EVALUATING THE TASK DEPENDENCE OF EYE MOVEMENTS FOR COMPRESSED VIDEOS

Anush K. Moorthy, Wilson S. Geisler and Alan C. Bovik

The University of Texas at Austin,
Austin, Texas - 78712, USA.

ABSTRACT

The dependence of eye movements on the assigned task when a subject views an image/video is a topic of considerable interest. In the recent past, researchers have focused on how distortions in an image affect eye movements and fixations. Surprisingly little work has been done on how distortions in video affect the direction of gaze. Here we seek to evaluate how distortions in compressed video affects eye movements. We study the effect that two particular tasks (quality assessment and summarization) have on eye movements when subjects view videos. Further, we also evaluate the effect of the amount of distortion on eye movements conditioned on a particular task. The results from this study are relevant for researchers developing video quality assessment algorithms based on human eye moments. The results may also be useful for improving pooling strategies for quality indices.

Index Terms— Eye movements, Quality Assessment, Videos, Task-dependence, Compression

1. INTRODUCTION

When shown an image or a video, a human rapidly scans the scene using ballistic eye movements called saccades linked by fixations [1]. Most information is gathered during a fixation and little to no information is gathered during a saccade. The study of eye movements has been an active area of research for a long time. Understanding how humans fixate is an important step toward understanding human vision.

Image/video quality assessment has been another area of interest in the recent past [2]. Development of quality assessment metrics is closely related to eye movements, in that, information gathered from fixations is used to perform the complex task of assessing the quality of images. Recently, pooling strategies for collapsing local quality scores based on fixations and eye movements have been proposed to improve the performance of algorithms which seek to emulate human perception of quality [3].

In this paper we focus on the area of research which falls between these highly complex tasks. In particular, we evaluate the effect of a given task on eye movements. It is known that the task heavily influences eye movements [1]. Specifically we evaluate the effect of a quality-assessment task on eye movements. Further, we also evaluate the effect of the presence of artifacts on eye movements. Although we are unaware of any such work for distorted videos, there exists literature on similar studies for images.

2. PREVIOUS WORK

Miyata *et al.* conducted a study where they tracked the eyes of subjects who were asked to rate the quality of images degraded by blurring, addition of noise or color-shifting [4]. They concluded that for the types of distortions considered, visual attention patterns do not change in the presence of distortion.

Vuori and Olkkonen [5] studied the effect of image sharpness on visual attention when viewing natural images to evaluate quality. They evaluated the possibility of applying eye movement information to predict subjective quality preferences. A total of 24 images were shown to eight observers and each participant was asked to evaluate the general quality of the image or of its colors on a 7-point scale. Further, the study involved the use of another specific task for each image - this task was memory oriented; for example, how many buildings are there in the image? Using the saccade duration as the test parameter, the authors concluded that image quality had a significant effect on eye movements - for example, saccade duration increased with increasing blur. They also reached the conclusion that on the image-specific task there was a significant difference in the saccade duration.

In [6] the authors evaluate the following: 1. influence of task on occulomotor behavior and 2. visual attention patterns when viewing a pristine image vs. when viewing a degraded image. In order to investigate the task impact, the authors classify the tasks as ‘free-viewing’ and ‘quality assessment’. In the ‘free-viewing’ task, the subjects were instructed to “look around the image”. The authors claim that free-viewing reduces the top-down effects in attention. They conclude that fixation duration increases when a pristine image is used for a quality assessment task. Further, the type of degradation modifies visual attention - specifically, the authors notice that JPEG and JPEG2000 distortions affect fixations. Finally, they note that multiple viewing of the same image does not alter the viewing strategy. Details of this study were also presented in [7].

In [8], the authors build upon the research in [6], where they vary the *degree* of distortion and investigate the influence of distortion type and degree when judging image quality. Their conclusions mirror those in [6] - JPEG and JPEG2000 distortions do change visual attention strategies. Further, they also note that the degree of distortion appears to change fixations as well. However, these conclusions do not hold when the distortion pattern is a global one (eg., blur or noise).

Related work for videos was performed by Abdollahian *et. al.* in [9], where the authors studied eye movement patters for different camera motions including pan, tilt and zoom.

Our aim in this paper is to perform an analysis similar to that in [8], but for videos. Specifically, we consider two tasks - summarization and quality assessment - and evaluate the effect that the task has on eye movements. Further, given a particular task, we evaluate the effect of distortion on eye movements. These analyses are carried out for H.264/AVC compressed videos which exhibit blocking and blurring artifacts.

