https://doi.org/10.2352/ISSN.2470-1173.2018.14.HVEI-529

Measuring Video Quality by Eye Response

Deepti Pappusetty, Hari Kalva; Florida Atlantic University; Boca Raton, Florida/ USA

Abstract

This paper presents an emerging research area of using eye response for video quality evaluation. Experiments were conducted to study how eye movements such as saccades and fixations in a video, which represents visual information processing, react to changes in video quality. Discussed in this paper are two sets of experiments on eye movements. The first experiment had multiple original movie video clips with quality variation applied across the entire scene. The second round of tests had multiple original movie video clips with transitioning between extreme qualities half way through within the video. The key contributions of this paper are an exploration of gaze response to video quality variation and results of the experiment show a significant difference in Fixations/ Saccades ratio between low and high-quality videos. Gaze response to video quality shows a potential to be considered as one of the quality evaluation metrics for future studies.

Introduction

Psychological studies on perception provide us an understanding of how processing of information occurs in Human Visual System (HVS). Perceptual and cognitive factors affect how users acquire and process visual information. Psychological studies related to HVS, Vision Science, Behavioral Sciences, and Neuroscience have extensive applications in the field of image processing, computer vision and video processing [1]. These studies can be utilized to develop efficient, perceptually optimized video processing frameworks, algorithms, and applications [2].

In past few years, impressive progress has happened in research for the transmission and storage of videos using compression algorithms [3]. Properties of human visual system play a crucial role in perception and experience of viewing videos. HVS characteristics are imperative in designing efficient coding algorithms. Effective video compression techniques exploit four types of redundancies: 1. Perceptual, 2. Temporal, 3. Spatial and 4. Statistical [4]. Video quality assessment metrics are used to predict viewer's quality of experience [5]. It includes subjective and objective quality metrics. Subjective quality assessments include Mean Opinion Score (MOS) giving the average rating over all the viewers for a given clip. Recommended testing procedures include Double Stimulus Continuous Quality Scale (DSCQS), Double Stimulus Impairment Scale (DSIS), Single Stimulus Continuous Quality Evaluation (SSCQE) or Absolute Category Rating (ACR) [6]. Objective quality assessments include metrics like Mean Squared error (MSE), peak signal-to-noise ratio (PSNR), Structural similarity (SSIM), and Visual information Fidelity (VIF) etc.

This study mainly explores gaze (fixations and saccades) analysis to video quality evaluation. Eye response can be measured non-intrusively using an eye tracker and offers a potential new approach to understanding video structure and content. Gaze response to video stimuli can be used to develop new approaches to video analysis, scene exploration & quality evaluation. The key contributions of this study are an exploration of gaze response to video quality variation and results of the experiments show that fixations/ saccades ratios are significantly different with various perceived quality. Gaze response to video quality shows a potential to be considered as one of the quality evaluation metrics for future studies.

Related Background

The eye has often been compared to like a camera which is self-focusing, adjusting automatically for light intensity, has a self-cleaning lens and feeds into a computer with parallel-processing capabilities so advanced that engineers are only just starting to consider similar strategies for the hardware they design [7].

Eye movement such as saccades (rapid movement of eyes) and fixations (the focus of the eyes on a point) represents the visual processing of information presented in scenes. Fixations lead to data extraction and processing [8]. Saccades depend on the saliency of objects in a scene and the tasks and intentions of the observer [9]. Viewers usually demonstrate short fixation durations and large saccade amplitudes in the earlier phases of viewing an image followed by longer fixations and shorter saccades in the latter phases of scene viewing [10].

Eye tracking has gained recognition in image and video quality analysis. The research included gaze information into the image quality assessment system by introducing artifacts in the regions of higher saliency for a potential improvement of assessment accuracy in image quality metric [11]. Eye tracking data was used to provide insights into the optimal use of visual attention in image quality research by obtaining data with a large degree of stimulus variability [12]. You *et al.* proposed an advanced foveal imaging model to generate the perceived representation of video by integrating visual attention into the foveation mechanism [13]. These studies show integrating eye tracking data into image and video quality metrics. QOE using eye trackers often include analysis of eye tracking data like fixation location, duration, and order [14]. [15] [16] provide survey studies on physiological approaches of quality assessments.

Most of the QOE research focus on salient regions and added distortion to those regions. We conducted experiments by adding distortion to the entire scene and getting free – viewing responses. We found fixations to saccades ratio significantly different and larger for distorted video when compared to original uncompressed videos. We presented a preliminary work of this research at European Conference on Eye Movements [17].

Experiment

Subjects

Twenty subjects participated in this experiment. All the participants were naïve graduate and undergraduate students of Florida Atlantic University. We divided 20 subjects who participated in this experiment into four groups of 5 participants each. All the subjects had normal or corrected-to-normal vision. The subjects' age ranged between 20-30 years with normal or corrected-to-normal vision.

