

Automatic object identification and analysis of eye movement recordings

MATTHIAS RÖTTING¹, MATTHIAS GÖBEL² AND JOHANNES SPRINGER¹

¹ *Lehrstuhl und Institut für Arbeitswissenschaft der RWTH Aachen
Bergdriesch 27, D-52062 Aachen, Germany*

² *Institut für Arbeitswissenschaften der TU Berlin
Steinplatz 1, D-10623 Berlin, Germany*

*Schlüsselwörter: Blickbewegungen, Blickbewegungsregistrierung, Kopfbewegungs-
registrierung, Automatische Auswertung, Busfahrer, Straßenbahnfahrer*

1. Introduction

The use of eye movement recording can be helpful in the design and evaluation of workplaces, products and many other research topics. As long as only parameters related to the movements of the eyes within the head of the subject (e.g. saccades and fixations) are to be used, an automatic evaluation seems to be the standard procedure. However, often these data should be related to the objects the subject interacts with. An identification of the objects is hindered by severe drawbacks: Either the head of the subject must be fixed or a manual analysis of the eye movement recordings is necessary (requiring roughly 30 to 60 times the acquisition duration).

For these reasons a system was developed that allows automatic object definition for all research applications where the relevant objects are fixed in space and a semi-automatic identification for moving or not previously definable objects.

2. System overview

The system consists of an eye movement recorder, a 6 degree of freedom (DOF) head tracker, and a personal computer. The computer is equipped with a custom converter board and/or an analog-to-digital (A/D)-converter for acquiring the eye movement data and a video overlay board. Currently the system works together with the eye movement recorder NAC EMR 600, but should be adaptable to all other eye trackers that output eye data either as analog signal or coded in the blanking period

of the video signal. As head tracker the Flock of Birds[®] of Ascension Inc. is used. Basic output of the system is a list of the fixations and the fixated objects along with the description of the saccade (width, mean velocity, accelerations) leading to the fixation. These data are used to identify, among others, transition frequencies and fixation- and action-sequences.

The video overlay board enables the visualization of the data in relation to the stimulus material.

Figure 1 gives an overview of the system components.

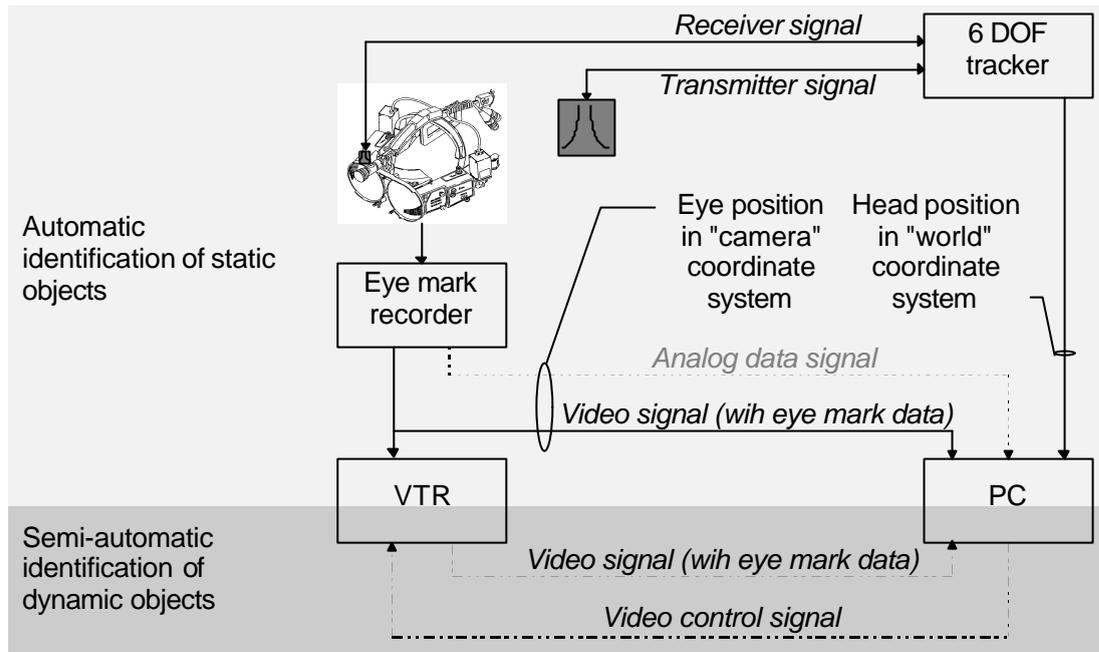


Figure 1: Overview of the relevant components for automatic and semi-automatic object identification.

3. Automatic identification of static objects

The automatic identification of objects requires that the objects are fixed in space. This is usually the case for machine interfaces, control panels, advertisements and often for graphical computer interfaces. It should be possible to describe different sets of static objects, e.g. sequences of advertisements or a succession of different screen layouts of a computer program. An automatic object identification is impossible if the objects are moving unpredictably in space. A typical example are traffic situations.

The following description will refer to two different coordinate systems:

The eye movements are registered with an eye movement recorder (EMR) worn on the head of the subject and are superimposed on the image recorded by the field camera of the EMR. Therefore, the eye position is specified in relation to the field camera. This coordinate system will be called "camera" coordinate system.

The transmitter of the 6 DOF tracker constitutes the origin of the coordinate system that later will be referred to as the "world" coordinate system. The objects are defined in relation to the "world" coordinate system (see below). Mounted on the field

camera of the EMR is the receiver of the 6 DOF tracker. The position of the field camera in "world" coordinates is recorded by the computer in parallel to the eye movement data.