3. EVALUATING THE TASK DEPENDENCE OF EYE MOVEMENTS

In this paper we propose to address the following questions:

1. Does the task impact eye movements in video?
2. Given a distorted video, for a particular task, will the eye movements be any different than for pristine (undistorted) reference video?
3. Does the level of distortion have any effect on eye movements?

3.1. The Tasks

Most of the work cited above operated under a "free viewing" condition where the subject was asked to simply view the images. Those authors contend that this task reduces top-down effects and hence enhances bottom-up parameters responsible for fixations and gaze. We believe that completely detaching top-down influences from a visual task is almost impossible. When asked to view an image or a video freely, we assert that the human develops his own top-down instruction and hence the free-viewing task is not completely free of top-down influences. As it has been noted before [5], the specific task has a definite impact on the number of fixations and viewing time [1]. Given that this is true, under our hypothesis that a 'free viewing' condition does not isolate top-down influences, eye movements then correspond to an implicit task that the human undertakes. Comparing these fixation patterns against those from a quality assessment task hence does not seem ideal. Therefore, we strictly define the alternate task so as to ensure uniformity across human subjects.

In order to answer the task-dependence question, we record fixations as the subjects were shown videos with two alternate tasks : (1) a Summarization task and (2) a Quality assessment task.

The instructions to the subject for the summarization task were as follows: *You will be shown a set of videos. At the end of a set of presentations, we shall show you a frame from a video and will be asked to explain/summarize the scene that corresponds to the video. You will have to explain the series of events/say what happened in that video.* Our hypothesis is that such a strict definition emphasizing the same top-down influences would serve to remove possible randomness associated with inter-subject eye movements.

The quality assessment task is as follows: *You will be shown a set of videos. At the end of each video, you will be asked to rate the quality of the video on a five-point scale (Bad -1/5, Poor - 2/5, Fair - 3/5, Good - 4/5, Excellent - 5/5).* Our hypothesis here is that that a quality assessment task will influence eye movements, since the human will be more critically viewing the video than in the summarization task.

3.2. The Videos

We selected 20 clips from various foreign (French, German) films, each 30 seconds long. The clips were chosen such that each clip had sufficient content to summarize the video. Further, the reason for choosing foreign films was to minimize the effect of memory or recall in the task when seeing a clip that the subject has already seen before. A frame from a sample of the videos used in the study is shown in Figures 1.

In order to introduce distortion in videos we compressed each video with the H.264/AVC encoder [10]. H.264/AVC is the latest standard for video compression. Videos were compressed using the JM reference software encoder [11] and the baseline profile [10] was used. Each video was compressed at two different bit-rates, 1.25 Mbps and 2.25 Mbps. The encoder parameters were: 50 macroblocks/slice, 27 slices/frame, 3 slice groups, dispersed Flexible Macroblock Ordering (FMO) mode. The two such compressed videos were labeled 'High' and 'Low' compression/distortion respectively. Thus we created a total of 60 compressed videos, each of resolution 720×480 with a frame-rate of 30 frames per second (fps).

3.3. Experiment design and display

A total of 12 subjects participated in the study. Six subjects were assigned to each of the two tasks - quality assessment and summarization. The subjects were naive concerning the field of quality assessment. Eyesight was not tested, but a verbal confirmation of visual acuity was accepted. Each subject

Fig. 1. (a)-(f) Frames of video sequences used for the study. All of the chosen films were foreign films (non-english) in order to minimize memory effects.

saw 20 videos. The videos seen by the subject were such that no subject saw the same content twice. The video playlists were arranged such that amongst three subjects all 60 videos were seen; where each one of the three subjects saw all the possible 20 different contents but with varying levels of quality (pristine, low distortion, high distortion) which was randomly assigned. Since we assigned had six subjects per task, we obtained two sets of fixations for each of the 60 videos for each task.

We used the eyetracker manufactured by Cambridge Research Systems with a 50 Hz rate. The viewing distance was fixed at 1000 mm. The eyetracker was controlled using the provided MATLAB Video Eye Tracker (VET) toolbox. The stimuli were displayed on a 21" LCD monitor with a resolution of 1600×1200 and a refresh rate of 60 Hz. The videos were displayed at the center of the screen at their native resolution using the XGL toolbox developed at the University of Texas at Austin [12]. The XGL toolbox allows for precise display of videos without any jitter or delay so that no additional artifacts are introduced in the video. Each frame in the video was displayed twice (this is because the refresh rate is set at 60 Hz) so that the frame rate is 30 Hz. Calibration was performed using the provided routine. Calibration was performed at the beginning of the study as well as after every two presentations. The subject was explained the task beforehand, where the same task as mentioned before was read out each time. Care was taken to ensure that the subject understood that 'quality' did not mean quality of acting/directing or of content, but quality of the video. For the summarization task, after 6 or 8 videos the subject was shown randomly chosen frames from a subset of the videos that that subject had seen and was asked to describe the video. All subjects seemed to have no problems recalling the videos and provided a concise explanation. For the quality assessment task, after each presentation the subject was asked to rate the video verbally - the scores of which were noted down.

