DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-117

Role of spatio-temporal distortions in the visual periphery in disrupting natural attention deployment

Yashas Rai, Marcus Barkowsky, Patrick Le Callet
LUNAM University, University of Nantes, IRCCyN UMR CNRS 6597
Polytech Nantes, Rue Christian Pauc BP 50609 44306
Nantes Cedex 3, France
{ *Yashas.Rai, Marcus.Barkowsky, Patrick.Lecallet* } @ *univ-nantes.fr*

***Abstract*—Human visual system based quality metrics and perceptually optimized video coders often use principles of foveation and saliency to weigh the distortion in certain regions more heavily or hide the artefacts in regions where they are less noticeable. These approaches however fail to consider the impact such a tuning produces on the non-salient surroundings usually incident on the para, peri and extra-peri foveal visual regions. Vision studies on the other hand, have highlighted the enhanced sensitivity of these peripheral visual regions towards spatio-temporal artefacts: more so in the supra-threshold region. Because such analysis has often been performed using controlled synthetic stimuli and forced fixation based experimental approaches, that assume perfect luminance adaptation, tracking and semantic comprehension of underlying content, a thorough understanding of the impact of peripheral disturbances in a natural viewing scenario is missing. The present work therefore uses a Gaze Contingent Display to study the impact of spatio-temporal distortions in the peri foveal and extra-peri foveal regions in a free-viewing scenario, using natural scene stimuli. Using four state of the art gaze analysis-techniques to analyze the gaze data collected from 48 observers, spatio-temporally and semantically, confirms and extends our previous understanding of distortion perception in the periphery. Our observations indicate that non-flickering spatial distortions seem to have less of a disruptive effect in the visual periphery as compared to the temporally flickering artefacts and second, the threshold at which disruptions begin to occur is higher in the visual periphery as compared to that of the fovea, both of these effects being strongly scene dependent and prone to natural scene masking. The results highlight the need for sufficient consideration of the supra-threshold effects of peripheral distortions, in order to achieve an optimum perceptual experience.**

Keywords*—Attention Disruption, Free Viewing, Peripheral Sensitivity, Para-Fovea, Peri-Fovea, Spatio-Temporal Supra-Threshold Distortions, Gaze Contingent Display, H.265 Encoder, Gaze Analysis Techniques, Semantic Gaze Analysis

I. INTRODUCTION

Presented a visual scene spanning several degrees of viewing angle, the visual system is not able to comprehend this scene in full detail throughout the spatial extent and crisp sampling of spatial information is instead limited to a few degrees around the point of gaze[1]. We constantly make explicit eye movements (overtly attend) in order to bring the area of interest in a scene into this central region, the fovea, the other surrounding regions being incident onto the para, peri and extra-peri foveal regions (attended covertly) which in turn play a major role in determining future attention points[2][3]. Each retinal region onto which the scene is projected, has a very unique anatomy[4] and also a spatio-temporal sensitivity characteristic. A perceptually tuned system must therefore aim to optimize each of these regions separately.

There have been several earlier attempts[5][6] to study the spatio-temporal response of the peripheral regions using controlled homogenous stimuli like Gabor patches or Sine wave gratings. From the perspective of spatial frequencies, most works highlight the concept of *Contrast Constancy* at Supra-Threshold levels where foveal spatial sensitivities are very similar to the peripheral spatial sensitivities after being scaled by their respective threshold sensitivities[6][7]. From a temporal perspective however, the peripheral area is observed to be almost as sensitive as the fovea, to temporal artefacts[5].

Amongst others, three important aspects that earlier studies using synthetic stimuli do not consider are that, there is a spatio-temporally localized and rapid adaptation of the eye to the varying luminance levels[8], that semantic comprehension of the underlying content is important to evaluate quality[9] and that a variety of inter channel-orientation masking effects also influences the sensitivity levels[10]. In addition, the forced tracking experimental methodology disturbs the naturalised temporal luminance adaptation patterns of the eye and also introduces an unnaturalness in the tracking capability[11], such a difference clearly visible in the experiments of Dorr et al[12]. To the best of our knowledge, an experiment investigating the peripheral spatio-temporal sensitivities in naturalistic scenes using a free viewing-condition has not been performed so far.

The drop in attention across the retinal region is often exploited in the form of a saliency window (in case of modelling the drop in physical sensitivity : also called the foveation filter) that models the decreased importance of the peripheral areas. Some attempts to apply such a foveation filter adaptively, in accordance to the gaze patterns of a subject, also known as a *Gaze Contingent Display(GCD)*, can be found in [13]. Several early works have used such a GCD setup to alter the spatial[14] or temporal[15] resolution of the video in accordance to the retinal eccentricity. A similar test-setup has also been used by the authors in an earlier study, to particularly analyze the effects of peripheral quality flicker in certain temporal frequency bands [16].

In the current work, we study spatio-temporal disturbances of varying strengths, frequency and amplitudes in the para, peri and extra-peri foveal retinal regions using a Gaze contingent display, the setup briefly introduced in Section II. While spatial distortions in this context refer to the various blurring, blocking and ringing artefacts[17] produced by a H.265 coder at coarse quantization levels, temporal flicker, also referred to as coarse/

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-117

fine grain flickering in literature, is a *highly eye-catching and perceptually annoying luminance variation along the temporal dimension, that is actually absent in the original sequence*[18].

