EYEC3D: 3D VIDEO EYE TRACKING DATASET

Philippe Hanhart and Touradj Ebrahimi

Multimedia Signal Processing Group, EPFL, Lausanne, Switzerland

ABSTRACT

Despite efforts of the scientific community in recent years, little is known about the mechanisms of the human visual system that control visual attention when watching 3D content. To help understanding these mechanisms and develop more accurate visual attention models, we created a public 3D video eye tracking dataset. The dataset provides the eye tracking information corresponding to eight stereoscopic video sequences. The eye tracking information includes the fixation points and fixation density maps measured during subjective experiments. This paper describes the dataset in details, including the stereoscopic video sequences, the eye tracking experiments, and the computation of the fixation density maps.

Index Terms— Eye tracking, human fixation, visual attention, 3DTV, stereoscopy, dataset

1. INTRODUCTION

Understanding visual attention in 3DTV is essential for many applications, e.g., capture, coding, visual comfort enhancement, 2D-to-3D conversion, retargeting, and subtitling [1]. Visual attention models can be exploited to design more efficient compression algorithms, better objective quality metrics, and more accurate computational models of visual saliency. Therefore, public datasets of 3D content with associated ground truth eye tracking data are needed. However, to the best of our knowledge, there are very few publicly available 3D eye tracking datasets: two for stereoscopic images [2, 3] and none for stereoscopic video sequences.

To overcome the lack of publicly available 3D video eye tracking datasets, we created the EyeC3D dataset[1]. Eight stereoscopic video sequences were used in the eye tracking experiments. The sequences were selected based on the availability of stereoscopic video content with associated depth maps. For each video, eye movement data was recorded via a set of subjective experiments using 21 naïve subjects and a professional eye tracking device. From the eye movement data, the fixation density maps (FDMs) were computed for each frame of the stereoscopic video sequences. Figure 1 illustrates one example of resulting FDM.

In summary, the dataset provides the following:

1. Lists of fixation points;
2. Fixation density maps.

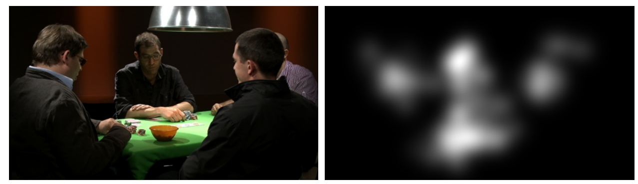

Fig. 1. Example of FDM from our dataset.

Table 1. Stereoscopic video sequences.

Sequence	Frames	Length [s]	Views (L-R)	$d_{\min}$ [pixels]	$d_{\max}$ [pixels]
Boxers	$1-250$	10	$0-1$	-14	29
Hall	$1-250$	10	$0-1$	-15	20
Lab	$151-400$	10	$0-1$	-100	44
News report	$1-250$	10	$0-1$	-45	71
Phone call	$151-400$	10	$0-1$	-35	39
Musicians	$1-250$	10	$0-1$	0	176
Poker	$1-250$	10	$0-1$	0	176
Poznan Hall2	$1-200$	8	$7-6$	16	118

2. STEREOSCOPIC VIDEO SEQUENCES

Eight stereoscopic video sequences were used in the eye tracking experiments (see Table 1). Sequences *Boxers*, *Hall*, *Lab*, *News report*, and *Phone call* were obtained from the NAMA3DS1 database[2] [4]. Sequences *Musicians* and *Poker* were obtained from the European FP7 Research Project MUSCADE[3] [5]. Sequence *Poznan Hall2* was obtained from the Poznań multiview video database [6].

Since the cameras used for recording sequences *Musicians*, *Poker*, and *Poznan Hall2* were set in a parallel configuration, they are assumed to converge at an infinite point. This setup leads to stereoscopic window violation and does not sufficiently exploits the depth range, as the 3D content appears only in front of the screen plane. Therefore, for these

This work has been conducted in the framework of the Swiss National Foundation for Scientific Research (FN 200021-143696-1), EC funded Network of Excellence VideoSense, and COST IC1003 European Network on Quality of Experience in Multimedia Systems and Services QUALINET.

[1]The dataset can be downloaded here: http://mmspg.epfl.ch/eyec3d

[2]http://www.irccyn.ec-nantes.fr/spip.php?article954

[3]http://www.muscade.eu

sequences, horizontal image translation was applied with a shift to converge the scene defined as [3]

$$s = \frac{d_{\min} - d_{\max}}{2}$$

where $d_{\min}$ and $d_{\max}$ are the minimum and maximum disparity values computed from the camera parameters (see Table 1), respectively. Note that no shift was applied to the sequences from the NAMA3DS1 database, as the cameras were already converged during recording.

