

A. *Video Database*

The video database used in this experiment was generated from seven high-definition videos (the same videos have been used in previous experiments). All them with 1280×720, 50 fps and 10 seconds duration. Representative frames of these originals are shown in Figure 1.

Fig. 1. First frame of the sequences included in Experiment.

B. Artifacts

New versions of each original video were generated using different combinations of artifacts. To generate test sequences, we generate blockiness, blurriness and packet loss artifacts and combine these artifacts at different proportions and strengths. In summary, the following parameters are used to generate the test sequences:

- 7 original videos Basketball, Romeo and Juliet, Park Run, Cactus, Park Joy, Barbecue and Into Trees;
- 4 different strengths of blockiness and blurriness: 0.4, 0.6 and 0.8;
- 3 different packet loss ratios: 0.7%, 2.6%, 4.3% and 8.1%.

Due to a large number of combined test videos (over 200 videos) makes it impossible to perform an experiment in a reasonable amount of time. Therefore, a reduction of the number of test sequences was achieved by selecting a subset of the original group. In this process, we eliminated most of the sequences already used in the two previous experiment already performed. We also reduced the number of parameters considering the results of the previous experiments. In summary, to select the test sequences we considered the following:

1) Total time necessary to execute the overall experiment;
2) The results of two previous experiments showed that the mean annoyance and the mean strength scores are significantly different for packet loss ratios of 0.7 and 8.1.
3) Among the blockiness and blurriness strength values considered in the previous experiments, we chose only the strength 0.4 and 0.6, which were considered to be more representative of these artifacts.

Thus, taking into account these choices, they resulted into 19 combinations (Table I) $\times$ 7 (video versions) + 7 original videos = 140 videos. This is a reasonable number of sequences which made it possible to run the experiment in the allocated time of 50 minutes.

C. Methodology

For all experimental session, subjects are requested to score a set of test videos with different combinations of artifacts. The number of observers was chosen in accordance with ITU Recommendations [9]. A Single Stimulus setup with implicit reference [9] was adopted for the task. The experiment is running in a room with constant illumination of approximately 70 lux. Each subject watches the stimuli (test sequences) on a 23 inches LED monitor of resolution 1360×768. The distance between the subject's eyes and the video monitor is 3 times the monitor screen's height.

TABLE I
COMBINATIONS OF THE PACKET LOSS RATIO (0.7 AND 8.1) WITH BLOCKINESS AND BLURRINESS (STRENGTHS 0.4 AND 0.6) ARTIFACTS.

Combination	Packet-Loss	Blocky	Blurry
1	0.0	0.6	0.0
2	0.0	0.0	0.6
3	8.1	0.0	0.0
4	0.7	0.0	0.4
5	8.1	0.0	0.4
6	0.7	0.0	0.6
7	8.1	0.0	0.6
8	0.7	0.4	0.0
9	8.1	0.4	0.0
10	0.7	0.4	0.4
11	8.1	0.4	0.4
12	0.7	0.4	0.6
13	8.1	0.4	0.6
14	0.7	0.6	0.0
15	8.1	0.6	0.0
16	0.7	0.6	0.4
17	8.1	0.6	0.4
18	0.7	0.6	0.6
19	8.1	0.6	0.6

Subjects are seated straight ahead of the monitor, centered at or slightly below the eye height for most subjects. A chinrest is used to guarantee a constant distance between the subject's eyes and the monitor. A SensoMotoric Instruments iView RED Eye Tracker was used throughout the experiment to record the eye-movements of the participants (The eye tracker had a sampling rate of $50/60$Hz, a pupil tracking resolution of 0.1° and a gaze position accuracy of $0.5 - 1^{\circ}$). The user interface for the experiment is implemented using the *Neurobehavioral Systems* software Presentation.

All subjects are asked to watch the test sequences and indicate whether they perceived any impairment in the videos; if so, they were asked to enter how annoying the artifacts were on a continuous annoyance scale ranging between 0 to 100. During the experiment, the experimenter guides the subject through all sessions (calibration, free viewing, training, practice and a main experiment). Table II show the summary of the settings of each experiment.