Duchowski et al [19] define a *Perceptually Lossless GCD as the one in which, for a specified viewing distance and (instantaneous) gaze direction, the reconstructed stimulus (as viewed from the GCD) and the original stimulus appear identical to human observers*. To gauge the effects of peripheral disturbances, we examine the extent to which our test scenario statistically approaches a perceptually lossless GCD. Four state-of-the-art Spatio-Temporal and Semantic gaze data analysis techniques as described in Section III are used to analyze the otherwise noisy gaze data which confirms and extends our earlier findings performed by simpler techniques[16].

II. Experimental Setup

To maintain a natural (free-viewing like) gaze pattern, subjects were made to view the scene normally without the need to provide explicit quantitative measurements. Annoyance on the other hand, is quantified by examining the extent to which the gaze patterns are disturbed as compared to viewing a uniform high quality video[20]. The Gaze contingent display(GCD) discussed in Section I helps us regulate the exact position of the disturbance incident on the various retinal regions. Further description and details of the test setup maybe found in [16] and only the parts relevant to the discussion will be described here.

A. Test Preparation and Apparatus

The initial screening procedure included a Monoyer visual acuity test and an Ishihara color blindness test, in which three observers were rejected. The preparation also included a procedure to determine the dominant-eye of the observer considered.

The test equipment constituted of a TV Logic LVM401W display operated at 60Hz refresh rate and calibrated according to the specifications in BT.709. To obtain the gaze data, the SMI Hi-Speed eye-tracker was used in binocular viewing mode, thus providing 500 gaze samples per second. The test software was able to dynamically synthesize a frame containing the required peripheral disturbance in accordance with the gaze coordinates, and render it on the display at 60fps.

The test was performed with twelve sequences of full-HD resolution (1920x1080), 30 frames per second, each lasting ten seconds and covering a variety of genres ranging from outdoor scenes to a group of indoor objects, as shown in Figure 1. To reduce the overall dimensionality of the test without compromising on the ability to study scene dependent (spatio-temporal) effects, the sequence space was divided such that six sequences (ConstructionField, Fountains, Library, ResidentialBuilding, TallBuildings, NTIA-Redgold) were used for the peri-foveal stimulation while the remaining were used for the extra peri-foveal stimulation.

B. Generating the Test Stimuli

Spatial distortion of two different strengths and temporal disturbances of three types were tested in each: the peri and the extra-peri foveal regions. In addition to the high quality reference uniformly coded at $QP = 22$ (equal to that of the foveal region), a second uniformly degraded reference at

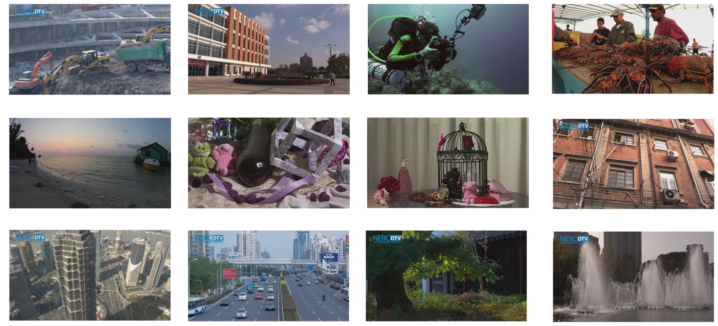


Fig. 1. Source sequences used for the test each chosen from various genres. From top-left row wise: Construction(SJTU), Library(SJTU), Diver(Liquid Assets), Lobsters(Liquid Assets), Evening Walk(Liquid Assets), NTIA Purple(NTIA), NTIA Red-Gold(NTIA), Residential Building(SJTU), Tall Buildings(SJTU), Traffic and Buildings(SJTU), Tree Shade(SJTU), Fountains(SJTU)

$QP = 39$ was also used in the test, for reasons that will become clearer in Section III. Figure 2 describes the spatial and temporal quality profile of the test stimulus as a function of its distance *d* from the centre of gaze (also called the foveola) and the time instant in ms, each test case referred to as a Hypothetical Reference Circuit (HRC). Regions in the peri-fovea are modulated in HRCs 3 to 7 whereas the extra peri-foveal regions are stimulated in HRCs 8-12 as shown in the parenthesis. From a spatial perspective, the quality rolls off in log-linear quantization step sizes starting from the foveal boundary till the tested peripheral region (para/peri foveal boundary), the reason being that past experiments observed QP to be linearly dependent on the observed perceptual quality[21]. Other than the foveal and peripheral streams, two other streams coded at intermediate QPs are used to build a smooth spatial quality transition. Adjacent zones in between these regions are filled up by linearly interpolating (weights indicated in figure) the stimulus obtained from the streams at either boundaries, so that blocking artefacts are not evident.

While this arrangement is invariant with time in the HRCs 3 and 4, the stimulus outside the central foveal region, is also varied with respect to time in case of HRCs 5-7, to create a temporal flickering effect: the exact QP variation profile determining the amplitude and frequency of flicker.

