Task Dependence of Visual Attention on Compressed Videos: Point of Gaze Statistics and Analysis

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

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

ABSTRACT

We tracked the points-of-gaze of human observers as they viewed videos drawn from foreign films while engaged in two different tasks: (1) Quality Assessment and (2) Summarization. Each video was subjected to three possible distortion severities – no compression (pristine), low compression and high compression – using the H.264 compression standard. We have analyzed these eye-movement locations in detail. We extracted local statistical features around points-of-gaze and used them to answer the following questions: (1) Are there statistical differences in variances of points-of-gaze across videos between the two tasks?, (2) Does the variance in eye movements indicate a change in viewing strategy with change in distortion severity? (3) Are statistics at points-of-gaze different from those at random locations? (4) How do local low-level statistics vary across tasks? (5) How do point-of-gaze statistics vary across distortion severities within each task?

1. INTRODUCTION

Task dependence of eye movements when a subject views an image/video is a topic of considerable interest,[1,2] however, little work has been done in analyzing how the presence of distortions in an image/video affect visual eye movements.[3–6] We have undertaken a study on how distortions in compressed videos affect eye movements, when the subject views videos compressed using the recently proposed H.264 compression standard[7] for two different tasks – quality assessment and summarization.

Actually, there are five major types of eye movements, at least two i.e. saccade and smooth pursuit might be relevant here. Saccades are rapid, ballistic movements of the eyes that abruptly change the point of fixation while smooth pursuit movements are much slower tracking movements of the eyes done to keep a moving stimulus on the fovea. Human eye movements are made up of a series of fixations and rapid ballistic eye-movements.[2] Humans are believed to gather the most visual information during a fixation and little-to-no-information during a saccade. Researchers in vision science have studied human eye movements and fixations and point-of-gaze statistics in natural environments in order to understand how humans capture and process information about their environments, and how different low-level statistics attract points of gaze.[1,2,8]

Eye movement pattern analysis has also attracted some attention from the image and video quality assessment community.[3–6]Quality assessment refers to development of algorithms that are capable of predicting human judgement on the perceptual palatability of the stimulus.[9] Given that quality assessment is closely tied-in with points of gaze and modeling perceptual mechanisms using low-level indicators, it is not surprising that researchers in the field of quality assessment have attempted to use insights from human points of gaze modeling to better the performance of quality assessment (QA) algorithms. Generally, QA algorithms compute a perceptual map of local quality scores, which are combined using a 'pooling' technique (in many cases, the simple mean[9]), in order to produce a holistic measure of quality for the stimulus. If one could predict points of gaze using an algorithm and weight these local perceptual quality scores with weights proportional to the attention a region is likely to receive, one would imagine that some improvement in performance of QA algorithms in terms of correlation with human perception may be attained. Indeed, researchers have explored this area, with limited success,[10] where points of gaze probability map weights from the undistorted reference image were used to weight the quality scores of a distorted image. Here, however, our goal is different. We wish to understand how distortions in stimuli affect points of gaze. Specifically, we seek to understand how the local statistics of distorted stimuli vary

Further author information: (Send correspondence to Anish Mittal) : E-mail: mittal.anish@gmail.com

as a function of points of gaze. Although points of gaze study has not received much attention, especially for videos, some researchers have explored it for distorted images.

Miyala *et al.* conducted a subjective study where they tracked eye movement patterns of human observers when viewing distorted images. The images viewed were degraded by blur, noise and color shifts.[4] The authors concluded that distortions do not affect visual fixation patterns for the set of distorted stimuli that they considered. Ninassi *et al.*[3] introduced JPEG2000 and JPEG compression distortions in images and observed that these distortions indeed affect fixation; although they do not discuss how distortion severity affects it. Vu *et al.* undertook a comprehensive evaluation of how various distortion severities and various distortions (blur, noise, JPEG and JPEG2000 compression) affect human fixations.[5] The distorted images were viewed by human subjects performing two different tasks – free viewing and quality assessment. Based on the results from the study, the authors conclude that subjects tend to look at the same region, immaterial of the task for global distortions such as blur and noise; however for local distortions such as JPEG and JPEG2000 that manifest local artifacts and for suprathreshold distortion severities, there exist marked differences in viewing patterns. This does not hold for local distortions with below threshold distortion severities however.