3. EYE TRACKING EXPERIMENTS

The eye tracking experiments were conducted at the MMSPG test laboratory, which fulfills the recommendations for subjective evaluation of visual data issued by ITU-R [7]. The viewing conditions were set according to Rec. ITU-R BT.2021 [7]. Table 2 presents the detailed summary of the experiments.

As few fixation points can be recorded during each frame, the video sequences were presented twice to each subject, in a nonconsecutive order. To reduce contextual effects, the stimuli orders of display were randomized applying different permutation for each subject.

The left and right eye movements were recorded independently on a separate computer. Synchronization between eye movement data and video frames was ensured using a specific video player that records the current timestamp of the eye tracking system, which was sent by UDP on a local network, when displaying each frame of the video.

Table 2. Overview of the eye tracking experiments.

Category	Details	Specification
Participants	Number (♂/♀)	21 (16/5)
	Age range (average age)	$18-31$ (21.8)
	Screening	Snellen chart, Ishihara chart, and Randot test
Viewing conditions	Environment	Laboratory
	Illumination	Low
	Color temperature	6500 [K]
	Viewing distance	1.8 [m]
	Task	Free-viewing
Display	Manufacturer	Hyundai
	Model	S465D
	Type	Polarized LCD
	Size	46 [inch]
	Resolution	1920×1080 [pixel]
	Angular resolution	60 [pixel/degree]
Display calibration	Probe	X-Rite i1Display Pro
	Profile	D65 white point, 120 [cd/m^2] brightness, minimum black level
Eye tracker	Manufacturer	Smart Eye
	Model	Smart Eye Pro 5.8
	Mounting position	1.28 [m] from the display
	Sampling frequency	60 [Hz]
	Accuracy	< 0.5 [degree]
	Calibration points	4 points on screen
Video presentation	Presentation order	Random
	Presentation time	$8-10$ [s]
	Repetitions	2
	Grey-screen duration	2 [s]

4. FIXATION DENSITY MAPS

The eye tracking system used in our experiments (see Table 2 for details) automatically discriminates between saccades and fixations, based on the gaze velocity information. Blinks are also detected automatically based on the distance between the two eyelids of each eye. All detected saccades and blinks were excluded from the eye movement data and only the gaze points classified as fixation points were used.

For each frame of the video sequence, the corresponding fixation points were processed as follows. First, the right-eye fixation points were shifted horizontally according to the right-to-left disparity map [3]. Then, these points were combined with the left-eye fixation points and filtered with a Gaussian kernel to account for the eye tracking inaccuracies and the reduction of the visual sensitivity, which depends on the distance from the fovea. The standard deviation of the Gaussian filter used for computing the FDMs was set to 1 degree of visual angle, which corresponds to $\sigma = 60$ pixels in our experiments. This is based on the assumption that the fovea of the human eye covers approximately 2 degrees of visual angle [3]. Therefore, for each frame of the stereoscopic video sequence, only one FDM, corresponding to the left view, was produced from the left and right eye movements.

5. CONCLUSION

This paper describes the EyeC3D public dataset, which consists of the fixation points and fixation density maps, obtained via eye tracking experiments, corresponding to eight stereoscopic video sequences.

6. REFERENCES

[1] Q. Huynh-Thu, M. Barkowsky, and P. Le Callet, "The Importance of Visual Attention in Improving the 3D-TV Viewing Experience: Overview and New Perspectives," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 421–431, June 2011.

[2] C. Lang, T. V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli, and S. Yan, "Depth Matters: Influence of Depth Cues on Visual Saliency," in *Computer Vision*, Lecture Notes in Computer Science, pp. 101–115. Springer Berlin Heidelberg, 2012.

[3] J. Wang, M.P. Da Silva, P. Le Callet, and V. Ricordel, "Computational Model of Stereoscopic 3D Visual Saliency," *IEEE Transactions on Image Processing*, vol. 22, no. 6, pp. 2151–2165, June 2013.

[4] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. Le Callet, J. Gutierrez, and N. Garcia, "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," in *Fourth International Workshop on Quality of Multimedia Experience*, July 2012, pp. 109–114.

[5] ISO/IEC JTC1/SC29/WG11, "Proposed Stereo Test Sequences for 3D Video Coding," Doc. M23703, Feb. 2012.

[6] ISO/IEC JTC1/SC29/WG11, "Poznań Multiview Video Test Sequences and Camera Parameters," Doc. M17050, Oct. 2009.

[7] ITU-R BT.2021, "Subjective methods for the assessment of stereoscopic 3DTV systems," ITU, Aug. 2012.