Camera motion influence on dynamic saliency central bias

Etienne Baudrier, Vincent Rosselli, Mohamed-Chaker Larabi

CAMERA MOTION INFLUENCE ON DYNAMIC SALIENCY CENTRAL BIAS

*E. Baudrier**

LSIIT UMR CNRS 7005
University of Strasbourg
FRANCE

V. Rosselli and M. C. Larabi

XLIM-SIC UMR CNRS 6172
University of Poitiers
FRANCE

ABSTRACT

Saliency models have been extensively studied for static images and the focus is now on moving images. There is a central bias in both cases that is emphasized in the dynamic case. One aspect in this latter is the camera motion that influences the scene interpretation. The movie director exploits this motion to make the observer focus on the targeted object which is often in the center of the scene. This aspect is not taken into account in current saliency dynamic models. In this paper, we study the camera motion influence on the gaze distribution in order to include it in a new saliency model. Observers' gazes are recorded with an eye tracker, camera motions (e.g. tracking, zoom...) are calculated thanks to a polynomial projection of the motion field and the motion influence is statistically tested on the recorded gazes.

***Index Terms*—** Visual system, Image motion analysis, Image analysis

1. INTRODUCTION

Saliency map is a useful tool that has taken a great place in different user-centric applications such as marketing, machine learning, computer vision and so on. Its usefulness lies in the fact that an application developer is able to predict the principal areas on which a user will focus with a high probability. The saliency map is considered as a bottom-up, stimuli driven process taking part in the gaze formation jointly with the top-down process. It allows the Human Visual System (HVS) to partially predict the Regions Of Interest (ROI) and the gaze points.

First models have been developed by *Treisman et al.* [1] followed by a still-image computer model presented by *Koch et al.* [2] that has been expanded several years later by *Itti et al.* [3]. Some other works have been devoted to this topic and authors have propose new models or extensions of the existing ones to still images [4, 5, 6] and then to image sequences [7, 8, 9]. In both cases, a central bias (CB) is found: the gaze points are localized around the scene center [10, 9]. Moreover,in [9], the model based only on the CB obtains the best results. This is surprising because salient points can be found all over the image(s). Nevertheless, it is clear that the CB influence should be taken into account in the saliency models and, in this aim, assessed. The point of this paper is to study statistically the CB in the frame of the images sequences.

The CB can be explained by the fact that a third party exists between the scene and the observer This can be the photograph for still images and, the cameraman or the director for moving images. This third party focuses the camera and then guide the observer's gaze to the important areas in the scene and more frequently to the center of the scene. The video introduces another parameter which is the motion that can influence this central concentration of the gaze points. It is indeed well known that the faster a car, the smaller the field of the driver's vision. For the gaze, a moving car corresponds roughly to a zoom.

The main issue in this paper is to evaluate how the spreading of the gaze points around the scene center is correlated to the speed of camera motion in dynamic scene. The remainder of this paper is organized as follows: The second section is dedicated to the description of the experimental protocol including the devices, observers and test setup. The data extraction approach and its exploitation are detailed in the third section. The fourth section is devoted to the description and the analysis of the experimental results and finally this paper ends by some conclusions and future works.

2. EXPERIMENTAL SETUP

In such a work, the experimental setup plays an important role in the accuracy and the consistency of the obtained results. This section is intended to describe the used devices and adopted conditions of our test setup.

The selected test sequences are presented to the observer on a calibrated CRT display. The eye-tracking is performed with an acquisition frequency of 50 Hz. The observations take place in a psychophysical room, constructed in our lab with respect to the ITU recommendations [11]. The observer constitutes an important link in our assessment chain. So his visual performances have to be confirmed by appropriate tests. Thus, the panel has undergone a vision checkup (*Snellen* test

*The author performed the work during the European project EDCine

for visual acuity and *Ishihara* test for color blindness). Ten observers have performed the test; most of them are male. They were asked to make a novice observation, i.e. to watch the test sequences as they do it when watching television or movies and no other task has been assigned. Moreover, the observers have not seen previously the test sequences to avoid memory effects. The test is composed of 16 sequences coming from various sources such as the free movie Elephants Dream, VQEG sequences, and so on. All these sequences have an HD resolution (1920×1080) and are available frame by frame, which enables a high range of possible uses. The sequences are compressed by using MPEG-4 with a high bitrate and this to ensure a visually-lossless quality so as no compression artifact will disturb the observers. The test set has been chosen so as to have natural and artificial sequences, indoor and outdoor scenes, including camera motions such as tracking, zoom-in and/or zoom-out at various speeds.