Although researchers have studied the points of gaze when viewing natural videos (for eg., see ref.[11]), not much has been done in the area of points of gaze when viewing distorted videos. The closest in concept to ours is the work by by Abdollahian *et. al.* who analyze eye movement patterns for different camera motions.[12] Here, we analyze human visual fixations when viewing distorted stimuli for two specific tasks – quality assessment and summarization. Most of the research cited above utilize a 'free-viewing' task as the alternative task. As we pointed out in the an early of this work,[13] completely detaching top-down influences from a visual task may be impossible, however, the influence of top-down strategies may be minimized to a certain extent by proper definition of the alternative task. A 'free-viewing' condition results in the subject utilizing his own top-down instruction, thereby rendering any objective comparison of attentional results from these tasks with quality assessment impossible. A strict definition of the alternative task however, will aid such a comparison. Here, we greatly extend the preliminary work in ref. 13, by not only analyzing eye movements locations using simple statistics, but also by extracting relevant low-level features at point of gaze and analyzing the relationship of these point-of-gaze statistics across tasks and distortion severities. To the best of our knowledge, the present work is the first systematic attempt at evaluating point-of-gaze statistics for distorted videos.

In the rest of the paper we summarize the tasks and the videos used in the study as well as the distortions introduced in the view. We then describe the various low-level statistical indicators we extract at point-of-gaze and analyze these point-of-gaze statistics across tasks and distortion severities. We then discuss how we plan to extend this work to improve quality assessment algorithm performance.

2. SUBJECTIVE EVALUATION OF THE TASK DEPENDENCE OF EYE MOVEMENTS

2.1 Task Description

Human eye movements were recorded for two different tasks – summarization and quality assessment. In the former, the subject was shown a frame from one of the videos that the subject had viewed after a set of videos and asked to summarize what happened in that particular scene. In the latter, the subject was asked to rate the quality of the video after each presentation on a 5-point scale. As we have mentioned, such a rigid task description will reduce the effect of top-down influences on eye movements. Based on the data collected, we seek to assess if there exist any differences in viewing strategies between the two tasks.

2.2 Video Description

Twenty video clips from various foreign films (French, German), each 30 seconds long were chosen such that each clip had enough content to allow for summarization. Foreign films were chosen in order to minimize the effect of memory or recall in case the subject had viewed the films before. In order to create distortions in the videos, the H.264/AVC encoder[7] was used. The details of the parameters used are given in.[13] Each video was compressed at two different bit-rates and labeled 'high' and 'low' compression/distortion depending on the amount of bit rate allocation. Thus we created a total of 60 videos (40 compressed + 20 reference), each of resolution 720×480 with a frame-rate of 30 frames per second (fps). The compressed videos exhibit blocking and blurring artifacts.

2.3 Experiment Details

A total of 12 subjects participated in the study. Six subjects were assigned to each of the two tasks - quality assessment and summarization. Each subject viewed 20 videos in one session, lasting less than 30 minutes. 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) contents but with varying levels of quality (pristine, low distortion, high distortion) which was randomly assigned. Since we assigned six subjects per task, we obtained two sets of fixations for each of the 60 videos for each task.

The eyetracker used to record eye movement patterns was the one manufactured by Cambridge Research Systems with a 50 Hz rate, The stimuli were displayed on a 21" LCD monitor with a resolution of 1600×1200 and a refresh rate of 60 Hz, and viewed by subjects at a distance of 1000 mm. 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.[14] The subject was briefed before the actual study. Once the eye movements were obtained, we discovered that for one of the two subjects assigned to video 12 (summarization, pristine) the eye movements were un-reliable, as indicated by the eye-tracker software. Hence, we chose to neglect that particular video content across distortion severities (pristine, low distortion, high distortion) and tasks (quality, summarization), thereby reducing the number of videos in the study to 57.

3. EXTRACTING POINT-OF-GAZE STATISTICS

Although researchers in the field have used saccade duration[6] and fixations[5] previously, we choose to use eye movement patterns for analysis. This eliminates the need to arbitrarily define a fixation and its duration, while allowing for statistical analysis. Eye movements that were deemed un-reliable (as defined by the eye-tracker software) as well as those that were outside the video (recall that the videos were displayed at the center of a black screen) were discarded before any analysis was performed.