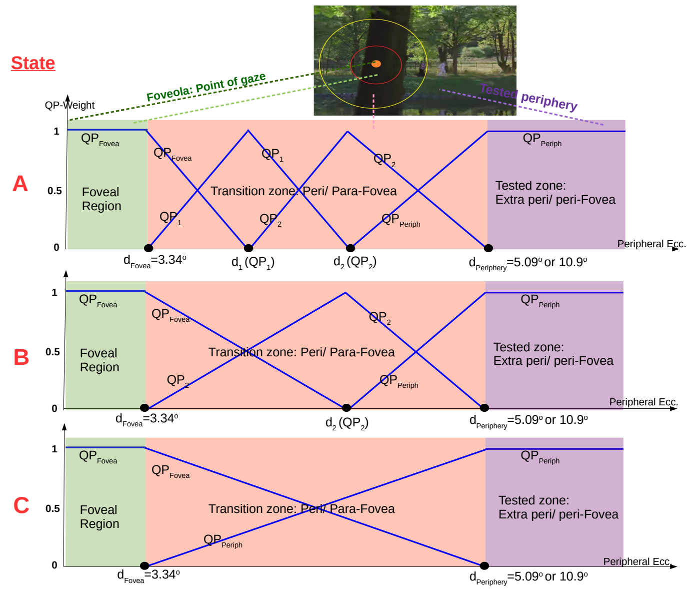
C. Performing the Test

The forty eight observers who passed the screening were then given a training, in which all the sequences to be presented in the test were displayed once, so that surprise (unexpected) based effects are minimized. These subjects were then instructed to watch the video naturally without being assigned any special task. An initial two seconds of uniform high quality helped the subject settle into the video content, following which, the mentioned distortions appear in the visual periphery.

III. Results

Local dispersion is first assessed using a sliding window approach that operates over 60ms which in turn helps classify the raw x-y data into saccades and non-saccades. Further, additional constraints are applied on the minimum amplitude of the saccades as well as the minimum time interval of the fixations to reliably categorize the data into fixations, smooth

©2016 Society for Imaging Science and Technology
DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-117



time	QP_{Fov}	HRC-3 (8) 0Hz, QP=[28,31,35]				HRC-4 (9) 0Hz, QP=[31,39,45]				HRC-5 (10) 7.5Hz, QP=[31,39,45]				HRC-6 (11) 15Hz, QP=[28,31,35]				HRC-7 (12) 15Hz, QP=[31,39,45]			
		St.	QP_1	QP_2	QP_{Per}	St.	QP_1	QP_2	QP_{Per}	St.	QP_1	QP_2	QP_{Per}	St.	QP_1	QP_2	QP_{Per}	St.	QP_1	QP_2	QP_{Per}
T+0ms	22	A	28	31	35	A	31	39	45	A	31	39	45	A	28	31	35	A	31	39	45
T+16.6 ms	22	A	28	31	35	A	31	39	45	A	31	39	45	B	-	28	31	B	-	31	39
T+33.3 ms	22	A	28	31	35	A	31	39	45	B	-	31	39	C	-	-	28	C	-	-	31
T+50 ms	22	A	28	31	35	A	31	39	45	B	-	31	39	B	-	28	31	B	-	31	39
T+66.6 ms	22	A	28	31	35	A	31	39	45	C	-	-	31	A	28	31	35	A	31	39	45
T+83.3 ms	22	A	28	31	35	A	31	39	45	C	-	-	31	B	-	28	31	B	-	31	39
T+100 ms	22	A	28	31	35	A	31	39	45	B	-	31	39	C	-	-	28	C	-	-	31

Fig. 2. Assuming that an observer is looking at the tree, the figure shows the variation of the QP across the visual regions. While the foveal region has a constant QP_{Fovea}, the tested peripheral zone has a spatially invariant, but in case of HRCs 5-7, a time varying QP of QP_{Periph} as illustrated by the states A,B,C. In the intermediate transition zone, the QP value linearly drops off with distance: intermediate QPs are used to aid the linear interpolation. The table describes the state at a given point of time and the intermediate QPs in each of these zones for all the HRCs. (-) in the table indicates the absence of an intermediate QP at the given time interval.

pursuits and saccades [22]. To lose as little data as possible, we consider all those saccades whose amplitude is greater than the extent of the central foveal region i.e $> 3.09^o$, and those fixations whose durations are greater than the minimum foveal object recognition time [23]. The consistency of the results obtained were also verified by varying these fixation-saccade threshold parameters.

After isolating the fast saccades from the gaze data[24], the remaining fixations and smooth pursuits are also used to compute a *saliency map* for every frame in the video, which indicates the relative probability with which the subjects gaze at various regions in that frame. A gaussian filter with $\frac{FWHM}{2} = 3.34^o$ (the foveal width) that models the drop in attention as we move away from the fovea, is then used to filter these fixation and smooth pursuit regions to compute a saliency map over the observers.