III. Data Analysis

We used the standard methods [9] for analyzing the annoyance judgments provided by the test subjects. We first computed Mean Opinion Score (MOS) for each test sequence. MOS is calculated by averaging the annoyance levels over all observers for each video. In this report MOS's values are treated as *MOS1*, *MOS2* and *MOS3* (i.e., MOS of the *Exp.1*, *Exp.2* and *Exp.3*, respectively). To test the relationship among the MOS's values we used the Spearman's correlation (ρ). To determine the significant effects between MOS's values

TABLE II
SUMMARY OVERALL EXPERIMENTS

Settings	Exp. 1	Exp. 2	Exp. 3
Numbers of Test Sequences	91	77	140
Numbers of Subjects	15	16	23
Calibration	–	*Yes*	*Yes*
Free Viewing	*Yes*	*Yes*	*Yes*
Training	*Yes*	*Yes*	*Yes*
Practice Trials	*Yes*	*Yes*	*Yes*
Main Experiment	*Yes*	*Yes*	*Yes*
Packet Loss	*Yes*	–	*Yes*
Blockiness	–	*Yes*	*Yes*
Blurriness	–	*Yes*	*Yes*

we used a non-parametric T-Test. We also are analysing only where settings (artifacts) are the same among the experiments:

- Case 1: from Exp. 1 until 3 considering only original videos;
- Case 2: between Exp. 1 and Exp. 3, considering only the 8.1% ratio of packet loss artefact;
- Case 3: between Exp. 2 and Exp. 3, considering only the 0.6 strengths of blockiness and blurriness artifacts (isolated).

A. *Case* 1

We report the MOS's computed across all 7 versions of video and all experiments, as shown in Figure 2. It is interesting to notice how the MOS's changes among the experiments: the MOS's of the experiments 1 until 3 are 1.43, 4.92 and 0.80, respectively. To know the strength and direction of association that exists among those MOS's, the Spearmans correlation has been calculated. Results has showed that exist a moderate correlation ($\rho = .595$) between $MOS1$ and $MOS3$. In other cases, there was one very weak correlation between $MOS1$ and $MOS2$ ($\rho = -.214$) and, between $MOS2$ and $MOS3$ ($\rho = .090$). Although, there are not statistically significant different ($p > .05$) among them.

B. *Case* 2

In this case, we report the $MOS1$ and $MOS3$ computed to packet loss artefact across all 7 versions of video, as shown in Figure 3. It is possible to notice that $MOS1 = 72.85$ is very higher than $MOS3 = 37.99$ with $\rho = .536$ and $p < .05$.

C. *Case* 3

Two different analyses were made between $MOS2$ and $MOS3$ across all 7 versions of video: first, we report the MOS's computed only to blockiness artefact and next, we report the MOS's computed only to blurriness artefact. Figure 4 shows a slightly difference between $MOS2$ and $MOS3$ to Park Joy, Into Tree and Park Run videos, but this difference increased when to Romeo & Juliet, Cactus, Basketball and Barbecue videos. Spearman's correlation coefficient has showed that there is a moderate and positive correlation ($\rho = .536$) between them. Although, there are statistically significant different only to Romeo & Juliet, Cactus and Basketball videos ($p < .05$).

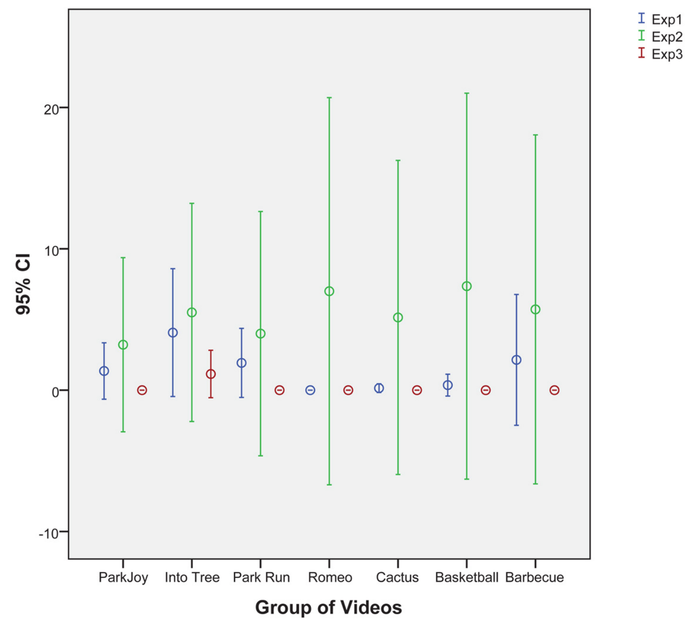

Fig. 2. Mean Annoyance Values averaged across all the original versions of each video.