Figure 1 shows the eye movement locations across tasks and distortion levels for one video from the study from a single subject. 200 eye movement locations are plotted on an intermediate frame in order to give an approximate idea of where the subject looks in a video for a specific distortion level and for a particular task. No scene change occurs for the set of frames chosen here. It is interesting to note that fixations are distributed all over the video when the subject is shown the pristine video for the QA task; however, with an increase in distortion severity, the attentional patterns seem to converge at particular locations, which may ostensibly correspond to distorted regions. For the summarization task, subjects tend to look more towards the regions of activity in the scene, however, the distribution of eye movement locations changes with the level of distortion. It is also interesting to note the actual fixation locations for this particular video. We draw the reader's attention to Fig. 1 (b) and (f) – summarization task, pristine vs. high distortion case. While in the pristine case the subject tends to look at obvious attractors such as faces and edges (shadow-edge in this case), the high distortion case leads to a change in strategy – the subject tends to look at smooth regions such as the walls. This is very interesting indeed, for compression distortions often manifest themselves in regions of low texture/activity. This gives us a hint of what to expect when we extract point-of-gaze statistics.

In Fig. 2, we plot all the eye movement locations, across videos and subjects for each task and for each distortion severity considered here. It should be clear to the reader that the density of eye movement locations is highest around the center of the scene for the summarization task, in accordance with previous research.[15] However, for the quality assessment task, there appears to be a greater spread in eye movement locations. This seems intuitive and our hypothesis is that the quality assessment task necessitates a deeper scrutiny of the video and hence the subject scans the entire video, as against the summarization task in which the subject tracks objects that attract his gaze.

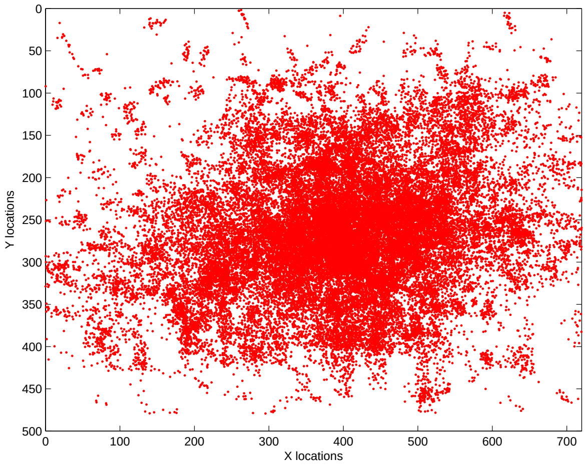
In order to understand how low-level features at point of gaze influence points of gaze, we computed a series of statistical features that may be attractors of points of gaze. The following statistics were computed at points of gaze using a $105 \times 105 \times 3$ - sized window around each eye movement location, which approximately corresponds to the size of fovea ($1.5°$) at a viewing distance of 1000 mm:

(a) (b)

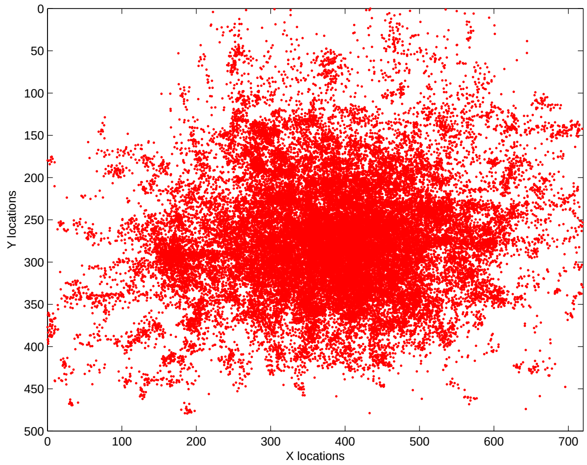
(c) (d)

(e) (f)

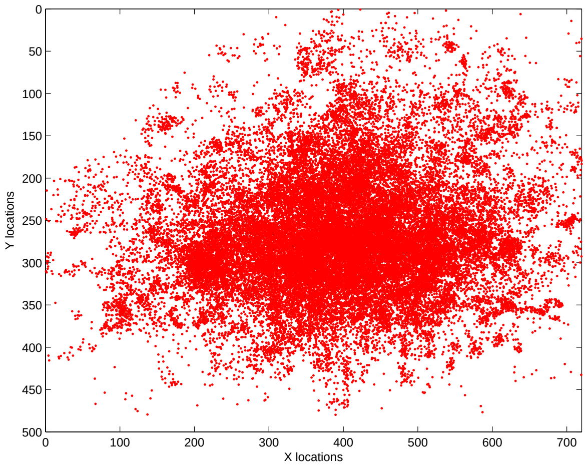
Figure 1. (a), (c) and (e) show the fixation patterns for the quality assessment task across a set of frames for (a) pristine, (c) low and (e) high distortion levels. (b), (d) and (f) show the same for the summarization task.



