Study of Center-Bias in the Viewing of Stereoscopic Image and a Framework for Extending 2D Visual Attention Models to 3D

Junle Wang, Matthieu Perreira da Silva, Patrick Le Callet, Vincent Ricordel

Study of Center-Bias in the Viewing of Stereoscopic Image and a Framework for Extending 2D Visual Attention Models to 3D

Junle Wang, Matthieu Perreira Da Silva, Patrick Le Callet, Vincent Ricordel

LUNAM Université, Université de Nantes, IRCCyN UMR CNRS 6597,
Polytech Nantes, Rue Christian Pauc BP 50609 44306 Nantes Cedex 3

ABSTRACT

Compared to the good performance that can be achieved by many 2D visual attention models, predicting salient regions of a 3D scene is still challenging. An efficient way to achieve this can be to exploit existing models designed for 2D content. However, the visual conflicts caused by binocular disparity and changes of viewing behavior in 3D viewing need to be dealt with. To cope with these, the present paper proposes a simple framework for extending 2D attention models for 3D images, as well as evaluates center-bias in 3D-viewing condition. To validate the results, a database is created, which contains eye-movements of 35 subjects recorded during free viewing of eighteen 3D images and their corresponding 2D version. Fixation density maps indicate a weaker center-bias in the viewing of 3D images. Moreover, objective metric results demonstrate the efficiency of the proposed model and a large added value of center-bias when it is taken into account in computational modeling of 3D visual attention.

Keywords: Visual attention, stereoscopy, 3DTV, center bias, eye-tracking, gaze pattern

1. INTRODUCTION

The human visual system (HVS), receives a considerably large amount of information beyond its capability to process all of it. To cope with large amounts of information, visual attention is one of the most important mechanisms deployed in the HVS to reduce the complexity of scene analysis.[1] Inspired by the HVS, numerous computational models of visual attention have been investigated during the last decades to predict salient areas .

Nowadays, stereoscopic 3D content increases the sensation of presence through the enhancement of depth perception. For simplicity of notation, from now on, we will use the term 3D to refer to stereoscopic 3D in the remainder of this article. Compared to the body of studies with regards to visual attention in 2D viewing condition, a relatively small number of studies and computational models of 3D visual attention can be found in the literature. Nevertheless, studies related to 3D visual attention have been recently gaining an increasing amount of attention because of the emergence of 3D content (in cinemas and at home) and the recent availability of high-definition 3D-capable acquisition and display equipment. Several new candidate applications of 3D visual attention models are also emerging,[2] such as: 3D video capture, 2D to 3D conversion, reframing and depth adaptation, subtitling 3D movie.

The enhancement of depth perception is one of the most significant changes that viewers can feel during watching 3D image/video. However, the additional depth information makes predicting salient areas in a 3D scene become a challenging task. Psychophysical studies[3,4] have shown that watching the 3D version of a scene can make the viewer's attention distribute differently as compared to watching the 2D version of the same scene. Fortunately, it has been demonstrated that 3D visual attention is still guided by many 2D visual features.[5] This consistence of the influence of 2D low-level features implies the possibility of adapting existing 2D models to 3D cases. This is also the reason why most of the existing computational models of 3D visual attention share a same step in which salient regions are first detected based on 2D visual features.[6–9]

In addition to the salient regions that result from 2D visual features, fixation patterns from eye-tracking experiments have also demonstrated a bias towards the screen center. This phenomenon is named as "center-bias" (or "central fixation bias"). The causes of this center-bias effect include the photographer bias, the viewing

strategy, the orbital reserve, the motor-bias, and the center of screen bias.[10] Studies[11–13] have indicated that the prediction of salient region can be largely improved by integrating the center-bias effect.

However, center-bias is not taken into account in most of the existing 3D visual attention models. There still exist several difficulties of applying center-bias in 3D visual attention models:

(1) The influence of center-bias in 3D viewing has not been confirmed. Several studies[3,4,14] draw inconsistent conclusions about how the spatial extent of exploration varies from 2D viewing to 3D viewing. This variation of extent implies the different degrees of center-bias in the two viewing conditions.