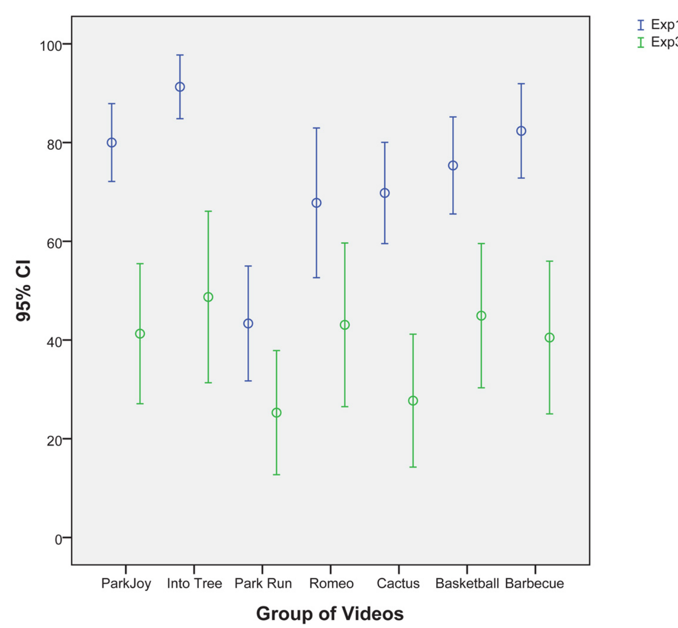

Fig. 3. Mean Annoyance Values averaged across of the packet loss versions of each video.

Next, it is considered blurriness artefact. As shown in Figure 5 there is a slightly difference between $MOS2$ and $MOS3$ to Park Joy, Into Tree, Cactus, Basketball and Barbecue videos increasing this different to Park Run and Romeo & Juliet videos. Despite of strong correlation ($\rho = .857$) between them, there was not statistically significant different ($p > .05$).

IV. CONCLUSION

The conclusion goes here.

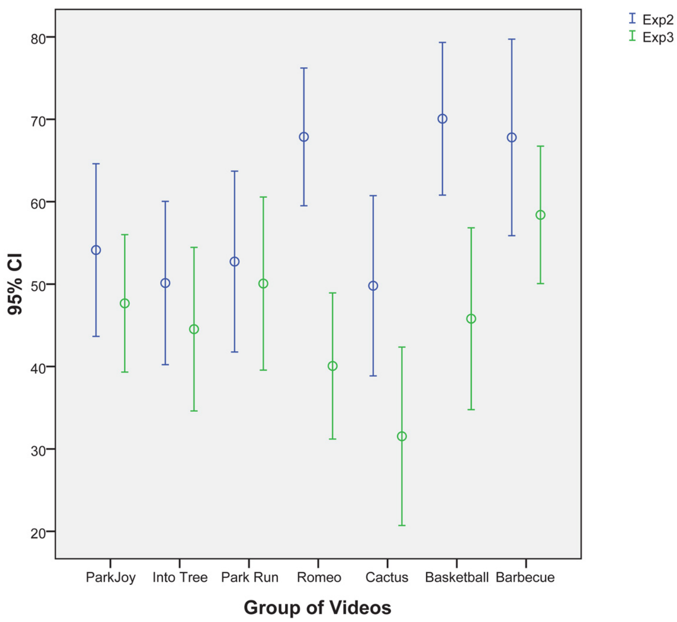

Fig. 4. Mean Annoyance Values averaged across of the blockiness versions of each video.