To further analyze the data semantically, we manually segment each of the ten second videos into clearly comprehensible objects and analyze the fixations that fall within these 3-D space-time bounding boxes. For attaining this objective, five naive observers first performed a memory recall task as in [25] in order to determine the important objects in

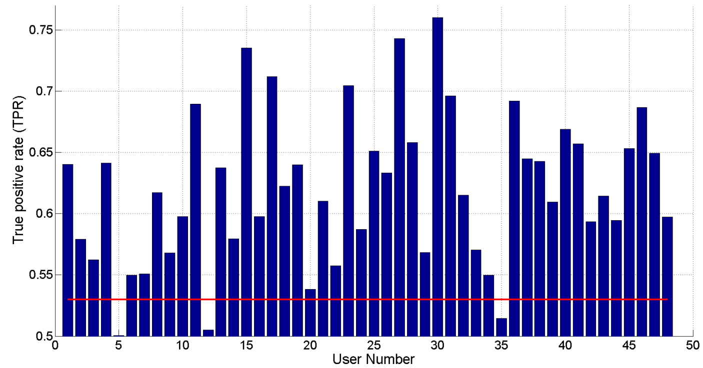


Fig. 3. The true positive rate(TP) at the equal error rate point(EER) averaged over all sequences. Those observers having a TP rate of close to 50% (below the red line) are rejected. The current analysis therefore rejects three inconsistent observers

the scene. These regions refer to the semantic boundaries of commonly identifiable objects rather than serving as a measure of saliency. Each ten second video is segmented into 10-12 spatio-temporal objects, with all regions outside these object boundaries considered as the background.

A. *Rejection of inconsistent observers*

In order to test the conformity of the gaze patterns of each observer to that of an average observer, a one against all *Receiver Operating Characteristic* (ROC) based hybrid approach[26] is used. We first compute a saliency map for the sequence under consideration, using all the 47 observers other than the observer being considered. The fixation patterns of the considered observer for the sequence under consideration is then laid out on a thresholded version of the saliency map, the hit rates of which form the true positive (TP). The fixation pattern of the same person for a different sequence other than the one being considered, is also laid out on the thresholded version of the saliency map to form the false positives (FP). With varying thresholds, an ROC curve as visible in Figure 4 is plotted. Finally, the value of TP rate at the point in the ROC where the miss rate (1-TP) equals the false positive rate (FP), a point also referred to as equal error rate (EER) is examined. This value that is averaged over all sequences, is considered to be a good indicator of the observer consistency. The Figure 3 shows us the TP rate of all the 48 observers at the EER point, based on which three observers need to be rejected.

B. *Vector similarity*

Simpler indicators like Fixation Duration and Saccade Amplitudes, although good indicators of perceived disturbance[16], tend to not exploit the spatial and temporal relationships inherent in gaze data. Therefore, a new method that simplifies the enormous gaze information, whilst still considering these relationships, in addition to the shape of the scanpath is essential[27]. Studying all the individual saccade shapes within a scanpath helps us understand not only the pattern of attention shifts for a subject, but is also an indirect indicator of the amount of period the observer spends fixating on a certain region.

For the current analysis, a scanpath simplification scheme similar to that proposed by Jarodzka et al [27] is used in

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-117

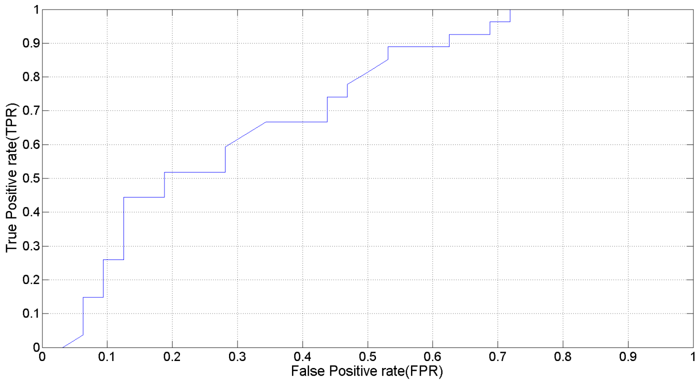


Fig. 4. Receiver operating characteristics for User 1 for the sequence ConstructionField. The false positives are obtained with the other sequences taken one at a time and averaging the results.

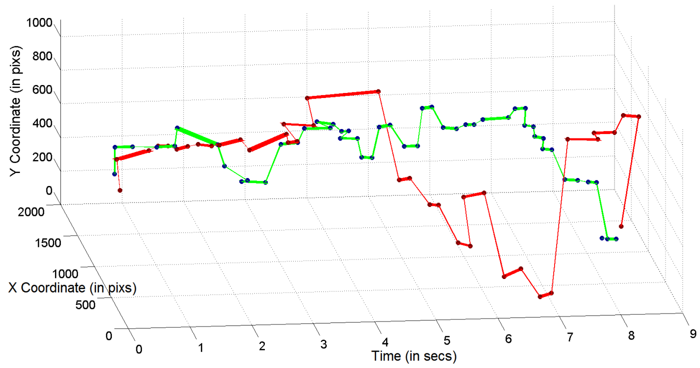


Fig. 5. Examples of simplified scanpaths for the test and the control sequence in which the thick lines indicate durations of fixations connected with saccadic thin lines. The dots indicate the transition between these two phases.