(2) The ways of integrating center-bias may not be consistent. Not all 3D visual attention models can integrate center-bias in the same way: in models[9] taking both views as input, center-bias can be added on both views; in models[7] taking one image and a depth map as input, center-bias has to be added as a post-processing step after the output of saliency map.

(3) Ground-truth data is still lacking. So far, there are few databases providing eye-tracking results for both the 2D and 3D versions of the same set of images. The lack of ground truth data limits the study of center-bias in 3D condition.

In this paper, we propose a simple framework of 3D visual attention which can easily take advantage of center-bias and existing 2D models. The degree of center-bias during 3D natural content images viewing is also quantitatively evaluated. Our results indicate a clear difference between the center-bias in 2D and 3D viewing conditions. By applying a proper degree of center-bias in the proposed framework, a significant added value is demonstrated in the prediction of saliency maps for 3D images. A new database containing (both the 2D and 3D versions of) natural content images and their corresponding binocular eye-tracking data is also introduced in this paper.

2. A FRAMEWORK OF 3D VISUAL ATTENTION

The proposed framework is inspired by the attentional framework for stereo vision proposed by Bruce and Tsotsos,[9] which is selected on the basis of its biological plausibility. Due to the complexity of Bruce's framework, a simplification was made in our study by keeping only layer 1 (to detect salient areas using 2D visual features) and layer 2 (to shift attention according to various binocular disparities).

In the proposed framework, the left-view image and the right-view image are taken separately as inputs. Firstly, a 2D visual attention model is applied independently on the two images, and creates a corresponding 2D saliency map for each view. Secondly, the left and right 2D saliency maps go through an attention shifting step in which two saliency maps are merged according to the local disparity information. It is worth noting that, in this framework, center-bias can be either added in both paths to weight the two 2D saliency maps before the attention shifting step (Figure 1.a); or be added after the attention shifting step to weight the fused saliency map (Figure 1.b). The details of these step are introduced as follows.

2.1 2D saliency computation

Since developing a completely new computational model of 2D visual attention is not in the scope of this paper, we leave the work of 2D visual features detection and 2D saliency map creation to existing models. In this study, three state-of-the-art models using different mechanisms have been tried:

(1) Bruce's AIM model[15] which is based on information maximization.

(2) Itti's model[16] which is the most widely used one in the literature. This model is based on three low-level features, including intensity, color and orientation.

(3) Hou's model[17] based on the computation of spectral residual.

To evaluate the performance of the proposed 3D visual attention framework, each of these three models is applied to perform 2D saliency prediction.

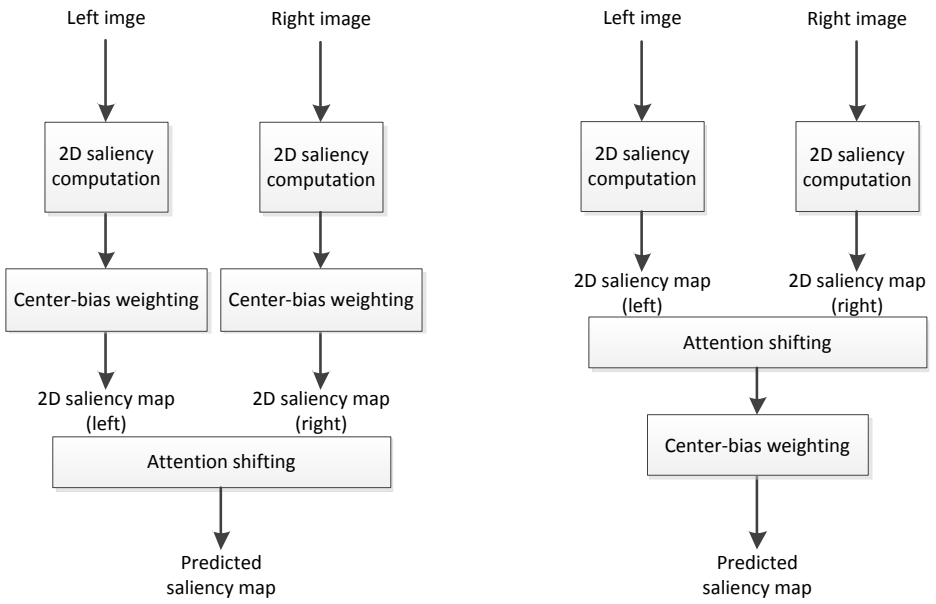

Figure 1: Overview diagrams of the proposed model with two ways of integrating center-bias. (a) Two independent center-bias models are added in the left and right paths to weight the two 2D saliency maps. (b) One center-bias model is added after the two 2D saliency maps have been merged.