The eye movements obtained from one of the two subjects for video 12 (summarization task, pristine case) were discovered to be completely un-reliable (as indicated by the eye-tracker software). Hence, instead of including partial results from that video, we choose to completely neglect it (and its low and high distorted versions) in our analysis. Thus, we acquired eye movement data for 57 videos for each task.

4. RESULTS

Previous work similar in concept to ours for images utilized saccade duration [5] and fixation locations [8]. In this study we utilize actual eye movements. The eye movements for each video and for each task were averaged across the two subjects to produce a mean set of eye movements for each of the videos. The eye movement locations are 2-d vectors – coordinates on the 2-d screen – and are functions of time. In order to collapse the spatial vector, we compute the distance of each of these eye movement locations from the center of the screen. This produces a scalar value for each sampled instant. In order to pool this time-series, and to compare eye movements across tasks we use the coefficient of variation (CV) [13], which is the ratio of the standard deviation of a vector to its mean. The CV captures the variations in eye movements better than the simple mean. Having said this, we also note that the CV may not be an ideal metric for time-series pooling. Future work will involve understanding how best to pool these time-series scores.

In our study the CV of eye movements across each video was computed to produce a single value for each video. This was computed across all videos to form a 57-dimensional vector for each task. A Kolomogorov-Smirnov (KS) test for Gaussianity [13] of the vectors corresponding to the videos for each of two tasks confirmed that the null hypothesis (the vectors are drawn from a Gaussian distribution) could not be rejected at 95% confidence level. Hence we applied the t-test [13] in order to determine if the two sample vectors are drawn from the same distribution. We found that the null hypothesis (the two vectors are drawn from the same distribution) was rejected at the 95% confidence level. In accordance with previous literature, we find that *the task clearly influences eye movements in video*. This answers the first question that we posed: *Does the task impact fixations in video?*.

In order to answer the next two questions: *Given a distorted video , for a particular task, will the eye movements be any different than for pristine reference video?* and *Does the level of distortion have any effect on eye movements?* the following analysis is performed. For each set of three videos sharing the same content (pristine, low distortion, high distortion) and for each task, the mean fixation location as a function of time between the two subjects is computed as described above[1]. Hence for each video, a vector of length equal to the number of (sampled) eye-movements is produced. This vector is utilized to study the relationship between two videos which share the same content. The KS test for Gaussianity indicates that for a majority of these vectors, the null hypothesis (the vectors are drawn from a Gaussian distribution) is rejected at the 95% confidence level.

At this point, there are two options - (1) to apply a non-parametric test (such as the Mann-Whitney U test for two independent samples) or (2) to apply the t-test but with a reduced confidence level. Sheskin provides arguments for both of these cases, and it seems that most researchers prefer to use a conservative t-test owing to its robustness as against a non-parametric one [13]. We utilized the t-test to infer if the two sets of eye movements were drawn from the same distribution, at the 90% confidence level. This is evalauted for each task. We realize that the number of subjects has an influence on the results of these statistical tests, and future work will involve an increase in the number of subjects.

The reader will notice that we perform the t-test on each content *individually*, instead of across contents. The reason for this is as follows. Since content of the video may influence the perceivability of distortions, our hypothesis is that the eye movements in videos where distortions are not perceivable will not differ significantly from those for the pristine video. In cases where the distortion is perceivable, the distributions of eye movements must differ significantly. Future work will involve quantifying the distortions at eye movement locations using objective quality indices [14].

The results of the t-test for each task are seen in Table 1. The null hypothesis is that the two sets of eye movements (pristine and low distortion or pristine and high distortion) are drawn from the same distribution. A value of '0' in these tables indicates that the null hypothesis cannot be rejected at the 90% confidence level. A value of '1' indicates that the null hypothesis can be rejected at the 90% confidence level and the two sets of eye movements are significantly different.

The results in Table 1 warrant further discussion. We find that for the summarization task, there seems to exist definite statistical differences between eye movements in the presence of distortion. This is extremely interesting. Under the hypothesis that the human is guided by a particular task when he views a video (in this case the summarization task), his eye movements change in the presence of distortion. In the case of the quality assessment task, however, there seems to be greater agreement between eye movements when viewing a video. This implies that the strategy that the subject adopts for the quality assessment task remains essentially the same in the presence of distortion. Of course, this depends heavily on the content of the video and the quality of direction.