order to pool consecutive saccades having similar directions as visible in Figure 5. Each of the **m** individual saccades in the control gaze path is compared to that of the **n** saccades in the test path by considering three separate aspects of similarity: spatial proximity of the saccade starting points i.e 2D euclidean separation between their starting points, Difference in direction and magnitude as indicated by the vector difference of the saccades and third, the temporal proximity of the two saccade midpoints in which the saccade initiation latency of 219ms[28] is modelled as a rectangular function. Each of the three measures is normalised by the maximum possible value, i.e. the screen diagonal for the first two measures and are then averaged together to produce an overall similarity score ranging from 0 to 1. These scores are then tabulated in a table of size $m \times n$ through which we compute a least cost path starting from the first node to the bottommost right node. Using the Dijkstras algorithm, we obtain a least cost path, the path traversal cost indicating the overall similarity of the two paths. This cost has to be normalised in accordance to the path length for further comparison.

Using such a measure, a similarity metric may be obtained for any given user between the control and a test sequence. However, to compute a test statistic, we accumulate such scores over all the 45 observers and use a repeated measures ANOVA to check if there were significant differences in the similarity scores in the $Ref_0 - Ref_1$ uniform distortion case, as compared to the peripheral distorted cases. The similarity scores averaged across the observers along with the statistically significant cases (marked in bold) are shown in Table I.

TABLE I. P VALUES OF THE VECTOR SIMILARITY METRIC

Sequence	$HRC_{3(8)}$	$HRC_{4(9)}$	$HRC_{5(10)}$	$HRC_{6(11)}$	$HRC_{7(12)}$
Construction Field	0.17	0.46	0.86	0.06	0.28
Fountains	0.73	0.32	**0.05**	0.99	0.42
Library	0.09	0.08	0.39	0.27	0.45
Resid. Building	0.16	0.90	**0.02**	0.63	0.94
Tall Buildings	0.16	0.67	0.47	0.31	0.84
NTIA-Redgold	0.53	0.30	**0.03**	0.67	0.09
Diver	0.50	0.73	0.77	0.95	0.56
Lobsters	0.12	0.88	0.33	0.17	0.07
Evening Walk	0.10	**0.05**	**0.02**	0.30	0.19
Traffic Building	0.70	0.16	0.82	0.71	0.34
NTIA-Purple	0.43	0.96	**0.00**	0.41	0.46
Tree Shade	0.36	0.54	0.66	0.74	0.41

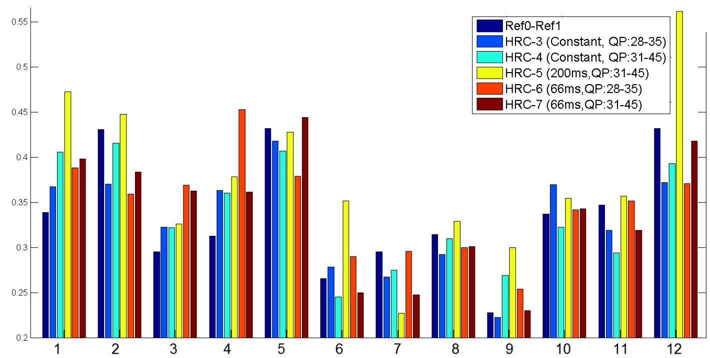


Fig. 6. Comparison of the KL-Divergence scores (averaged over all frames) calculated between the test and uniform quality foveal reference map.

C. Divergence of Saliency Maps

The reference and the test saliency maps can be compared like two 2D probability distributions using a K-L divergence criterion with the null hypothesis that the attention patterns in the reference scene I_{ref} and the test scene I_{test} are similar in nature. The KL-Divergence measures the amount of information (entropy based) that is lost by making such an assumption.

As Figure 6 indicates, non-flickering distortions $HRC_{3(8)}$ and $HRC_{4(9)}$ are in general slightly less annoying as compared to the temporal flicker in $HRC_{5(10)}$ and $HRC_{7(12)}$. The Ref_1 case, uniformly encoded at $QP = 39$, is often observed to be more divergent from the reference than the case when a distortion of $QP = 45$ in $HRC_{4(9)}$ is introduced in the peri or extra-peri fovea alone.

It is interesting to note that for several sequences, the Ref_1 having a foveal QP of 39 and also peripheral QP of 39, has a higher deviance than the HRC_4 case having a foveal QP of 22 and peripheral QP of 45, possibly indicating that the threshold for annoyance perception is much higher in the periphery as compared to the fovea.

D. ROI based disturbance measurement

For analyzing the data semantically, we examine the class of the region (Objects of interest / Background), most important to each of the observers during their viewing duration. Based on the opinion of the naive observers described in Section III, the regions in the video were categorized into objects of interest and the regions that were not considered as important by the observers (background region). The relative importance of the two regions were examined by observing *where* the user was mostly fixating in