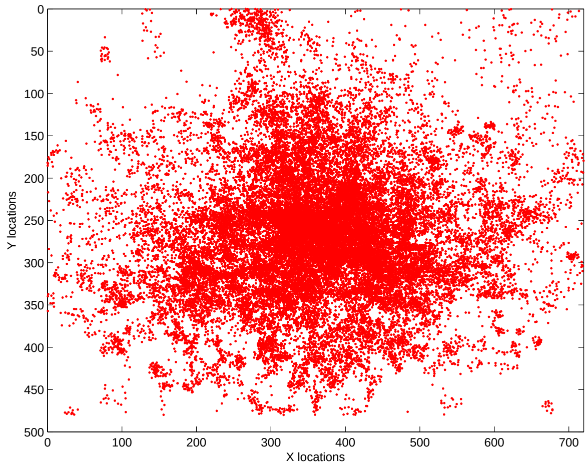
(a)



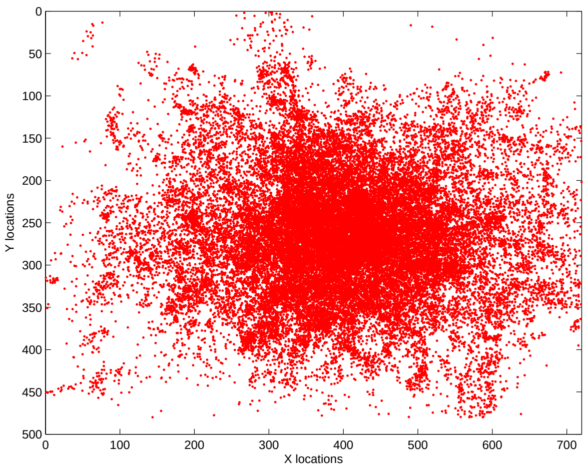
(b)



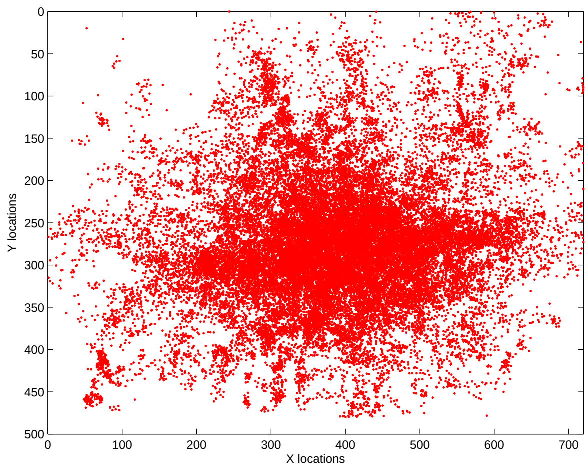
(c)



(d)