Test Stimuli & Procedure

We selected three movies "A walk in the woods," "Inception" and "The Lord of the Rings" to pick short 10 seconds video clips for

(a) Sample extreme quality variations for video A

(b) Video B

(c) Video C

(d) Video D

(e) Video E

(f) Video F

(g) Video G

(h) Video H

(i) Video I

(j) Video J

Figure 1. Ten natural movie videos, represented as A, B, C, D, E, F, G, H, I, and J from three movies “A walk in the woods”, “Inception” and “The Lord of the Rings”

the experiment (Figure 1). Each of the selected video clips was represented as A, B, C, D, E, F, G, H, I, and J respectively. Each of these clips had one or two actors in each shot and did not contain any vigorous movements. None of the clips had any scene cut. Resolution for videos was 1920x1080 and 24fps frame rate.

We used an x264 encoder/ decoder to process each video clip into four different quality levels. First, a Full 10 seconds Good (G) quality encoded at 30000 kbps. For example, “A” represents a full good quality movie. Second, a full 10 seconds Bad (B) quality encoded at 1000 kbps bitrate. “A̲” represents a bad quality movie. Third, a 5 seconds good quality to 5 seconds bad quality video – half-good-to-half-bad (HGHB). “A | A’” represents a half-Good-to-half-bad quality video. Lastly, a 5 seconds bad quality to 5 seconds good quality video – half-bad-to-half-good (HBHG). “A’|A” represents a half-bad-to-half-good quality video. Table I shows the sequence of movies played for each pool of participants.

We conducted a test with each participant individually. No video was shown more than once to a participant to avoid any expectation bias. Each participant first went through a 5-point calibration process before each trial followed by the ten 10-seconds video clips shown in sequence as a playlist.

Apparatus

We conducted the test in a dark room with no other source of light except for the display monitor. We showed the test stimuli on a 27-inch flat screen LCD monitor with 1920x1080 resolution. A Tobii X2-60 eye tracker was used to record the gaze responses of the subjects. The Tobii eye tracker consists of an infrared camera and is desktop mounted on the monitor. We did not use any chin rest and was a free viewing experimental condition. The participants were instructed to keep their head movements, and eye blinks to minimal. We set the viewing distance for the participants to approximately 65 cm as per the standard eye-tracking instructions. The sampling rate for eye-tracker is 60 Hz. We used OGAMA Gaze and Mouse Analyzer software version 4.5 to design our experiment and to obtain the pupil diameter.

OGAMA does not have a feature to play videos. For this, we developed a new software in Windows Presentation Foundation (WPF) C#.net to play videos in the embedded windows media player feature. We used OGAMA desktop module to record the screen while users watched the test stimuli playing on the screen. The desktop module of OGAMA collects gaze responses of individuals watching the content on the screen. To synchronize with OGAMA to get the start time and end time of each video, we coded the program in such a way that, when the video started playing, a mouse click event is triggered which is caught by OGAMA. Also, a mouse click event is registered when the video stops playing giving the end time for OGAMA pupil readings. Once all the videos in the playlist ends, “F1” event is triggered by the software. It marks as an indication to OGAMA (set in OGAMA settings) to stop the desktop reading and signifies the end of the experiment for a subject. Matlab scripts were written to get individual video stimuli responses based on the start and end timestamps recorded by OGAMA.

We exported raw responses for each participant per video. Raw data was split per video per subject and imported back to OGAMA as individual responses. OGAMA provides a statistical analysis feature to give metrics like fixations duration, saccades duration, duration mean, and so forth [17].

Table I: Test Stimuli Sequence for Each Group for Experiment I

Groups	Video Sequence
1	A, B, C, D, E, F, G, H, I, J
2	F, G, H, I, J, A, B, C, D, E
3	A|A’, B|B’, C’|C, D’|D, E|E’, F|F’, G’|G, H’|H, I’|I, J’|J
4	A’|A, B’|B, C|C’, D|D’, E’|E, F’|F, G|G’, H|H’, I|I’, J|J’

Results and Discussions

Full Good Versus Full Bad Video-wise Results

We obtained mean of all the participants video-wise for fixations/saccades ratios metric of OGAMA standard statistical analysis metrics. Full good quality videos versus full bad quality videos results are as shown in Figure 2. OGAMA defines the metric fixations/saccades ratio as a sum of fixation times divided by trial duration [18]. Our null hypothesis is that there is no average difference between eye responses of participants to high versus low quality videos. Results show statistically significant difference in fixation/saccades ratios between low and high-quality videos rejecting null hypothesis. Fixation/saccades ratios are higher values for low-quality videos (Z-value is -2.2934. The p-value is 0.02202. The result is significant at $p \leq 0.05$.). The results infer that viewers are fixating more in bad quality videos when compared to high-quality videos. Bad video quality videos mark more effort by users in discerning features. In a study with gaze comparison for high-quality images versus distorted images, the latter attracted fewer numbers of fixations but longer fixation durations, shorter saccade distance and stronger central fixations bias [19]. We see similar results in case of videos too.