2.2 Center-bias modeling

So far, there is still not a strong agreement on the ways of modeling the center-bias. In the literature, center-bias was usually modeled by using either an isotropic Gaussian kernel[11] or an anisotropic Gaussian kernel.[18] Tseng *et al.*[10] demonstrated that the image borders have a large impact on center-bias. It implies that the shape (i.e. the length to width ratio) of the images should be also taken into account when designing the Gaussian kernel. Consequently, in our study, we use an anisotropic Gaussian kernel[18]to model the center-bias. This kernel is used for weighting the saliency map. The weighted saliency map, named S', is then given by:

$$S'(x,y) = S(x,y)exp\left(-\frac{(x-x_0)^2}{2\sigma_x^2} - \frac{(y-y_0)^2}{2\sigma_y^2}\right)$$

where (x_0,y_0) represent the image's center coordinates. σ_x and σ_y denote the standard deviation related to the x-axis and the y-axis, respectively. The relationship between σ_x and σ_y is quantified according to the size of image viewed:

$$\sigma_y = \sigma_x \times \left(\frac{R_x}{R_y} Ind(R_x < R_y) + \frac{R_y}{R_x} Ind(R_x > R_y)\right)$$

where R_x and R_y are the image's width and height, and $Ind()$ is the indicatric function. Note that the standard deviations σ_x and σ_y, representing the degree of center-bias are measured in visual degree, since the measurement of visual degree takes into account the viewing distance.

2.3 Attention shifting

Due to the disparity between the left view and the right view, an area in a scene can thus correspond to two slightly different locations in the retinal images of two eyes. Moreover, since conflicts may exist between the two eyes due to occlusions in binocular viewing, the saliency maps of the left view and the right view may not necessarily be the same at all the locations. Consequently, the two saliency maps that come from the two eyes need to be merged by shifting each pixel's saliency value from one view to the other.

Figure 2: The eighteen images used in eye-tracking experiment.

The distance of shifting is processed according to the local disparity between the two views. Due to the symmetry of binocular disparity, a saliency map from either of the two views can be shifted to fit the other one. We thus arbitrarily shift the saliency map of the right view, and then combine it with the saliency map of left view. The resulting saliency map S" is obtained by Equation 1:

$$S"(i,j) = S_L(i,j) + S_R(i + D_x(i,j), j + D_y(i,j)) \quad (1)$$

where (i,j) represents the coordinate of each pixel in the image; S_L denotes the left-view saliency map; S_R denotes the right-view saliency map; D_x and D_y denote the horizontal and vertical disparity at each pixel.

3. EXPERIMENT

In this section, we introduce an eye-tracking experiment which created a database providing the fixation density maps of both the 2D and 3D versions of a set of natural content images.

3.1 Stimuli

Eighteen stereoscopic pairs of images[6,19] were collected (see Figure 2). Ten of the images and their disparity maps come from the Middlebury 2005/2006 image set.[20] They have a resolution about 1300*1100 pixels. The other eight pictures were taken in the campus of the University of Nantes using a Panasonic AG-3DA1 twin-lens 3D camera. These images have a resolution of 1920*1080 pixels. Their disparity maps were generated by a depth estimation algorithm using optical flow.[21,22] Both the 3D version (containing a left view image and a right view image) and the 2D version (containing two copies of the left view image) of these images were used in this binocular eye-tracking experiment.