3. DATA EXPLOITATION

3.1. Eye-tracker data

The data coming from the Eye-tracker are time series of gaze-point coordinates for each observer and each sequence. Points recorded out of the screen are not taken into account. For each point, the polar coordinates (ρ, θ) with the frame center as origin are computed. These coordinates are the most adapted to the central bias study.

3.2. Motion estimation

The definition of the central gaze-spreading has to be made clear and evaluated so as to evaluate the motion influence on it. The interesting motions are the camera ones (tracking, zoom, rotation) so called *principal motions*. The other motions (e.g. object or people motions) influence also the gaze but can not explain the central bias, as they are located everywhere in videos. The method developed by Druon et al. in [12] is used to detect the principal motions. This method is based on the speed vector field projection on a polynomial orthogonal basis. In this method, the camera motion is modeled with the 1-degree polynomial projection coefficients: let

$$\mathcal{F} : \begin{cases} \Omega \subset \mathbb{R}^2 & \rightarrow \mathbb{R}^2 \\ (x, y) & \mapsto (\mathcal{U}(x, y), \mathcal{V}(x, y)) \end{cases}$$

be the motion field (for a given frame) and $\mathcal{B} = (P_{0,0}, P_{0,1}, P_{1,0})$ an orthogonal basis of $\Omega_1[X, Y]$ (the vector space on Ω of polynomials of maximum degree 1). Then the projection π of $\mathcal{F} = (\mathcal{U}, \mathcal{V})$ on $\mathcal{B}$ is

$$\pi(\mathcal{U}) = a_{0,0}P_{0,0} + a_{0,1}P_{0,1} + a_{1,0}P_{1,0} \quad (1)$$

$$\pi(\mathcal{V}) = b_{0,0}P_{0,0} + b_{0,1}P_{0,1} + b_{1,0}P_{1,0} \quad (2)$$

	$a_{0,0}$	$b_{0,0}$	$a_{0,1}$	$b_{0,1}$	$a_{1,0}$	$b_{1,0}$
tracking	τ_{hor}	τ_{vert}	0	0	0	0
zoom	0	0	0	$\sigma/2$	$\sigma/2$	0
rotation	0	0	$-\rho/2$	0	0	$\rho/2$

Table 1. Polynomial projection coefficients for camera-motion identification

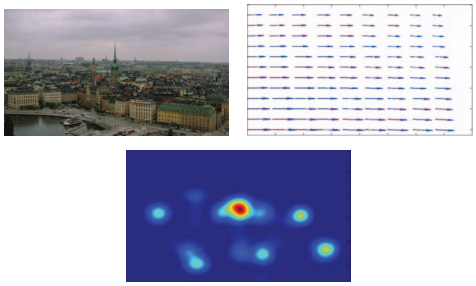

Fig. 1. An example of a frame (left) , the corresponding vector field (middle) and the corresponding map of the observers' gaze points recorded with the Eyetracker , smoothed with a Parzen window (right)

The motion is identified thanks to the projection coefficients as it is detailed in Tab.1 (cf [12]). τ_{hor} (resp τ_{vert}) is the vertical (resp. horizontal) tracking speed, σ is the zoom motion quantity and ρ is the rotation motion quantity.

In our application, the motion fields come from the MPEG standard. Different kinds of motion can be mixed (and some noise too) but in a first study, we focus on pure motions, that is tracking, zoom or rotation. So, for each clip, the pure-motion time intervals are identified and the corresponding gaze points are selected. Fig. 1 shows the different elements of the chain: a frame coming from the clip 10, the corresponding motion field and the recorded gaze points.

The rotation is a scarce camera motion and only one was found in our clips. Likewise, only three zoom motions were detected in our clips. This is not enough to exploit it. Thus we focus on tracking for which examples are numerous (more than 30, 000 frames). More precisely, there are 16 clips, their duration is between 10s and 123s, and the total duration is 14mn44s. The tracking durations for the clips are summarized in Tab. 2. The tracking motion speeds are between 1 and 384 pixels per second, and there are some speeds for which there are too few frames to exploit them (e.g. 336 pix/s).