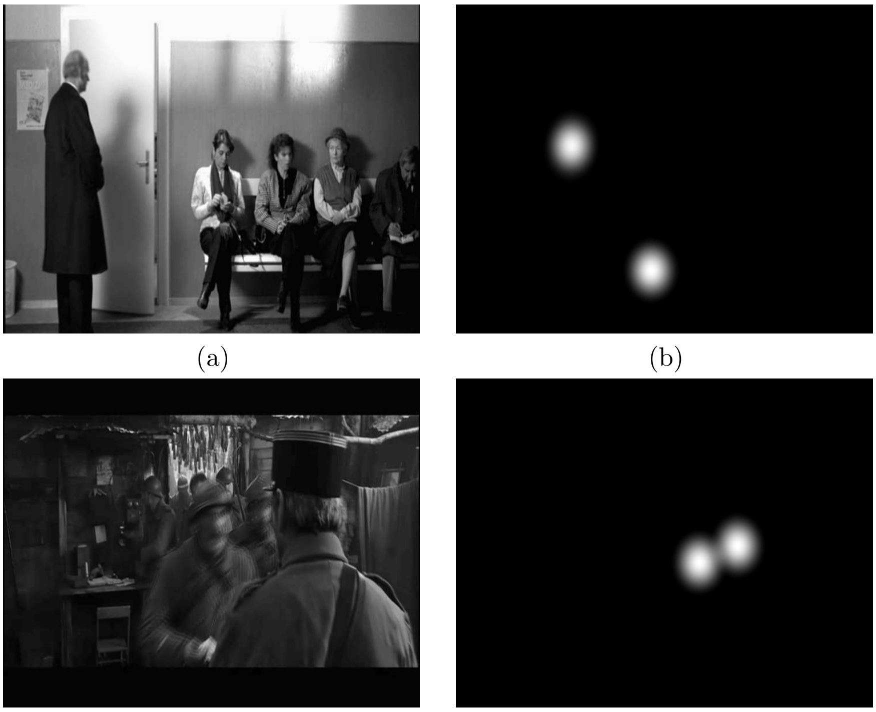


(e)



(f)

Figure 2. (a), (c) and (e) show the fixation patterns for the quality assessment task for all videos for (a) pristine, (c) low and (e) high distortion levels. (b), (d) and (f) show the same for the summarization task.



(a) (b)

(c) (d)

Figure 3. (a),(c) shows frames from 2 videos and (c),(d) shows corresponding average fixation probability maps.

1. Mean luminance = μ, the mean of the local patch luminance values,
2. RMS contrast = $\frac{\sigma}{\mu}$, the ratio of the standard deviation of patch luminance values to the mean luminance,[8]
3. Quality, using the full-reference structural similarity index (SSIM)[16] between the pristine and distorted videos (computed only for the distorted videos),
4. Average motion magnitude, where motion vectors are extracted using an optical flow estimation algorithm[17] and
5. Motion magnitude coefficient of variation(COV) = $\frac{\sigma}{\mu}$, the ratio of the standard deviation of patch motion magnitude values to the average motion magnitude

For each point-of-gaze, we first weight a region ($105 \times 105 \times 3$) around the point-of-gaze using a spatio-temporal weighting function, using separable spatio-temporal raised-cosine windows form the weighting function. Specifically, each patch is first weighted by a spatial raised-cosine window (details), followed by a one-dimensional temporal raised-cosine window (details). These cosine-weighted regions are then averaged across subjects to produce a mean point-of-gaze map. The computed statistics are then weighted by this spatio-temporal weighting function. For example, in Fig. 3, we plot the weighted point-of-gaze maps for 2 frames from 2 different videos. We note that there may be multiple points of gaze per subject for each frame of the video, since the eye-tracker rate is greater than the video frame-rate. However, this does not affect our analysis in any way. Further, some of the frames may have no recorded eye movement locations (due to lack of reliable estimates for example). Again, this does not affect our analysis since we consider only valid points-of-gaze.

4. RESULTS

We compute the above described low-level statistics around points-of-gaze across videos and then concatenate all observations to form a large vector for each distortion severity and for each task. Further analysis is based

Table 1. Results of a significance test between tasks: Variance of Eye Movements. A value of '1' or '-1' indicates that null hypothesis can be rejected at 95% confidence level and the two sets of eye movements are significantly different. A value of 0 indicates that the null hypothesis could not be rejected. A positive sign indicates that the statistic has greater mean for the summarization task than the quality assessment task, while a negative sign indicates the opposite. The number in brackets is the p-value.

Statistics	Significance		
	Pristine	Low Distortion	High Distortion
	Summarization vs Quality	Summarization vs Quality	Summarization vs Quality
Variance	-1(0)	0(0.0854)	-1(0.0393)

Table 2. Results of a significance test between tasks: Variance of Eye Movements. A value of '1' or '-1' indicates that null hypothesis can be rejected at 95% confidence level and the two sets of eye movements are significantly different. A value of 0 indicates that the null hypothesis could not be rejected. A positive sign indicates that the statistic has greater mean for the first distortion severity listed while a negative sign indicates the opposite. The number in brackets is the p-value.

Statistics	Significance					
	Quality Task			Summarization Task		
	Pristine vs Low	Pristine vs high	Low vs high	Pristine vs Low	Pristine vs high	Low vs high
Variance	0(0.0949)	0(0.2660)	1(0)	1(0)	0(0.1508)	0(.0982)

on statistical comparisons of these statistics. It is prudent to point out that our analysis implicitly assumes that each element of these vectors is independent from the other elements. In order to verify if there exist statistical differences between local features at points-of-gaze across tasks and distortion severities, we perform multiple experiments.