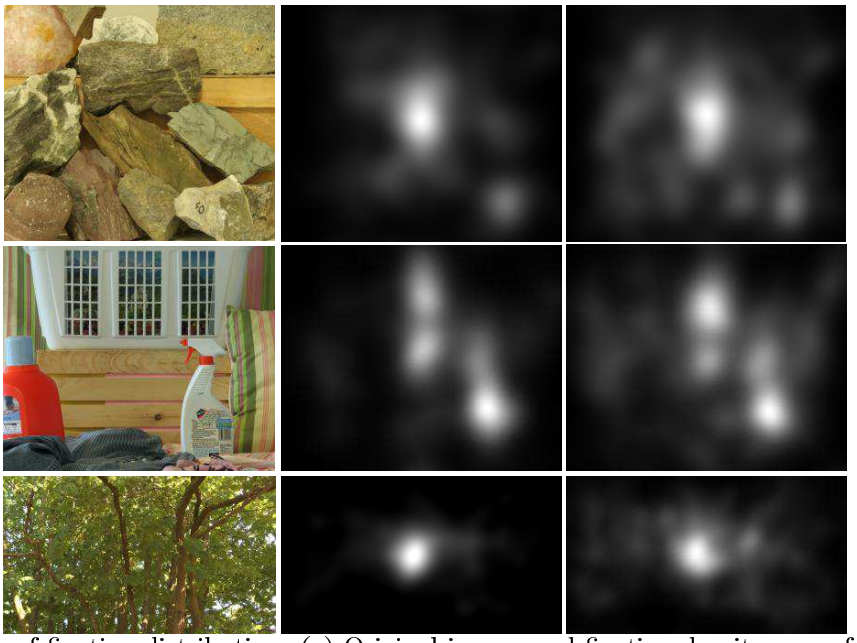

Figure 3: Examples of fixation distribution: (a) Original image; and fixation density maps from the viewing in (b) 2D condition, (c) 3D condition.

3.2 Apparatus and procedures

Stimuli were displayed on a 26-inch Panasonic BT-3DL2550 3D LCD screen, which has a resolution of 1920 * 1200 pixels and a refresh rate of 60 Hz. The maximum luminance of the display was 180 cd/m2, which yielded a maximum luminance of about 60 cd/m2 when watched through glasses. A SMI RED 500Hz remote eye-tracker was used to record the eye movements.

Subjects watched the screen at a viewing distance of 97 cm through a pair of passive polarized glasses. The screen subtended 33 degrees * 19 degrees of visual angle at this viewing distance. Subjects were required to do a free-viewing task. Each image was presented for 15 seconds. Between every two scenes, a center point was showed for 500 ms at the screen center with zero disparity. A nine-point calibration was performed at the beginning of the experiment, and repeated every five scenes. To deal with the problem of visual fatigue, each subject was required to have at least three rests during the whole observation.

3.3 Participants

Thirty-five subjects participated in the experiment, including 25 right-eye dominant subjects and 10 left-eye dominant subjects. The mean age of subjects was 24.2 years old. All subjects had either normal or corrected-to-normal visual acuity (checked by the Monoyer chart), color vision (checked by the Ishihara test) and 3D acuity (checked by the Randot stereo test).

3.4 Post-processing of eye-tracking data

Two fixation maps were firstly created separately for both eyes. Note that the fixations from the right eye were shifted for the same reason introduced in section 2.3. Secondly, the fixation maps were filtered using a two-dimensional Gaussian kernel to account for the decrease in visual accuracy with increasing eccentricity from the fovea. The standard deviation of the Gaussian kernel used for creating of fixation density maps was equal to 2 degrees of visual angle.

