AN EYE TRACKING DATABASE FOR STEREOSCOPIC VIDEO

Yuming Fang, Junle Wang, Jing Li, Romuald Pépion, Patrick Le Callet

HAL Id: hal-01095107

https://hal.science/hal-01095107

Submitted on 15 Dec 2014

AN EYE TRACKING DATABASE FOR STEREOSCOPIC VIDEO

Yuming Fang[1], Junle Wang[2], Jing Li[3], Romuald Pépion[3] and Patrick Le Callet[3]

[1]School of Information Technology, Jiangxi University of Finance and Economics, Nanchang, China
[2]Ars Nova Systems, Nantes, France
[3]LUNAM Université, Université de Nantes, IRCCyN UMR CNRS, Polytech Nantes, Nantes, France

ABSTRACT

We present a large-scale eye tracking database for stereoscopic video. A set of participants were involved in this eye tracking experiment. The human fixation maps were created as the ground truth for stereoscopic video from the gaze data from participants. To the best of our knowledge, this is the first large-scale eye tracking database of visual attention modeling for stereoscopic video. The details of the processing operations and properties of the database are described in this paper.

Index Terms— stereoscopic video, visual attention, eye tracking database, human fixation.

1. INTRODUCTION

Visual attention is an important characteristic in the Human Visual System (HVS) for visual information processing. It is a cognitive process to selectively process the significant visual information while ignoring others in visual scenes. Generally, the visual information captured by human eyes is much more than that the HVS can process. During viewing the visual scene, the HVS cannot recognize the components and their relationship in the visual scene immediately. Selective attention would filter out redundant visual information to focus on the most important visual information to process. These important visual information are also termed as salient regions.

In the past decades, there have been many studies investigating visual attention mechanisms from various fields of biology, psychology, neuroscience, computer vision, etc. Many computational models of visual attention have been proposed for saliency detection in 2D images/video. To evaluate the performance of computational models of visual attention, many eye tracking databases have been built on 2D images/video in the public domain. With the rapid development of stereoscopic display, there is an increasing interest in various emerging applications for 3D multimedia in the research community. Currently, there are several saliency detection models proposed for 3D images/video [1]. To evaluate the performance of the saliency detection models for stereoscopic images, some studies have conducted eye tracking experiments to obtain the ground truth.

At present, there are less studies focusing on building saliency detection models of stereoscopic video in the research community, compared with visual attention modeling for stereoscopic images. Different from visual attention modeling for 3D images, computational models of stereoscopic video have to consider the influence of the motion factor and this causes the visual attention modeling for stereoscopic video complicated. Another reason is that the lack of ground truth of eye tracking dataset for stereoscopic video have limited the studies of computational models of stereoscopic video. In this study, we investigate the human prediction on stereoscopic video based on eye tracking experiments. The study tries to provide some cues for visual attention modeling for stereoscopic video. Furthermore, the built eye tracking database provide a benchmark for performance evaluation of computational models of stereoscopic video for the research community.

2. THE EYE TRACKING DATABASE

2.1. Stereoscopic Video Sequences

The stereoscopic video dataset in this study includes 41 video sequences in total: 24 video sequences are from RMIT3DV database [2]; 3 video sequences are from Elephant Dream project; 2 video sequences are from Okugai; 12 video sequences are from the IVC stereoscopic video database.

2.1.1. RMIT3DV Database

The database is composed of video sequences by considering diverse content with different factors such as level of 3D effect, aesthetic composition, variations in colour, environment, motion, texture,light, etc. [2]. The stereoscopic video in this database were natively captured by a Panasonic AG-3D camera using Blackmagic Hyperdeck Shuttles and Intel SATA3 SSDs to record the video content. These video sequences are recorded with the resolution of 1920×1080 and the format is 10-bit 4:2:2 YUV at 25 fps without audio content. More details about this database can be referred to the original paper [2].

2.1.2. IVC Stereoscopic Video Database

A set of 12 stereoscopic video were captured by using a Panasonic AG-3DA1 3D camera [3]. Each video consists of two sequences of left and right views with full-HD resolution (1920 × 1080). This database includes both indoor and outdoor scenes taken in University of Nantes. Without using any depth range sensors during recording the stereoscopic video, the depth maps of the IVC stereoscopic video database are estimated by the optical flow approach on the stereo-pair video frames. The ground truth disparity maps are computed as the 'left-to-right' disparity map which represents the displacement of each pixel in the left view.