The analysis below is performed using the same procedure. The null hypothesis is that the two sets of data under considerations are drawn from the same distribution. In order to verify this hypothesis, we pool the set of observations for each of the two datasets, depending on the analysis (for eg., pristine quality vs. pristine summarization) to form a large vector. We then bootstrap sample from this vector *twice*, 10,000 times and compute the mean/median difference between the two sets of bootstrapped samples. Once such a sampling distribution of mean/median differences is formed, we check if the sample mean/median difference (for the two sets of data under question) lies in the tail of the sampling distribution. All our analysis is performed at the 95% confidence level (i.e, a significance level of $\alpha = 0.05$). If the sample mean/median difference lies in the tail, then the null hypothesis fails and we conclude that the two sample data are not drawn from the same distribution. This technique not only indicates if the two datasets are statistically different, but also allows us to specify if the mean/median of one set of observations is greater than or lesser than the mean/median observations from the other set.

4.1 Experiment I: Variance of Eye Movements

We seek to answer two questions: (1) Are there statistical differences in variances across videos between the two tasks? and (2) Does the variance in eye movements indicate a change in viewing strategy with change in distortion severity? To verify if the variance of eye movement locations are different across tasks and distortions, we first calculate the distance of each point-of-gaze from the top-left most point (i.e., $(0, 0)$) and then find the variance of these distances. Notice that there is no bias due to the absolute location of the eye-movements, since the variance is inherently computed about the mean. Such a computation is performed for each video separately and the vector so formed is used for the bootstrap-based analysis describe above. Realize that we seek to evaluate whether the mean variance of one task is greater than/lesser than/equal to the mean variance of the other.

The results are tabulated in Tables 1 and 2. From Table 1 statistics, we conclude that the mean variance is significantly larger for the quality assessment task as against that for the summarization task for the pristine and high distortion case. The null hypothesis could not be rejected for the low distortion case, however the p-value is not that much larger than the significance level. This is intuitive, since this implies that the subject scans the entire video in the quality assessment task, across distortion severities, while the summarization (i.e., normal viewing) leads to a more concentrated viewing strategy.

Table 3. Results of a significance test for each task: Statistics at Points-of-Gaze vs. Random Locations. A value of '1' or '-1' indicates that null hypothesis can be rejected at 95% confidence level and the two sets of eye movements are significantly different. A value of 0 indicates that the null hypothesis could not be rejected. A positive sign indicates that the statistic has greater mean for the first distortion severity listed while a negative sign indicates the opposite. The number in brackets is the p-value.

Statistics	Significance	
	Quality Task	Summarization Task
	Fixation vs Random Loc	Fixation vs Random Loc
Mean luminance	-1(0.0055)	0 (0.4385)
RMS contrast	1(0)	1(0.0005)
Average motion magnitude	-1(0)	-1(0)
Motion magnitude COV	0(0.6133)	1(0.003)

4.2 Experiment II: Statistics at Points-of-Gaze vs. Random Locations

We ask the following question: Are statistics at points-of-gaze different from those at random locations? In order to compute low-level statistics at random locations, we use the strategy proposed in.[8] Specifically, for a video under consideration, we select random points-of-gaze, where the locations are drawn from one of the rest of the videos in the set, uniformly at random. The reason to extract random patches in this fashion is summarized in.[8] A comparison between point-of-gaze statistics and random location statistics are computed for the pristine-only case for all the low-level measures considered here, except for the quality task – where quality statistics are available only for the distorted videos. We compare the median of the low-level statistic in question, across videos using the boot-strapping strategy described before. The choice of the median relates to the relative robustness of the median compared with the simple mean and is relevant in the random local statistics case. The results of such an analysis are tabulated in Table 3.

Based on the results in Table 3, one would infer that for the quality assessment task, regions with lower luminance draw gaze, while such a difference is not visible for the summarization task. For local contrast, on the contrary, regions with higher contrast seem to draw gaze, regardless of the task. Regions with lower motion seemingly draw gaze across tasks – this finding has deep implications for video quality assessment algorithm design, since this implies that a masking model that adheres to this bias of low-motion regions, would ideally exhibit greater correlation with human perception. Local magnitude variance for the summarization task is seemingly lower than those at random locations; quality assessment does not exhibit any bias of this sort.