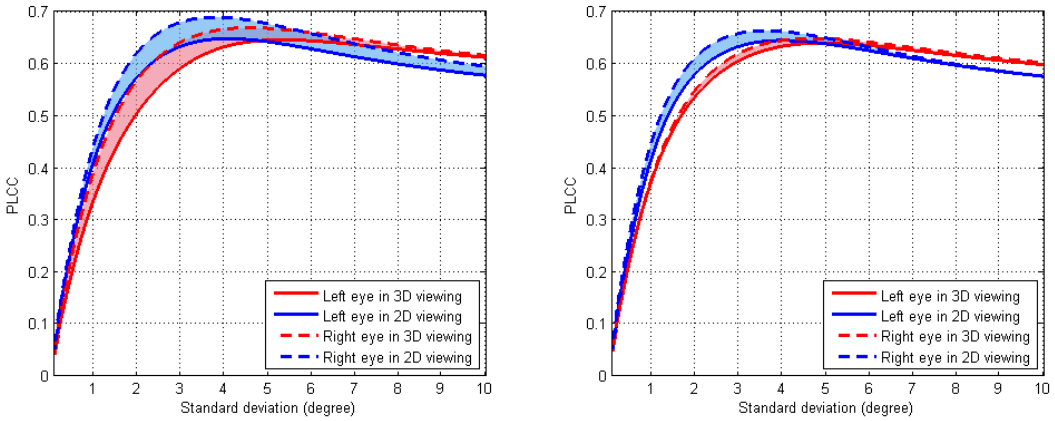

Figure 4: The left (resp. right) figure is obtained from subjects who are left-eye (resp. right-eye) dominant. Solid lines represent the degree of center-bias of right eye's data. Dash lines represent the degree of center-bias of left eye's data. Blue color and red color are, respectively, assigned to the lines and the areas between them in order to distinguish the 2D and 3D viewing conditions.

4. RESULTS AND ANALYSIS

4.1 Center-bias in 2D viewing and 3D viewing

Figure 3 shows some examples of the fixation density maps generated from our eye-tracking experiment. They were obtained during the viewing of the 2D version and the 3D version of the same set of images. From these fixation density maps, clear difference of fixation distribution can be observed. The fixations are more widely distributed in the 3D images than in the 2D images.

To quantitatively examine the degree of center-bias in 2D viewing and 3D viewing, we apply a similar method as the one used in.[18] A set of center-bias maps are firstly created by using only the center-bias model introduced in section 2.2 with a standard deviation σ_x ranging from 0 degree to 10 degree. Each of these center-bias maps is compared with the left and right fixation density maps of each image (both 2D and 3D version) by the Pearson Linear Correlation Coefficient (PLCC). These PLCC values are then averaged over observers. The averaged PLCC evolution is plotted in Figure 4.

In Figure 4, the value of PLCC represents the similarity between the real fixation distribution and a 2D Gaussian distribution; the value of standard deviation represents how much the Gaussian distribution is concentrated to the center. Therefore, a smaller standard deviation and a higher PLCC value mean that the distribution of fixations is more concentrated at the center, and thus correspond to a higher degree of center-bias.

From the two colors in Figure 4, one can clearly observe a higher degree of center-bias (of both eyes) in 2D viewing than in 3D viewing. This finding holds for (1) both the left eye and the right eye, and (2) for both the left-eye dominant viewers and the right-eye dominant viewers. This result is consistent with the result from Hakkinen *et al.*[3]and Jansen *et al.*[5] It is worth to note that, regarding the development of computational models, a lower degree of center-bias in 3D viewing means that the Gaussian kernel applied in the 3D visual attention model should have a larger standard deviation.

The observers involved in our experiment have different dominant eyes. Nevertheless, we find out that the relative difference between 2D viewing and 3D viewing for these two groups of observers are similar. For simplicity, we merge the data from these two groups of observers in the following analysis.

4.2 Integration of the center-bias

Curves in Figure 4 show differences of center-bias between left eye and right eye. These differences imply the plausibility of modeling the center-bias by two different Gaussian kernels for the saliency maps from both views. Several objective metrics have been applied to assess the performance of such a strategy. However, the results shows only marginal effects on the accuracy of the final maps by applying Gaussian kernels with different sizes,

as compared to either (1) applying two identical Gaussian kernels, or (2) applying a Gaussian kernel after the fusion of two views.

The reason of this similarity among different ways of applying center-bias could be due to the absence of occlusion areas and areas with extreme disparities in the images of our database. In the following quantitative analysis, only the way of adding center-bias after the fusion of two views (as shown in Figure 1.b) is used for the performance assessment.