4. RESULTS

The Eye Tracker has recorded the gaze points of 10 observers on the 16 clips (cf. Sec.2). The tracking motion is detected thanks to the projection method. The results concern the gaze behavior in function of the tracking motion speed. A first test shows a slight dependency on the tracking motion speed norm. The gaze behavior can also depend on the motion speed

Sequence	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Total length (sec)	69	68	92	35	123	99	67	103	82	40	56	10	10	10	10	10
Traveling length (sec)	20	24	39	14	29	23	9	13	22	6	13	0	5	9	0	10

Table 2. Sequence total durations and the corresponding tracking total duration

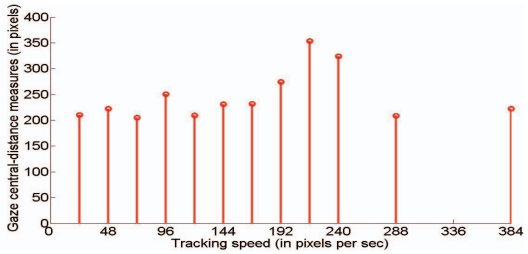

Fig. 2. Mean gaze-point distances to the frame center in function of the tracking speed vector norm $||\tau||$

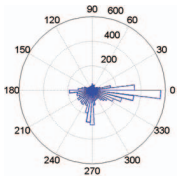

Fig. 3. Rose of the tracking speed direction angles (in degrees)

direction, so a second test is performed to establish a strong dependency of the gaze behavior on it; a last test is performed to assess the dependency of the gaze behavior to the image content.

4.1. Tests on the gaze-point motion tracking speed dependency

For each speed norm value, there are more than 100 recorded gaze points. So only statistical features can give an insight of the speed dependency of the gaze-point distances. In the first test, for each value of tracking speed norm, the corresponding clip frames and the recorded gaze-points are selected. The results of the test are gathered in Fig. 2 that shows the mean distances of the gaze points to the frame center in function of the tracking speed vector norm $||\tau||$. This result shows that the mean distance do not decrease when the tracking motion speed increases.

In the case of tracking, the motion vector has several distinct directions as illustrated Fig. 3. So the influence of tracking direction cannot be highlighted by the first test that makes an average on all the directions. This piece of information can be important for the gaze behavior analysis. As a consequence, a second test is performed. In this second test, the gaze-point central distances are decomposed in their projection on the tracking speed vector and on its orthogonal. Their

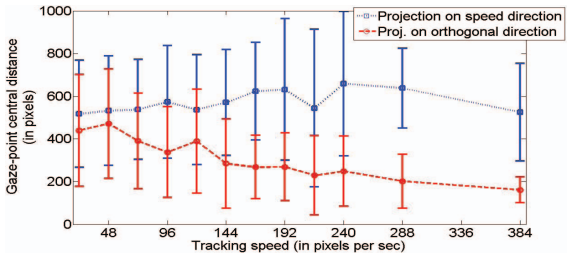

Fig. 4. Average values and standard deviation of the gaze-point (distance to the frame center) projections on the tracking speed axis and on its orthogonal.

averages and standard deviations are represented Fig. 4. The observers do not focus more on the center of the frame in the tracking speed direction (see Fig. 4). On the contrary, the observers tend to bring their gazes closer to a motion central axis (which can be the line including the frame center point and directed by the tracking speed vector). Moreover, the standard deviation decreases, which confirms the obtained result. Thus, it seems to be an interesting piece of information that could be taken into account in a dynamic saliency model.

4.2. Explanations

The different behaviors along and orthogonally to the tracking speed vector could be explained as follows: At low speed, the gaze fixes attention points. At high speed, it is more difficult for the gaze to find attention points, so it seeks points to fix: 1) in the camera speed direction, the focus is difficult because the speed is high, 2) in the speed orthogonal direction, the camera speed is negligible so the gaze try to fix the most probable place for attention points, that is the center. A high camera speed could be due to an object tracking. In this case, the tracked object should be centered in the image and attracts the gaze. Nevertheless, as the speed is high, the centering precision in the speed direction may be lower than in the orthogonal speed-direction. Then the gaze attracted by low speed area may also be less centered in the speed direction than in the orthogonal speed-direction. Thus, this could explain the results showed Fig. 4. In this case, our results would mean that the gaze is attracted by low speed object in a high speed field. To decide if it is the case, we look for the low speed parts in the tracking sequences and compute their average distance from the frame center in the speed and orthogonal speed directions in function of the tracking speed. The results show that the positions of low speed points in the

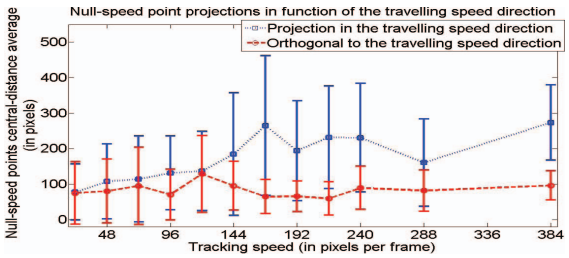

Fig. 5. Average distance of the null-speed points from the frame center in the speed and orthogonal speed directions in function of the tracking speed

image do not decrease neither in the speed orthogonal direction (Fig. 5) nor in the speed direction when speed increases. Thus the gaze behavior to explain is not lead by low-speed parts but is proper to vision in high speed tracking situations.

