DOI: 10.2390/biecoll-saga2013_10

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

Location-based Online Identification of Objects in the Centre of Visual Attention using Eye Tracking

Kai Harmening, Thies Pfeiffer

1 Introduction

Modern mobile eye trackers calculate the point-of-regard relatively to the current image obtained by a scene-camera. They show where the wearer of the eye tracker is looking at in this 2D picture, but they fail to provide a link to the object of interest in the environment. To understand the context of the wearer's current actions, human annotators therefore have to label the recorded fixations manually. This is very time consuming and also prevents an online interactive use in HCI.

A popular scenario for mobile eye tracking are supermarkets. Gidlöf et al. (2013) used this scenario to study the visual behaviour in a decision-process. De Beugher et al. (2012) developed an offline approach to automate the analysis of object identification. For usage of mobile eye tracking in an online recommender system (Pfeiffer et al., 2013), that supports the user in a supermarket, it is essential to identify the object of interest immediately. Our work addresses this issue by using location information to speed-up the identification of the fixated object and at the same time making detection results more robust.

2 Approach

The purpose of our framework is to identify the object in the focus of visual attention based on the scene-camera image and the point-of-regard in a mobile interaction. In our application example, we want to identify the fixated cereal package in a decision situation in a supermarket.

Kai Harmening, kharmening@techfak.uni-bielefeld.de
SFB 673, Faculty of Technology, Bielefeld University, Germany

Thies Pfeiffer, tpfeiffe@techfak.uni-bielefeld.de
CITEC – Cognitive Interaction Technology Center of Excellence, Bielefeld University, Germany

Figure 1: Example of an annotated panorama image of a shelf. The objects relevant for the study are manually identified once for the setup using shapes (here rectangles). The annotation is done using a general purpose vector graphics format (SVG).

We use a hierarchical database to store our objects of interest while preserving topological information as far as possible. For the supermarket, the top-level elements are panorama-pictures taken from the relevant shelves. The second level holds all relevant objects, here cereal packages, which are pre-labelled by human annotators in the panorama picture (see Figure 1). Using this process, we get image examples of relevant objects together with topological information. For each extracted image, multi-dimensional fingerprints are computed (computer vision, feature extraction) using a set of pre-configured methods. Currently we support SIFT, SURF and ORB.

To find the focused object, an area in the scene-image around the point-of-regard is extracted. Based on the assumption that the user has not moved very much since the last frame (i.e. within 40 ms), the search process starts within the neighbourhood of the previously fixated and successfully identified object. This step uses the topological information encoded in the hierarchical database, here the panorama images, to significantly speed up the search process. Only if this approach is not successful, the search domain is expanded by climbing up the hierarchy. Finally, the found object and useful information like the name of the product is visualized (see Figure 2).

3 Preliminary Results

In a first evaluation of the system, the area in the panorama-picture fixated by the participant could be successfully identified in 416 of 470 consecutive images (88.51%). An object could be classified in 336 of these images (71.49%) and in 329 images (70%) the classified object was the one being focussed.

4 Discussions and Further Work

The presented approach uses a database of objects to identify fixated objects of interest. Using information about the topology of the objects, the system is able to classify sequences of fixations in the same area, e.g. within one shelf, in real-time. Only when the environment completely changes, a re-localization based on computer vision may take some time. To solve this problem, the database is prepared to support other means of localizations, such as GPS.

The presented approach can be used with online and offline data. While our current focus is on the online capabilities for human-computer interaction the offline modus works even better because there is more time for the computer vision processes and when the position in the environment is lost, the full database can be searched in the worst case. This is not feasible in the online approach.

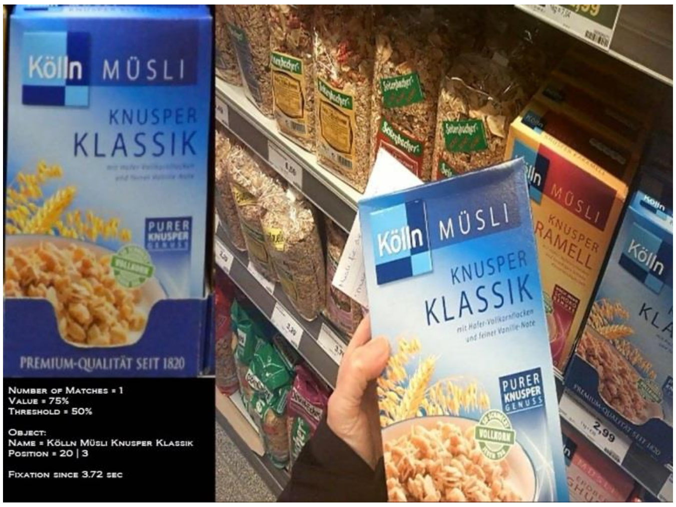

Figure 2: Application of the present work in a supermarket, Right: Consumer with an object of interest, Left: Found object in the database with information

References

1. Gidlöf K, Wallin A, Dewhurst R, Holmqvist K (2013). Using Eye Tracking to Trace a Cognitive Process: Gaze Behaviour During Decision Making in a Natural Environment. Journal of Eye Movement Research 6(1):3, 1-14
2. De Beugher S, Younes I, Brône G, Goedemé T (2012). Automatic analysis of eye-tracking data using object detection algorithms. UbiComp ’12 677-680
3. Pfeiffer T, Pfeiffer J, Meißner M (2013). Mobile Recommendation Agents Making Online Use of Visual Attention Information at the Point of Sale. Gmunden Retreat on NeuroIS