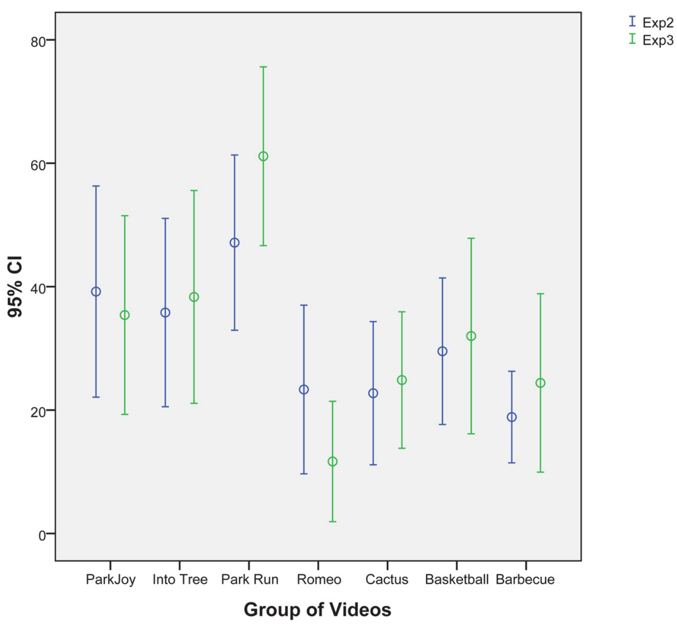

Fig. 5. Mean Annoyance Values averaged across of the blurriness versions of each video.

REFERENCES

[1] Farias, M. C. Q.; Heynderickx, I.; Bruno, L. M. E.; Redi, J. A., *Visual Artifacts Interference Understanding and Modeling (VARIUM)*. In: Seventh International Workshop on Video Processing and Quality Metrics for Consumer Electronics, 2013, Scottsdale.

[2] Redi, J. A.; Heynderickx, I.; Bruno, L. M. E.; Farias, M. C. Q., *On the impact of packet-loss impairments on visual attention mechanisms*. In: ISCAS, pp. 1107-1110. IEEE, 2013.

[3] Farias, M. C. Q.; Foley, J. M.; Mitra, S. K., *Perceptual contributions of blocky, blurry and noisy artifacts to overall annoyance*. In: Proc. IEEE International Conference on Multimedia & Expo, Baltimore, MD, USA, 2003, Vol. 1, pp. 529532.

[4] Farias, M. C. Q.; Foley, J. M.; Mitra, S. K., *Some Properties of Synthetic Blocky and Blurry Artifacts*. In: Proceedings of the SPIE, Human Vision and Electronic Imaging VIII, Vol. 5007, pp. 128-136 (2003).

[5] Farias, M. C. Q.; Foley, J. M.; Mitra, S. K., *Perceptual analysis of video impairments that combine blocky, blurry, noisy, and ringing synthetic artifacts*. In: Proceedings of the SPIE, Human Vision and Electronic Imaging VIII, Vol. 5666, pp. 107-118 (2005).

[6] Farias, M. C. Q.; Foley, J. M.; Mitra, S. K., *Detectability and Annoyance of Synthetic Blurring and Ringing in Video Sequences*. In: International Conference on Acoustics, Speech, and Signal Processing - ICASSP, Vol. 3, pp. 697-700, 2004.

[7] Farias, M. C. Q.; Foley, J. M.; Mitra, S. K., *Detectability and Annoyance of Synthetic Blocky and Blurry Video Artifacts*. In: Journal Sid Symposium Digest of Technical Papers, Vol. 33, No. 1, 2002.

[8] Farias, M. C. Q.; Foley, J. M.; Mitra, S. K., *Detectability and annoyance of synthetic blockiness, blurriness, noisiness, and ringing in video sequences*. In: Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, Philadelphia, PA, Vol. 2, pp. 553556, 2005.

[9] International Telecommunication Union, *ITU-T Recommendation BT.500-8: Methodology for the subjective assessment of the quality of television pictures*. 1998.

[10] Farias, M. C. Q.; Mitra, S. K., *No-reference video quality metric based on artifact measurements*. In: ICIP, IEEE, pp. 141-145, 2005.

[11] Landau, S.; Everitt, B. S., *A Handbook of Statistical Analyses Using SPSS*. ISBN: 9781584883692, Statistics Chapman & Hall/CRC, Taylor & Francis, 2004.

[12] Montgomery, D., C., *Introduction To Statistical Quality Control*. ISBN: 9780470233979, John Wiley & Sons, Incorporated, 2009.

[13] Seshadrinathan, K.; Soundararajan, R.; Bovik, A. C.; Cormack, L. K., *Study of Subjective and Objective Quality Assessment of Video*. In: IEEE Transactions on Image Processing, Vol. 19, No. 6, Piscataway, NJ, USA, 2010.