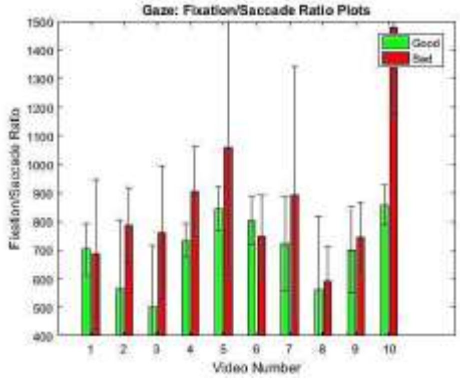

Figure 2. Full Good videos versus Full Bad videos fixations/saccades ratios video-wise mean for all the participants

HGHB and HBHG Video-wise Results

Figure 3 and Figure 4 show results from HGHB and HBHG scenarios. HGHB and HBHG video sequences showcase a more random order of video combinations, i.e., good to bad and bad to good when compared to Full good and Full bad videos. We divide it into two scenarios when analyzing results video-wise (i) First Half Good Quality to First Half Bad Quality transition (ii) Second Half Good Quality to Second Half Bad Quality transition.

Figure 3 shows video-wise mean for first-half extreme quality videos. We see close to the significant difference in fixations/saccades ratio for bad versus good (Denoted by Bad1 and Good1 for the first half). The Z-value is -1.8857. The p-value is 0.05876. The result is not significant at $p \leq 0.05$. Figure 4 shows the results for the second half. Not very remarkable results are seen with random order of videos (The Z-value is -0.4587. The p-value is 0.64552. The result is not significant at $p \leq 0.05$). A plausible explanation for this could be the effect of a different quality presented in first half and viewer being familiar with the events happening in the scene.

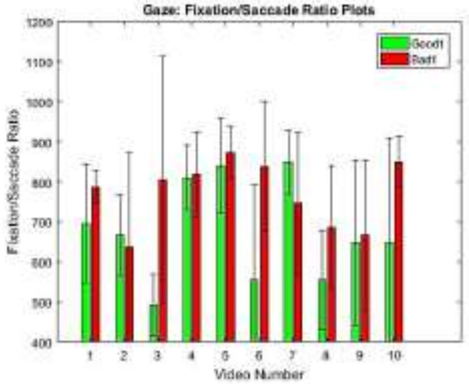

Figure 3. Fixations/saccades ratios Video-wise mean for first half good versus first half bad quality videos

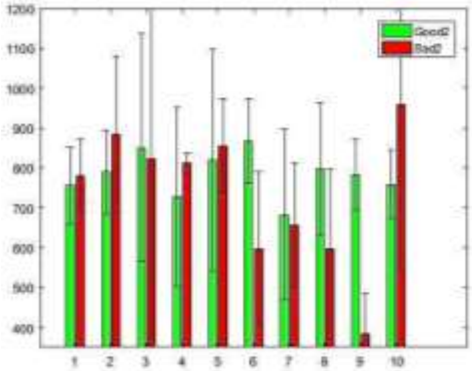

Figure 4. Fixations/saccades ratios Video-wise mean for second half of good versus second half bad quality videos

Conclusions

The experiment shows a significant difference in fixations/ saccades ratios for low and high-quality videos. Higher fixations/saccades ratios are found for low-quality videos compared to high quality in case of full good quality videos versus bad quality videos. Longer fixation times show more effort to extract information for scene perception. These results show a substantial effect of low-quality videos in scene-viewing gaze behavior. The results are not very significant in the events of half-way transitions. This experiment forms the basis for a potential to explore more and establish a relationship between the magnitude of the response and variations in video quality. Applications of this research include automatic assessment of subjective video quality and video content & structure analysis.

References

[1] A. T. Duchowski, “A breadth-first survey of eye-tracking applications,” Behavior Research Methods, Instruments, & Computers, vol. 34, no. 4, pp. 455–470, Nov. 2002.

[2] H. Kalva, A. Bovik, H. Chen, K. Egiazarian, and Z. Wang, “Introduction to the Issue on Perception Inspired Video Processing,” IEEE Journal of Selected Topics in Signal Processing, vol. 8, no. 3, pp. 355–357, Jun. 2014.