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-117

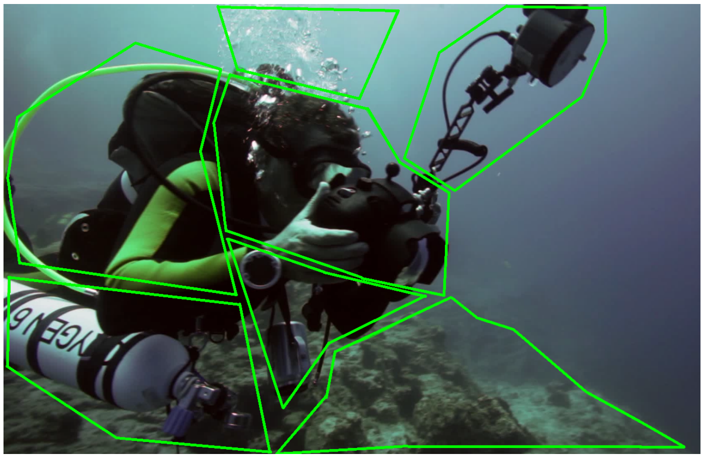


Fig. 7. A still frame from the diver sequence which shows some of the objects visible in the frame: The objects in this case are classified into Oxygen cylinder, the divers head and body, shooting gear, rocks and air bubbles

each viewing. The number of fixations that fall inside each of these 3D bounding boxes are counted, each 400ms considered as a separate fixation, in order to indicate the deployment of increased attention for longer fixations. Based on the region where most fixations are concentrated, we determine the preferred region of the observer.

To then perform a statistical test on the paired data, we count the number of observers who change their preferred region from object to background (and vice versa) between the control and the test videos for each of the videos presented. Table IV for example shows two such contingency tables generated for the sequence Construction Field in case of the $Ref_0 - Ref_1$ condition (on the left) and for the $Ref_0 - HRC_7$ case (on the right) respectively. A paired, Mc-Nemar Chi-Square test then compares the 2×2 contingency table with the null hypothesis that the introduction of peripheral disturbances has no special influence on the frequently fixated region.

TABLE IV. CONTINGENCY TABLES

Test / Ref	ROI	BG
ROI	26	10
BG	11	1

Test / Ref	ROI	BG
ROI	24	12
BG	6	5

With the fixation locations from 45 subjects used in the test, a high degree of test power was achieved so that we avoid missing any effects that exists in reality. Although sequence dependent effects were very prominent throughout the test, Table II describes the p-Values and the power of the test in those cases where a statistical significance is found (as indicated in bold). A strong difference in the distribution is found in most cases, especially in HRC_5 possibly indicating a high sensitivity of the periphery in the 7.5Hz band.

E. Attention Transitions

In order to exploit the sequential nature of gaze patterns in videos, it is essential to model the thought process of the subject by analyzing his attention transitions among the different objects of interest and finding similarities in the scan patterns[29] and the frequency of occurrence of the numerous local sub-patterns[30]. It has already been claimed that these scan paths are a strong function of the underlying task and stimulus and also that scan paths somehow recur with the repetition of the stimuli[3]. In case of the present experiment, if the GCD is a perceptually lossless one, we would assume that there would be significant similarities in the gaze patterns of the test video as compared to that of the reference.

TABLE II. OBJECT ATTENTION DISTRIBUTIONS- CHI.SQ TEST

Sequence	$Ref_0 - Ref_1$	$HRC_{3(8)}$	$HRC_{4(9)}$	$HRC_{5(10)}$	$HRC_{6(11)}$	$HRC_{7(12)}$
Construction Field	0.83	0.62	1.00	0.16	0.47	0.64
Fountains	0.78	0.23	0.37	0.29	0.23	0.32
Library	0.48	0.56	0.13	0.62	0.20	**0.05 (0.97)**
Resid. Building	0.18	0.16	0.65	0.64	1.00	0.59
Tall Buildings	0.44	0.16	0.11	0.83	0.47	1.00
NTIA-Redgold	0.59	0.35	0.29	**0.05 (0.98)**	0.62	0.56
Diver	0.81	0.32	0.80	0.49	0.81	1.00
Lobsters	0.76	0.80	0.53	0.76	0.13	0.74
Evening Walk	0.56	0.17	0.80	0.81	0.81	0.47
Traffic Building	1.00	0.74	0.32	**0.01 (0.43)**	0.20	0.11
NTIA-Purple	0.17	0.09	0.06	0.09	0.23	0.47
Tree Shade	**0.02 (0.95)**	**0.01 (0.99)**	0.37	**0.01 (0.99)**	0.21	0.74

TABLE III. LEVENSHTEIN SIMILARITY SCORES

Sequence	$Ref_0 - Ref_1$	$HRC_{3(8)}$	$HRC_{4(9)}$	$HRC_{5(10)}$	$HRC_{6(11)}$	$HRC_{7(12)}$
Construction Field	0.32	0.32	0.32	0.31	0.31	0.31
Fountains	0.23	0.23	0.24	0.25	0.23	0.23
Library	0.35	0.36	0.32	0.35	0.31	0.31
Resid. Building	0.34	0.31	0.31	0.32	**0.28**	**0.28**
Tall Buildings	0.21	**0.25**	**0.25**	0.24	**0.28**	**0.28**
NTIA-Redgold	0.38	0.37	**0.33**	0.36	**0.34**	**0.34**
Diver	0.38	0.36	0.37	**0.43**	0.36	0.36
Lobsters	0.36	**0.28**	0.33	0.36	0.36	0.36
Evening Walk	0.40	0.38	**0.35**	0.40	**0.36**	**0.36**
Traffic Building	0.27	0.27	0.25	0.26	**0.22**	0.22
NTIA-Purple	0.25	0.23	0.26	0.26	0.26	0.26
Tree Shade	0.41	0.40	0.42	0.41	0.40	0.40

A common way to compare two scanpaths, is to compute the Levenshtein distance which in turn indicates the number of string replace/insert/delete operations needed to transform the control string into the tested string[3]. A Levenshtein score normalised to the maximum length of the two compared strings, was used to compute the similarity of the scan patterns in the control condition as compared to that of the six other test cases.