4.3 Experiment III: Point-of-Gaze Statistics for Quality Assessment vs. Summarization

We ask the following question: How do local low-level statistics vary across tasks? In order to determine how the task influences low-level features at points-of-gaze, we compare the means between the 2 groups of statistics – low-level features from summarization and quality assessment, for each distorted severity in the study. The bootstrap-based approach described before was used to evaluate statistical significance. The results are tabulated in Table 4.

Based on the tabulated results, one would conjecture that on an average, higher local intensity at pristine and low distortion levels draws eye movements in the summarization task, as against the quality task; this trend is not reflected in the high-distortion case. Local contrast does not seem to follow any such general trends. SSIM quality index differences are significant when summarization and quality task are compared at both distortion levels. But humans tend to look at lower quality regions for high distortion level videos for quality assessment task than summarization. A reverse trend is seen at low distortion levels. Points of gaze during summarization, in the absence of distortion is seemingly drawn toward high-motion regions; but the statistics are not significantly different with the addition of distortions. These results seem to indicate that although there is a difference in the search strategy across tasks, there does not seem to be a trend across distortion severities.

4.4 Experiment IV: Within Task, Across Distortion Statistics

The question we answer is: How do point-of-gaze statistics vary across distortion severities within each task? Our analysis proceeds along the same lines as described above, and the results are tabulated in Table 5.

Table 4. Results of a significance test for each task: Statistics at Points-of-Gaze vs. Random Locations. A value of '1' or '-1' indicates that null hypothesis can be rejected at 95% confidence level and the two sets of eye movements are significantly different. A value of 0 indicates that the null hypothesis could not be rejected. A positive sign indicates that the statistic has greater mean for the summarization task than the quality assessment task, while a negative sign indicates the opposite. The number in brackets is the p-value.

Statistics	Significance		
	Pristine	Low Distortion	High Distortion
	Summarization vs Quality	Summarization vs Quality	Summarization vs Quality
Mean luminance	1(0)	1(0)	-1(0)
RMS contrast	-1(0)	1(0)	-1(0.0005)
SSIM (wrt pristine)	-	-1(0.0002)	1(0)
Average motion magnitude	-1(0.03)	0(0.20)	-1(0.02)
Motion magnitude COV	1(0)	0(0.13)	0(0.23)

Table 5. Results of a significance test for each task: Within Task, Across Distortion Statistics. A value of '1' or '-1' indicates that null hypothesis can be rejected at 95% confidence level and the two sets of eye movements are significantly different. A value of 0 indicates that the null hypothesis could not be rejected. A positive sign indicates that the statistic has greater mean for the first distortion severity listed while a negative sign indicates the opposite. The number in brackets is the p-value.

Statistics	Significance					
	Quality Task			Summarization Task		
	Pristine vs Low	Pristine vs high	Low vs high	Pristine vs Low	Pristine vs high	Low vs high
Mean intensity	1(0)	-1(0)	-1(0)	1(0.0002)	1(0)	1(0)
RMS contrast	1(0)	1(0)	-1(0)	1(0.0013)	0(0.8048)	-1(0)
SSIM (wrt pristine)	-	-	1(0)	-	-	1(0)
Average motion magnitude	0(0.70)	1(0)	1(0)	-1(0.04)	1(0)	1(0)
Motion magnitude COE	0(0.73)	-1(0)	-1(0)	1(0)	0(0.57)	-1(0)

In the quality assessment task, mean luminance and contrast statistics are significantly different across distortion severities. This implies that even within a particular task, the presence of distortion affects viewing behavior. Similar results generally hold for the summarization case. Local motion magnitudes are significantly higher in the presence of distortion, especially in the high distortion case – the subject seems to notice regions of high motion. This observation coupled with the generic observation on low-motion-magnitude regions implies that, on an average, humans tend to look at regions with low motion, however, the presence of distortion draws human gaze away from these regions. A reason for this could be that error-concealment fails at high-motion regions and hence the extreme nature of the distortions draw points of gaze. Similar arguments hold for the summarization task as well.