[3] A. B. Watson, Digital images, and human vision. Cambridge, Mass. : MIT Press, 1993.

[4] H. Kalva, “The H.264 Video Coding Standard,” IEEE MultiMedia, vol. 13, no. 4, pp. 86-90.

[5] S. Winkler and P. Mohandas, “The Evolution of Video Quality Measurement: From PSNR to Hybrid Metrics,” IEEE Transactions on Broadcasting, vol. 54, no. 3, pp. 660–668, Sep. 2008.

[6] ITU, “ITU-T P.910 Subjective video quality assessment methods for multimedia applications,” Tech. Rep., International Telecommunication Union, 2006.

[7] H. H. David, “The Eye,” Eye, Brain, and Vision, 6th ed., vol. 22, Scientific American Library, New York: W. H. Freeman & Co., ch 3, pp. 42-57, 1988.

[8] R. Groner and M. T. Groner, “Attention and eye movement control: an overview,” Eur Arch Psychiatry Neurol Sci, vol. 239, no. 1, pp. 9–16, 1989.

[9] J. M. FINDLAY, “Saccade Target Selection During Visual Search,” Vision Research, vol. 37, no. 5, pp. 617–631, Mar. 1997.

[10] A. Helo, S. Pannasch, L. Sirri, and P. Rämä, “The maturation of eye movement behavior: Scene viewing characteristics in children and adults,” Vision Research, vol. 103, pp. 83–91, Oct. 2014.

[11] K. Fliegel, “Eyetracking based approach to objective image quality assessment,” in 2008 42nd Annual IEEE International Carnahan Conference on Security Technology, pp. 371–376, 2008.

[12] W. Zhang and H. Liu, “Toward a Reliable Collection of Eye-Tracking Data for Image Quality Research: Challenges, Solutions, and Applications,” IEEE Transactions on Image Processing, vol. 26, no. 5, pp. 2424–2437, May 2017.

[13] J. You, T. Ebrahimi, and A. Perkis, “Attention Driven Foveated Video Quality Assessment,” IEEE Transactions on Image Processing, vol. 23, no. 1, pp. 200–213, Jan. 2014.

[14] U. Engelke et al., “Psychophysiology-Based QoE Assessment: A Survey,” IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 1, pp. 6–21, Feb. 2017.

[15] U. Engelke, A. Maeder, and H. J. Zepernick, “Visual attention modeling for subjective image quality databases,” in 2009 IEEE International Workshop on Multimedia Signal Processing, pp. 1–6, 2009.

[16] P. Le Callet and E. Niebur, “Visual Attention and Applications in Multimedia Technologies,” Proc IEEE Inst Electr Electron Eng, vol. 101, no. 9, pp. 2058–2067, Sep. 2013.

[17] D. Pappusetty, H. Kalva, “Eye Response to Blockiness Artifacts in Video,” in European Conference on Eye Movements, Wuppertal, Germany, Aug. 2017 (in press).

[18] A. Voßkühler, V. Nordmeier, L. Kuchinke, and A. M. Jacobs, “OGAMA (Open Gaze and Mouse Analyzer): Open-source software designed to analyze eye and mouse movements in slideshow study designs,” Behavior Research Methods, vol. 40, no. 4, pp. 1150–1162, Nov. 2008.

[19] F. Röhrbein, P. Goddard, M. Schneider, G. James, and K. Guo, “How does image noise affect actual and predicted human gaze allocation in assessing image quality?,” Vision Research, vol. 112, pp. 11–25, Jul. 2015.

Author Biography

Deepti Pappusetty received her BS in Computer Science and Engineering from Jawaharlal Nehru Technological University (2008), India, and her Masters in Bioengineering (2011) from Florida Atlantic University. She has around two years of experience as a Software Developer in Oracle Pvt. Ltd. She is currently pursuing her Ph.D. in Computer Science from FAU. Her work is focused on exploring ways to apply models of human perception and cognition to video processing. Applications of this research include new approaches to video analysis, quality evaluation and visual search.

Hari Kalva is a Professor, Associate Chair, and the Director of the Multimedia Lab in the Department of Computer & Electrical Engineering and Computer Science at Florida Atlantic University (FAU). Dr. Kalva has over 20 years of experience in multimedia research, development, and standardization. He has made key contributions to technologies that are now part of MPEG-4 standards. His current research focuses on understanding and applying human visual perception, cognition, and social context to optimize visual information processing. Dr. Kalva received a Ph.D. and an M.Phil. In Electrical Engineering from Columbia University in 2000 and 1999 respectively. He received an M.S. in Computer Engineering from Florida Atlantic University in 1994, and a B. Tech. in Electronics and Communications Engineering from N.B.K.R. Institute of Science and Technology, S.V. University, Tirupati, India in 1991.