2.2. Procedures

The stereoscopic video sequences were displayed on a 26-inch (552 × 323 mm) Panasonic BT-3DL2550 LCD screen with the resolution of 1920 × 1200 pixels and the refresh rate of 60Hz. Each screen pixel subtends 61.99 arcsec at a 93 cm viewing distance. The maximum luminance of the display is 180 cd/m^2, which yields a maximum luminance of about 60 cd/m^2 when watched through glasses. The stereoscopic video sequences were viewed by subjects through a pair of passive polarized glasses. The environment luminance was adjusted for each subject with an appropriate size for the pupil during the eye tracking experiment. The eye movements were recorded by the SMI RED (4 firewire) remote eye-tracker working at 60Hz. During the eye tracking experiments, we asked the observers not to move and check them in live.

The stereoscopic video sequences were presented in a random order for the subjective test. During the experiment, subjects were required to do a free-viewing to the screen. The calibration test was performed at the beginning of the experiment, and repeated every eight video sequences. Subjects could ask for a break before each calibration. Each subject was required to have at least one break during the viewing test. All the experiments were conducted from 09:00 to 13:00 a.m. or 2:00 to 6:00 p.m. to avoid the feeling of fatigue as much as possible.

There were 40 subjects ranging from 19 to 44 years old involving the experiment. The average age of subjects is 26.02 years. All the subjects has either normal or corrected-to-normal visual acuity, verified by three pretests prior to the start of the experiment: the Snellen chart was used to check the visual acuity of subjects; the Ishihara color test was used to check color vision of subjects; the Randot stereo test was used to check 3D visual acuity of subjects. They were also naive to the purpose of the experiment and were compensated for their participation.

2.3. Fixation Density Map

After recording the gaze points by the eye tracker, we create the fixation density map as the ground truth for saliency es-
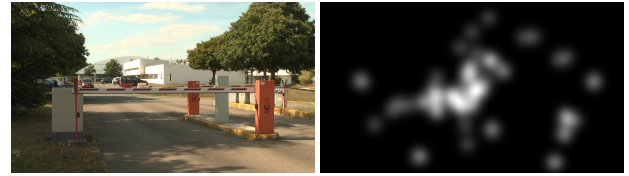

Fig. 1. First: the stereoscopic video frame; Second: the fixation density map.

timation for stereoscopic video. The gaze point maps from each eye were first created. The left gaze point maps were created by directly using the coordinates of the gaze positions of the left eye, while the right gaze maps were created by adding a displacement, horizontally and vertically, on the coordinates of each right-eye gaze point. We calculated the displacement of each gazed point from the 'right-to-left' disparity map computed by the same approach as the one used to crete the ground truth disparity maps of the IVC stereoscopic video database.

We summed these two gaze point maps as the final gaze point map and filtered it by a two-dimension Gaussian kernel based on the following two factors: (1) the decrease of visual acuity with increasing eccentricity from the fovea, and (2) the decrease in the accuracy of the eye tracker. The standard deviation of the Gaussian kernel is set as 2 degrees of the visual angle to guarantee the best visual acuity within the fovea. We provide one sample of the original video frame and its fixation density map in Fig. 1.

3. SUMMARY

This paper presents a large-scale eye tracking database for stereoscopic video sequences. The procedures and details of the database are illustrated. The database provides a benchmark for performance evaluation of computational models of visual attention for stereoscopic video. We will release this database to the research community in the future.

4. REFERENCES

[1] Y. Fang, J. Wang, M. Narwaria, P. Le Callet, and W. Lin, "Saliency detection for stereoscopic images," *IEEE T-IP*, Vol. 23, No. 6, pp. 2625-2636, 2014.

[2] E. Cheng, P. Burton, J. Burton, A. Joseski, and I. Burnett, "RMIT3DV: Pre-Announcement of a Creative Commons Uncompressed HD 3D Video Database," QoMEX 2012.

[3] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricordel, 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," QoMEX 2012.