To analyze if there were any significant differences in the Levenshtein similarity scores in the reference - uniformly degraded versus the reference - peripheral distorted cases, we perform a repeated measures ANOVA for the 45 observers. The average normalised Levenshtein similarity scores are indicated in Table III, where a higher number indicates a better Levenshtein similarity and statistically significant ($p < 0.1$) cases are marked in bold. The concentration of a large number of statistically significant differences in the last three columns of the table shows that, while spatial non-flickering disturbances in the periphery does not cause much deviance in viewing as compared to a uniformly degraded reference, temporally flickering distortions in the periphery significantly alters the viewing pattern of the observers.

IV. CONCLUSIONS

Presence of a gaze contingent setup helps us study the gaze patterns of a subject in a naturalistic manner whilst introducing spatio-temporal distortions in the visual periphery at the exact eccentricity we wish to investigate. The gaze data obtained from this setup was then analyzed spatio-temporally and semantically.

As content dependent effects are visible in several instances

DOI: 10.2352/ISSN.2470-1173.2016.16HVEI-117

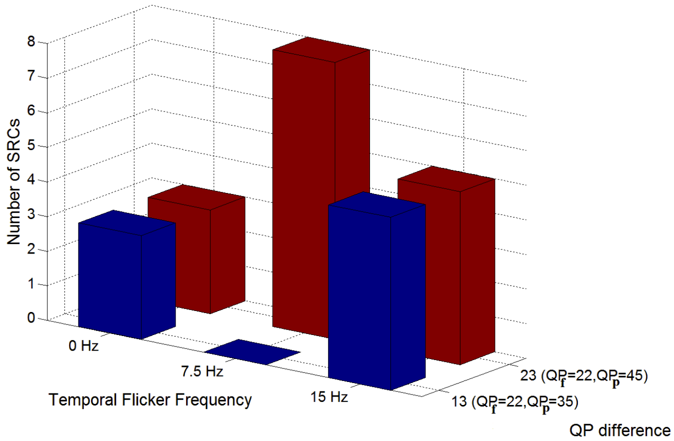


Fig. 8. Number of SRC sequences where a significant difference was found for the various temporal frequencies and distortion amplitudes in the periphery

due to the sequence specific features, the number of sequences where disturbances were statistically ascertained, either spatio-temporally or semantically is represented in Figure 8. The results indicate worse disruptions in case of temporal flicker (esp. in the 7.5Hz band), as compared to the non-flickering spatial distortions, in both: the peri and extra peri-fovea. The KL-Divergence values and semantic similarity scores seem to also indicate that the threshold at which the spatial artefacts starts to cause viewing disruptions are much higher in the visual periphery as compared to the foveal case.

Region of interest and *Foveation* based video coders that perform a local quality optimization must therefore be careful to restrict these peripheral artefacts within the threshold level, so that the natural way in which an observer watches a video remains unaltered.

V. Acknowledgment

This work was supported by the Marie Curie Initial Training Network under the PROVISION (PeRceptually Optimized VIdeo CompresSION) project.

References

[1] L. Itti and C. Koch, "Computational modelling of visual attention," *Nature reviews neuroscience*, vol. 2, no. 3, pp. 194–203, 2001.

[2] U. Rajashekar, I. Van Der Linde, A. C. Bovik, and L. K. Cormack, "Gaffe: A gaze-attentive fixation finding engine," *Image Processing, IEEE Transactions on*, vol. 17, no. 4, pp. 564–573, 2008.

[3] T. Foulsham, "Saliency and eye movements in the perception of natural scenes," Ph.D. dissertation, University of Nottingham, 2008.

[4] M. Iwasaki and H. Inomata, "Relation between superficial capillaries and foveal structures in the human retina." *Investigative ophthalmology & visual science*, vol. 27, no. 12, pp. 1698–1705, 1986.

[5] M. Wright and A. Johnston, "Spatiotemporal contrast sensitivity and visual field locus," *Vision Research*, vol. 23, no. 10, pp. 983 – 989, 1983, neural Interactions in the Vertebrate Retina Proceeding of an {ARVO} Electrophysiology Symposuim.

[6] M. W. Cannon Jr *et al.*, "Perceived contrast in the fovea and periphery," *JOSA A*, vol. 2, no. 10, pp. 1760–1768, 1985.

[7] G. E. Legge and D. Kersten, "Contrast discrimination in peripheral vision," *JOSA A*, vol. 4, no. 8, pp. 1594–1598, 1987.

[8] F. Rieke and M. E. Rudd, "The challenges natural images pose for visual adaptation," *Neuron*, vol. 64, no. 5, pp. 605–616, 2009.