4.3. Future integration in saliency models

In the case of still images [10], the central bias is modeled by weighting the proposed saliency map S by a two-dimensional Gaussian filter. For the pixel $\mathbf{p} = (x, y)$, the center coordinates $\mu = (\mu_x, \mu_y)$ and the bias variance σ, $S'(\mathbf{p}) = S(\mathbf{p}) \frac{1}{2\pi\sigma^2} \exp \frac{||\mathbf{p}-\mu||^2}{2\sigma^2}$

For the existing dynamic saliency maps, we propose also to integrate the highlighted effect by weighting the current saliency map by a two-dimensional Gaussian filter. As the effect is anisotropic and depends on the tracking speed norm $t = ||\tau||$, the Gaussian filter should include the bias variance-covariance matrix $\Sigma(t)$:

$$S'(\mathbf{p}) = S(\mathbf{p}) \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} \exp\left(\frac{1}{2}(\mathbf{p}-\mu)^T \Sigma(t)^{-1}(\mathbf{p}-\mu)\right)$$

5. CONCLUSION

This paper gives an insight of the camera motions influence on the observer's gaze-point center distance.The gaze points of 10 observers have been recorded with an eye tracker on 16 sequences. The camera motions have been extracted from the sequences thanks to the interpretation of the speed vector field projection on a polynomial basis. This allows us to extract the motion kind and the motion speed. For this experiment, only the tracking motions were sufficiently numerous to be studied. For this kind of motion, the observers' gaze points were represented in function of the tracking speed norm. It shows that motion does not influence the gaze point distribution according to the direction of the tracking speed vector but tends to concentrate the distribution around the center according to the direction orthogonal to the tracking speed vector. An explanation has been presented for this result and a test on the null-speed points has shown that these points cannot explain the latter result which is proper to the gaze in speed tracking sequences. The extension of this study is a second dynamic test which includes numerous zoom and rotation motions so as to evaluate also this kinds of motion. Then it will be interesting to include this gaze behavior in a dynamic saliency map model and to assess its contribution on the final predictions.

6. REFERENCES

[1] A.M. Treisman and G. Gelade, "A feature-integration theory of attention," *Cognitive Psychology*, vol. 12, pp. 97–136, 1980.

[2] C. Koch and S. Ullman, "Schifts in visual selective attention: toward the underlying neural circuitry," *Hum. neurobiol.*, vol. 4, no. 4, pp. 219–227, 1985.

[3] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE T-PAMI*, vol. 20, no. 11, pp. 1254–1259, 1998.

[4] É. Dinet and A. Bartholin, "A spatio-colorimetric model of visual attention," in *proc. of the Expert Symp. on Visual Appearance*, Paris, Oct 2006, CIE, pp. 97–105.

[5] H. Shi and Y. Yang, "A computational model of visual attention based on saliency maps," *Applied Mathematics and Computation*, vol. 188, pp. 16711677, 2007.

[6] A. Torralba, "Modeling global scene factors in attention," *J. Opt. Soc. Am. A*, vol. 20, no. 7, pp. 1407–1418, Jul 2003.

[7] E. Barth, J. Drewes, and T. Martinetz, "Dynamic predictions of tracked gaze," in *Proc. of ISSPA*, 2003.

[8] L. Itti, "Models of bottom-up attention and saliency," in *Neurobiology of Attention*, L. Itti, G. Rees, and J. K. Tsotsos, Eds., pp. 576–582. Elsevier, San Diego, CA, Jan 2005.

[9] O. Le Meur, P. Le Callet, and D. Barba, "Predicting visual fixations on video based on low-level visual features," *Vis. res.*, vol. 47, pp. 2483–2498, 2007.

[10] D. Parkhurst, K. Law, and E. Niebur, "Modeling the role of salience in the allocation of overt visual attention.," *Vis. res.*, vol. 42, no. 1, pp. 107–123, Jan 2002.

[11] ITU-R, "Methodology for the subjective assessment of the quality of television pictures," Tech. Rep. BT.500-10, ITU, Geneva, Switzerland, 2002.

[12] M. Druon, B. Tremblais, and B. Augereau, "Vector fields modelization using basis of polynomials: application to the analysis of simple face movements.," in *ICASSP*. IEEE, 2006, vol. 2, pp. 661–664.