5. CONCLUSION AND FUTURE WORK

We tracked the points-of-gaze of human observers as they viewed videos drawn from foreign films for two different tasks: (1) Quality Assessment and (2) Summarization. Each video was subjected to three possible distortion severities – no compression (pristine), low compression and high compression – using the H.264 compression standard. Here, we analyzed these eye-movement locations in detail. We extracted local statistical features around points-of-gaze and based on the analysis we concluded that: (1) There exist definite statistical differences in variances of eye movements across tasks and that the subject scans more of the video in the quality assessment task; (2) The viewing strategy changes as function of the distortion severity; (3) When compared with random gaze locations, there seems to be a definite difference in points-of-gaze and the differences are functions of the task; (4) Low-level statistics have certain characteristic properties for each distortion severity and task and (5) Within each task, point-of-gaze statistics have distortion-dependent properties, although a clear trend as a function of distortion severity could not be unearthed.

Future work will involve increasing the number of subjects in order to increase the accuracy of the point-of-gaze locations. The experiments presented in this work have applications in developing pooling strategies for video quality assessment algorithms and our future work will involve an approach similar to that in ref.,[8] where a point-of-gaze prediction algorithm that is a function of the task and takes distortion severity into account will be created.

REFERENCES

[1] Wooding, D. S., "Eye movements of large populations: Ii. deriving regions of interest, coverage, and similarity using fixation maps," *Behavior Research Methods, Instruments, & Computers* **34**(4), 518 (2002).

[2] Yarbus, A. L., [*Eye movements and vision*], Plenum press (1967).

[3] Ninassi, A., Le Meur, O., Le Callet, P., Barba, D., and Tirel, A., "Task impact on the visual attention in subjective image quality assessment," *14th European Signal Processing Conference, Florence* (2006).

[4] Miyata, K., Saito, M., Tsumura, N., Haneishi, H., and Miyake, Y., "Eye movement analysis and its application to evaluation of image quality," in [*The Fifth Color Imaging Conference: Color Science, Systems, and Applications*], 116–119 (1997).

[5] Vu, E. C. L. and Chandler, D. M., "Visual fixation patterns when judging image quality: Effects of distortion type, amount, and subject experience," in [*Image Analysis and Interpretation, 2008. SSIAI 2008. IEEE Southwest Symposium on*], 73–76, IEEE (2008).

[6] Vuori, T. and Olkkonen, M., "The effect of image sharpness on quantitative eye movement data and on image quality evaluation while viewing natural images," in [*Proceedings of SPIE*], **6059**, 605903 (2006).

[7] Richardson, I. E. G., [*H. 264 and MPEG-4 video compression*], Wiley Online Library (2003).

[8] Rajashekar, U., van der Linde, I., Bovik, A. C., and Cormack, L. K., "GAFFE: A gaze-attentive fixation finding engine," *Image Processing, IEEE Transactions on* **17**(4), 564–573 (2008).

[9] Wang, Z. and Bovik, A. C., "Modern Image quality assessment," *Synthesis Lectures on Image, Video, and Multimedia Processing* **2**(1), 1–156 (2006).

[10] Moorthy, A. K. and Bovik, A. C., "Visual importance pooling for image quality assessment," *Selected Topics in Signal Processing, IEEE Journal of* **3**(2), 193–201 (2009).

[11] Meur, O. L., Callet, P. L., and Barba, D., "Predicting visual fixations on video based on low-level visual features," *Vision Research* **47**(19), 2483 – 2498 (2007).

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

[13] Moorthy, A. K., Geisler, W. S., and Bovik, A. C., "Evaluating the task dependence of eye movements for compressed videos," *Fifth International Workshop on Video Processing and Quality Metrics for Consumer Electronics, Scottsdale, Arizona* (2010).

[14] Perry, J., "Xgl toolbox." http://fi.cvis.psy.utexas.edu/software.shtml (2008).

[15] Tatler, B. W., "The central fixation bias in scene viewing: Selecting an optimal viewing position independently of motor biases and image feature distributions," *Journal of Vision* **7**(14) (2007).

[16] Wang, Z., Bovik, A. C., Sheikh, H. R., and Simoncelli, E. P., "Image quality assessment: From error visibility to structural similarity," *Image Processing, IEEE Transactions on* **13**(4), 600–612 (2004).

[17] Black, M. J. and Anandan, P., ""the robust estimation of multiple motions: Parametric and piecewise-smooth flow fields"," *Computer Vision and Image Understanding* **63**(1), 75–104 (1996).