DOI: 10.2390/biecoll-saga2013_1

In Pfeiffer and Essig (2013): Proceedings of the First International Workshop on Solutions for Automatic Gaze-Data Analysis 2013 (SAGA 2013). DOI: 10.2390/biecoll-saga2013_0

A Pilot Study on Eye-tracking in 3D Search Tasks

Ekaterina Potapova, Valsamis Ntouskos, Astrid Weiss, Michael Zillich, Markus Vincze and Fiora Pirri

1 Introduction

Eye-tracking is an important step in the evaluation of computational visual attention models for comparison to human visual perception. There exist two types of visual attention: bottom-up and top-down attention. Bottom-up attention is driven by object properties, while top-down attention takes place when there is a specific search task at hand. There exist several datasets for bottom-up attention, less experiments were made to collect eye-tracking for specific search tasks, such as driving a car by Borji *et al.* [1]. In this work we are interested in collecting statistically valuable eye-tracking data for visual search tasks that are typical in daily life and of great interest to researchers in robotics and HRI.

The majority of existing datasets is concerned with the analysis of static images or video sequences. However, it has been shown that the inclusion of depth perception changes human attention behavior [2] and several attempts have been made to collect eye-tracking data while free-viewing stereo image pairs [3], artificially created 3D scenes [4], and even real 3D environments [7]. Therefore, we want to create a dataset in which people perform tasks in real 3D environments. Collecting such experimental 3D eye-tracking data will strongly benefit the community, and will show the influence of depth and motion on human perception.

Ekaterina Potapova, Astrid Weiss, Michael Zillich, Markus Vincze
ACIN, Vienna University of Technology, e-mail: {potapova,weiss,zillich,vincze}@acin.tuwien.ac.at

Valsamis Ntouskos, Fiora Pirri
ALCOR, University of Rome, e-mail: {ntouskos,pirri}@dis.uniroma1.it

The research leading to these results has received funding from the Austrian Science Foundation (FWF) under grant agreements No. TRP 139-N23, InSitu and No. TRP 623-N23, V4HRC, and the European Communitys Seventh Framework Programme FP7/2007-2013 under grant agreement No. 600623, STRANDS

We aim at producing a pilot dataset with 3D eye-tracking data from three participants to check if it is feasible to create a benchmark for saliency algorithms and computational attention models aimed at solving top-down search tasks.

2 Task Description

In the following we will present two exemplary tasks participants were asked to perform in our pilot study: a free-viewing and a counting task. We collected eye-tracking data in real cluttered environments using the Gaze Machine [5].

2.1 Experimental Setup

As a setup for our experiments we chose a cluttered scene, as it offers a variety of possibilities for bottom-up and top-down search tasks. The selected scene represented a pile of toys as it can be found in every child's room. Toys are stacked together and occluding each other. Therefore, 3D data is essential, because depth can potentially help to separate items. Our experimental setup consisted of a set of toys, randomly placed on the table. A white screen was placed in front of the participant and lifted for a specific amount of time, ensuring a precisely defined duration of viewing the scene. All questions to participants and their answers were documented. Additionally, experiments were video taped.

2.2 Free-viewing Task

In this task we studied bottom-up attention. In order to keep participants motivated the goal was to remember as many details as possible during a limited period of 30 seconds. After the completion of the experiment participants were asked several questions, to check their performance, such as 'Where can this type of scene be typically observed?', 'How many toys are in the scene?', 'Could you please name some toys you have seen?'. Points of regard were mapped onto image sequences for each participant separately to create preliminary fixation distance maps. We expected participants to be attracted by regions being in contrast with the surrounding, similar to [6].

2.3 Counting Task

In this task we studied top-down attention. We asked participants to search for cars in the scene and count them loudly when they find one. Search patterns were recorded. The experiment was stopped when a participant reported that all cars found. With such data it becomes possible to study human behaviour in terms of evaluation of directed top-down visual attention. We expected participants to sequentially search the scene from the most attractive region, i.e. saliency region, to the least attractive.

3 Participants

As a pilot study we invited three participants to check the feasibility of our methodological approach. All three were male students at the University of Rome, with age from 27 to 29. All of them had normal or corrected to normal vision.

4 Data Processing

To create a benchmark that can be used by other researchers, a 3D reconstruction model of the scene was created at first. Points of regard were localized and mapped onto this reconstruction. However, because of lighting conditions and image blurriness due to participants movements, it was not possible to localize points of regard on the unified model. Therefore, it was not possible to create hit maps in 3D. These issues will be addressed in the followed-up main study. Additionally, points of regard were mapped to one of two images in a stereo pair. Fixation distance maps were created for each frame for each participant. No averaged fixation distance maps were created at this points.

5 Preliminary Results

No hit maps were created in the pilot study, therefore it is not possible to quantitatively compare eye-tracking data with existing models of visual attention. Preliminary results show that attentional models designed to work on static images or videos fail to predict fixations when they are applied in real world environments.

The results of the free-viewing experiment were qualitatively compared to contrast-based attentional model described by Itti *et al.* [6]. Comparison showed that while contrast-based model predicts fixations on objects popping out from the scene based on color, human fixations were mostly distributed in areas with multiple tiny objects. One interesting question here is, whether this behaviour was influenced by the given task to motivate participants or by the nature of human attention. Potentially, it can show that humans are attracted by textured regions. This aspect should be investigated in more details.

Fixation data from counting experiment showed that humans tend to return to previous fixations. Currently existing models of human visual attention do not assume the possibility to go back and forth between already attended location. Potentially, this finding can indicate that humans tend to verify their visual experience by examining same regions with interesting objects several times.

In general, the pilot study greatly helped to adjust the setting for the upcoming studies. Data, obtained from the pilot study can lead to designing an attentional system comparable to human performance, when depth and orientation in space are taken into account.

6 Conclusion and Future Work

In this paper we addressed the issue of creating a benchmark for visual attention models based on eye-tracking data in realistic 3D environments. As an example a setup with cluttered toys was created. Participants performed two different tasks: free-viewing and counting. Pilot study showed that classical contrast-based models fail to predict human fixations in real 3D environments. Next steps will be to create a unified 3D reconstructed model of the scene and map fixation points onto this model. Our future work will also include the design of a general protocol to perform more experiments on visual search and algorithms to process eye-tracking data.

References

1. Borji A., Sihite D., Itti L. (2012). Probabilistic Learning of Task-Specific Visual Attention. In: Computer Vision and Pattern Recognition, 2012.
2. Wexler M., Ouarti N. (2008). Depth Affects Where We Look. In: Current Biology 18, 2008.
3. Lang C., Nguyen T., Katti H., Yadati K., Kankanhalli M., Yan S. (2012). Depth Matters: Influence of Depth Cues on Visual Saliency. In: European Conference on Computer Vision, 2012.
4. Vishwanath D., Kowler E. (2004). Saccadic localization in the presence of cues to three-dimensional shape. In: Journal of Vision 4, 2004.
5. Pirri F., Pizzoli M., Rudi A. (2011). A general method for the point of regard estimation in 3D space. In: Computer Vision and Pattern Recognition, 2011.
6. Itti L., Koch C., Niebur E. (1998). A Model of Saliency-Based Visual Attention for Rapid Scene Analysis. In: IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11), 1998.
7. Santner K., Fritz G., Paletta L., Mayer H. (2013). Visual Recovery of Saliency Maps from Human Attention in 3D Environments. In: IEEE International Conference on Robotics and Automation, 2013.