The results imply that for consumer applications, quality assessment algorithms that employ pooling strategies based on eye-movements will better predict the quality of experience that the end-user has. We propose to quantify the statistical differences as a function of the quality at points of gaze in future work.

5. CONCLUSIONS

We evaluated the task dependence of eye movements for two well-defined tasks - quality assessment and summarization. Further, we evaluated the effect that distortion (compression) had on eye movements for each of these tasks. The influence of the degree of distortion was also examined. Each of these analyses was carried out for a diverse set of videos. We concluded that across videos, the task influences eye movements. Specifically, the eye movements for a quality assessment task differs significantly from that for a summarization task. We also demonstrated how the degree of compression affects eye movements for each of these tasks for each video in the study. Future work will involve increasing the number of subjects, and evaluating the statistics at points of gaze. The study presented here has applications in developing pooling strategies for video quality assessment algorithms.

6. REFERENCES

[1] A. L. Yarbus, *Eye Movements and Vision*, Plenum press, 1967.

[2] Z. Wang and A.C. Bovik, *Modern Image Quality Assessment*, vol. 2, Morgan & Claypool Publishers, 2006.

[1] This leads to a time-series. Note that here the CV is not used to produce scalar value for the video

	Quality Assessment Task		Summarization Task	
Video	Low Distortion	High Distortion	Low Distortion	High Distortion
1	1	0	1	1
2	0	1	1	0
3	0	0	1	1
4	0	1	1	1
5	1	1	1	0
6	1	0	1	1
7	1	1	1	1
8	0	0	1	1
9	1	1	1	0
10	1	0	1	1
11	1	1	0	1
12	N/A	N/A	N/A	N/A
13	0	1	1	1
14	1	0	1	1
15	0	0	1	1
16	1	1	1	1
17	0	1	1	1
18	0	1	1	1
19	1	1	1	1
20	1	1	1	1

Table 1. Table evaluating the differences between distributions of the pristine and low distortion or pristine and high distortion version of each video. The null hypothesis is that the two sets of eye movements (pristine and low distortion or pristine and high distortion) are drawn from the same distribution. A value of ‘0’ in these tables indicates that the null hypothesis cannot be rejected at the 90% confidence level. A value of ‘1’ indicates that the null hypothesis can be rejected at the 90% confidence level and the two sets of eye movements are significantly different.

[3] A. K. Moorthy and A. C. Bovik, “Visual importance pooling for image quality assessment,” *IEEE Journal of Special Topics in Signal Processing: Visual media quality.*, April 2009.

[4] K. Miyata, M. Saito, N. Tsumura, H. Haneishi, and Y. Miyake, “Eye movement analysis and its application to evaluation of image quality,”,” in *Final Program and Proceedings of the Fifth IS&TSID Color Imaging Conference. Color Science, Systems and Applications*, 1997, pp. 116–119.

[5] T. Vuori and M. Olkkonen, “The effect of image sharpness on quantitative eye movement data and on image quality evaluation while viewing natural images,” in *Proceedings of SPIE*, 2006, vol. 6059, p. 605903.

[6] A. Ninassi, O. Le Meur, P. Le Callet, D. Barba, and A. Tirel, “Task Impact on the Visual Attention in Subjective Image Quality Assessment,” in *The 14th European Signal Processing Conference*, 2006.

[7] P. Le Callet, S. Perchard, S. Tourancheau, A. Ninassi, and D. Barba, “Towards the next generation of video and image quality metrics: Impact of display, resolution, content and visual attention in subjective assessment,” *Second International Workshop on Image media Quality and its Applications*, vol. 83, no. 73.05, pp. 10–51.

[8] C. T. Vu, E. C. Larson, and D. M. Chandler, “Visual Fixation Patterns when Judging Image Quality: Effects of Distortion Type, Amount, and Subject Experience,” in *IEEE Southwest Symposium on Image Analysis and Interpretation, 2008. SSIAI 2008*, 2008, pp. 73–76.

[9] G. Abdollahian, Z. Pizlo, and E.J. Delp, “A study on the effect of camera motion on human visual attention,” *15th IEEE International Conference on Image Processing, 2008. ICIP 2008*, pp. 693–696, 2008.

[10] I.E.G. Richardson, *H. 264 and MPEG-4 video compression*, Wiley Chichester, 2003.

[11] Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, “H.264/avc jm reference software,” *http://iphome.hhi.de/suehring/tml/*, August 2008.

[12] J. S. Perry, “Xgl toolbox,” *http://fi.cvis.psy.utexas.edu/software.shtml*, 2008.

[13] D. Sheskin, *Handbook of Parametric and Nonparametric Statistical Procedures*, CRC Pr I Llc, 2004.

[14] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.