[9] E. C. Larson and D. M. Chandler, "Most apparent distortion: full-reference image quality assessment and the role of strategy," *Journal of Electronic Imaging*, vol. 19, no. 1, pp. 011 006–011 006, 2010.

[10] M. M. Alam, K. P. Vilankar, D. J. Field, and D. M. Chandler, "Local masking in natural images: A database and analysis," *Journal of Vision*, vol. 14, no. 8, 2014. [Online]. Available: http://www.journalofvision.org/content/14/8/22.abstract

[11] J. Intriligator and P. Cavanagh, "The spatial resolution of visual attention," *Cognitive psychology*, vol. 43, no. 3, pp. 171–216, 2001.

[12] M. Dorr and P. J. Bex, "A gaze-contingent display to study contrast sensitivity under natural viewing conditions," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2011, pp. 78 650Y–78 650Y.

[13] L. Itti, "Quantitative modelling of perceptual salience at human eye position," *Visual cognition*, vol. 14, no. 4-8, pp. 959–984, 2006.

[14] W. S. Geisler and J. S. Perry, "Real-time foveated multiresolution system for low-bandwidth video communication," in *Photonics West'98 Electronic Imaging*. International Society for Optics and Photonics, 1998, pp. 294–305.

[15] M. Böhme, M. Dorr, T. Martinetz, and E. Barth, "Gaze-contingent temporal filtering of video," in *Proceedings of the 2006 symposium on Eye tracking research & applications*. ACM, 2006, pp. 109–115.

[16] Y. Rai, M. Barkowsky, and P. Le Callet, "Does h.265 based peri and para-foveal quality flicker disrupt natural viewing patterns?" in *Systems, Signals and Image Processing (IWSSIP), 2015 International Conference on*, Sept 2015, pp. 133–136.

[17] T. Wolff, H.-H. Ho, J. M. Foley, and S. K. Mitra, "Modeling subjectively perceived annoyance of H.264/AVC video as a function of perceived artifact strength," *Signal Processing*, vol. 90, no. 1, pp. 80–92, 2010.

[18] K. Zeng, T. Zhao, A. Rehman, and Z. Wang, "Characterizing perceptual artifacts in compressed video streams," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2014, pp. 90 140Q–90 140Q.

[19] A. T. Duchowski and A. Çöltekin, "Foveated gaze-contingent displays for peripheral lod management, 3d visualization, and stereo imaging," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 3, no. 4, p. 6, 2007.

[20] W. Osberger, N. Bergmann, and A. Maeder, "An automatic image quality assessment technique incorporating higher level perceptual factors," in *Image Processing, 1998. ICIP 98. Proceedings. 1998 International Conference on*. IEEE, 1998, pp. 414–418.

[21] M. Barkowsky, M. Pinson, R. Pépion, and P. Le Callet, "Analysis of freely available dataset for HDTV including coding and transmission distortions," in *Fifth International Workshop on Video Processing and Quality Metrics (VPQM)*, 2010.

[22] D. D. Salvucci and J. H. Goldberg, "Identifying fixations and saccades in eye-tracking protocols," in *Proceedings of the 2000 symposium on Eye tracking research & applications*. ACM, 2000, pp. 71–78.

[23] W. Seiple, K. Holopigian, Y. Shnayder, and J. P. Szlyk, "Duration thresholds for target detection and identification in the peripheral visual field," *Optometry & Vision Science*, vol. 78, no. 3, pp. 169–176, 2001.

[24] B. W. Tatler and T. Troscianko, "A rare glimpse of the eye in motion," *Perception*, vol. 31, no. 11, pp. 1403–1406, 2002.

[25] J. Radun, T. Leisti, T. Virtanen, G. Nyman, and J. Hakkinen, "Why is quality estimation judgment fast? comparison of gaze control strategies in quality and difference estimation tasks," *Journal of Electronic Imaging*, vol. 23, no. 6, p. 061103, 2014. [Online]. Available: http://dx.doi.org/10.1117/1.JEI.23.6.061103

[26] O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior research methods*, vol. 45, no. 1, pp. 251–266, 2013.

[27] H. Jarodzka, K. Holmqvist, and M. Nyström, "A vector-based, multidimensional scanpath similarity measure," in *Proceedings of the 2010 symposium on eye-tracking research & applications*. ACM, 2010, pp. 211–218.

[28] Q. Yang, M. P. Bucci, and Z. Kapoula, "The latency of saccades, vergence, and combined eye movements in children and in adults," *Investigative Ophthalmology & Visual Science*, vol. 43, no. 9, pp. 2939–2949, 2002.

[29] B. Steichen, M. M. Wu, D. Toker, C. Conati, and G. Carenini, "Te, te, hi, hi: Eye gaze sequence analysis for informing user-adaptive information visualizations," in *User Modeling, Adaptation, and Personalization*. Springer, 2014, pp. 183–194.

[30] J. M. West, A. R. Haake, E. P. Rozanski, and K. S. Karn, "eyepatterns: software for identifying patterns and similarities across fixation sequences," in *Proceedings of the 2006 symposium on Eye tracking research & applications*. ACM, 2006, pp. 149–154.