4.3 Performance of the proposed framework and added value of center-bias in 3D visual attention models

Table 1 gives the performance of the proposed 3D framework. Each of the three 2D models is combined with various levels of center-bias to predict the saliency maps of 3D images. Three degrees of center-bias are tested:

- Zero center-bias. In this case, no center-bias is considered. The saliency map is uniformly weighted.
- The 2D optimal value $\sigma_{2D} = 4.1$ degrees. This value of standard deviation results from a training based on the fixation patterns obtained in 2D viewing condition. This smaller value corresponds to a more concentrated distribution of fixations. Equally, it means a higher degree of center-bias.
- The 3D optimal value $\sigma_{3D} = 4.9$ degrees. This value of standard deviation results from a training based on the fixation patterns obtained in 3D viewing condition. This larger value indicates a wider spread distribution of fixations and a lower degree of center-bias.

Note that the performance of each 2D model in predicting saliency maps of 2D version images is also presented in Table 1 as a reference.

Three widely used objective similarity metrics are used in the performance assessment: (1) Pearson linear correlation coefficients (PLCC); (2) Kullback-Leibler divergence (KLD); (3) Area under the ROC curves. The results demonstrate that the performance of the proposed 3D visual attention model largely depends on the 2D saliency model adopted no matter whether center-bias is added. When no center-bias is taken into account, the proposed model generally has a comparable performance as the performance of the corresponding 2D model on 2D images. When the center-bias is considered, all these three metrics indicate great improvements of the performance of all the three 2D saliency models. Moreover, the results show the impact of the parameter σ on the contribution of the center-bias. The proposed model has a better performance when using the parameter σ particularly tuned for 3D condition. Those parameters previously used for 2D condition do not perform optimally for 3D content.

5. CONCLUSION AND PERSPECTIVE

Psychophysical studies about center-bias and a 3D visual attention framework are presented in this paper. The proposed framework can exploit existing 2D attention models, and has a good performance in predicting saliency maps of 3D still images. Our work also demonstrates that center-bias in 3D viewing condition is slightly weaker than in 2D viewing condition. Nevertheless, integrating the proper degree of center-bias can still make large added value to the proposed framework.

Moreover, if we take into account the depth value of each fixation, we find that the fixations are not only biased towards screen center but also towards to the objects closer to the observers. This phenomenon implies another bias, which was given the name of "depth-bias" in.[23] Our future works will focus on combining both the depth bias and the center bias for developing a 3D attention model.

6. ACKNOWLEDGMENTS

This work is supported by the French ANR-PERSEE project (project reference: no ANR-09-BLAN-0170).

2D saliency model	Image	Degree of CB	PLCC	KLD	AUC
Bruce's model	2D	No center-bias	0.2853	0.8135	0.6423
	3D	No center-bias	0.3423	0.5159	0.6397
		σ_{2D}	0.6717	0.4675	0.7358
		σ_{3D}	0.6913	0.3532	0.7377
Itti's model	2D	No center-bias	0.1370	2.8072	0.5480
	3D	No center-bias	0.1568	2.3740	0.5483
		σ_{2D}	0.2147	2.4450	0.5514
		σ_{3D}	0.2165	2.3438	0.5516
Hou's model	2D	No center-bias	0.2628	0.8576	0.6386
	3D	No center-bias	0.3003	0.5805	0.6273
		σ_{2D}	0.6120	0.5738	0.7326
		σ_{3D}	0.6232	0.4543	0.7323

Table 1: Performance of the proposed model on 3D images with different 2D saliency models and different degrees of center-bias (noted as CB in the table). $\sigma_{2D} = 4.1$ degrees and $\sigma_{3D} = 4.9$ degrees. The performance of these 2D attention models on 2D images is also presented. Note that a smaller KLD score means a better performance.

REFERENCES

[1] Wolfe, J., "Visual attention," *Seeing* **2**, 335–386 (2000).

[2] Huynh-Thu, Q., Barkowsky, M., Le Callet, P., et al., "The importance of visual attention in improving the 3d-tv viewing experience: Overview and new perspectives," *IEEE Transactions on Broadcasting* **57**(2), 421–431 (2011).

[3] Hakkinen, J., Kawai, T., Takatalo, J., Mitsuya, R., and Nyman, G., "What do people look at when they watch stereoscopic movies?," **7524**, 75240E, SPIE (2010).

[4] Ramasamy, C., House, D., Duchowski, A., and Daugherty, B., "Using eye tracking to analyze stereoscopic filmmaking," in [*SIGGRAPH'09: Posters*], 28, ACM (2009).

[5] Jansen, L., Onat, S., and König, P., "Influence of disparity on fixation and saccades in free viewing of natural scenes," *Journal of Vision* **9**(1) (2009).

[6] Wang, J., Perreira Da Silva, M., Le Callet, P., and Ricordel, V., "A computational model of stereoscopic 3d visual saliency," *IEEE Transactions on Image Processing* (to appear).

[7] Zhang, Y., Jiang, G., Yu, M., and Chen, K., "Stereoscopic visual attention model for 3d video," *Advances in Multimedia Modeling* , 314–324 (2010).

[8] Chamaret, C., Godeffroy, S., Lopez, P., and Le Meur, O., "Adaptive 3d rendering based on region-of-interest," in [*Proceedings of SPIE*], **7524**, 75240V (2010).

[9] Bruce, N. and Tsotsos, J., "An attentional framework for stereo vision," in [*Computer and Robot Vision, 2005. Proceedings. The 2nd Canadian Conference on*], 88–95, IEEE (2005).

[10] Tseng, P., Carmi, R., Cameron, I., Munoz, D., and Itti, L., "Quantifying center bias of observers in free viewing of dynamic natural scenes," *Journal of Vision* **9**(7) (2009).

[11] Zhao, Q. and Koch, C., "Learning a saliency map using fixated locations in natural scenes," *Journal of Vision* **11**(3) (2011).

[12] Ma, Y., Lu, L., Zhang, H., and Li, M., "A user attention model for video summarization," in [*Proceedings of the tenth ACM international conference on Multimedia*], 533–542, ACM (2002).

[13] Luo, Y., Yuan, J., Xue, P., and Tian, Q., "Saliency density maximization for object detection and localization," *Computer Vision–ACCV 2010* , 396–408 (2011).

[14] Wang, J., Le Callet, P., Ricordel, V., and Tourancheau, S., "Quantifying depth bias in free viewing of still stereoscopic synthetic stimuli," *16th European Conference on Eye Movements, Marseille, France* (2011).

[15] Bruce, N. and Tsotsos, J., "Saliency, attention, and visual search: An information theoretic approach," *Journal of Vision* **9**(3) (2009).

[16] Itti, L., Koch, C., and Niebur, E., "A model of saliency-based visual attention for rapid scene analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **20**(11), 1254–1259 (1998).

[17] Hou, X. and Zhang, L., "Saliency detection: A spectral residual approach," in [*Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*], 1–8, Ieee (2007).

[18] Le Meur, O., Le Callet, P., Barba, D., and Thoreau, D., "A coherent computational approach to model bottom-up visual attention," *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **28**(5), 802–817 (2006).

[19] Wang, J., Perreira Da Silva, M., Le Callet, P., and Ricordel, V., "IRCCyN/IVC 3DGaze database." http://www.irccyn.ec-nantes.fr/spip.php?article1102&lang=en (2011).

[20] Scharstein, D. and Pal, C., "Learning conditional random fields for stereo," in [*Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*], 1–8, IEEE (2007).

[21] Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., and Bischof, H., "Anisotropic huber-l1 optical flow," in [*Proceedings of the British machine vision conference*], (2009).

[22] Werlberger, M., Pock, T., and Bischof, H., "Motion estimation with non-local total variation regularization," in [*Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*], 2464–2471, IEEE (2010).

[23] Wang, J., Le Callet, P., Ricordel, V., Tourancheau, S., and Perreira Da Silva, M., "Study of depth bias of observers in free viewing of still stereoscopic synthetic stimuli," *Journal of Eye Movement Research* **5(5):1**, 1–11 (2012).