The automatic object identification can be described in six steps (see figures 2 to 6).

1. Enter object position in "world"-coordinates. The camera of the eye movement recorder is directed towards the scene with the objects and the video picture is displayed on the computer monitor with the help of the video overlay board. The position of the camera in "world" coordinates is registered. A simple drawing program is used to line out the different objects. Each object can be designated by a name. This procedure is repeated for a second camera position. The differences between the two sets of object data are used to calculate the relative depth of the objects (see figure 2). With this depth information it is possible to automatically identify the different objects even when the person wearing the EMR is moving in direction of the z-axis. A list of 2.5 D object coordinates is stored in the computer.

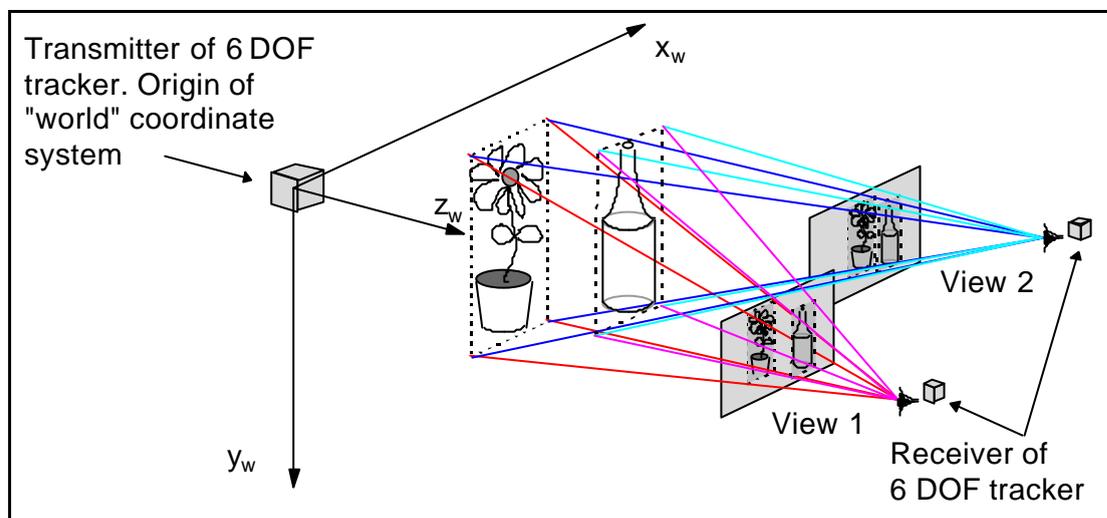


Figure 2: Two different camera positions are necessary to specify the objects in 2.5 D (see text for a description of the procedure).

2. Register eye movements relative to the "camera" coordinate system. Eye position data (relative to the "camera" coordinate system) with a time resolution of up to 600 Hz are stored by the computer.

3. Register head movements relative to the "world" coordinate system. Parallel to the eye data, the output of the 6 DOF tracker - the field camera or head position - is stored by the computer with a resolution of up to 60 Hz.

4. Determine fixations. On the basis of the eye position data fixations are determined. What constitutes a fixation? A fixation can be defined as a state, where the eye is in relative rest for a certain period of time. Hence the system uses a time and a velocity criterion to define a fixation (see e.g. Unema & Rötting 1990). Both criterions can be specified by the user to suit the experimental set up. The velocity criterion is fulfilled, if the eye does not move with a velocity higher than the specified maximum velocity. The time criterion is met, if the eye conforms to the velocity criterion for a time longer than the specified minimum amount of time. Only in case both criteria are met, a fixation is marked.

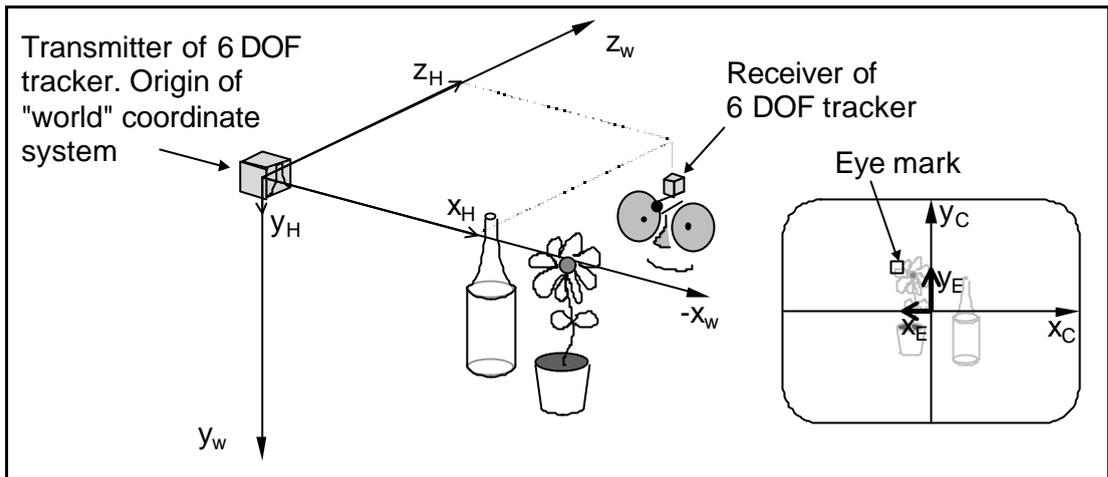


Figure 3: The eye movements are registered in relation to the field camera. By registering the position of the field camera in "world" coordinates it is possible to calculate the position of the eyes in "world" coordinates and relate the fixation point to an object.

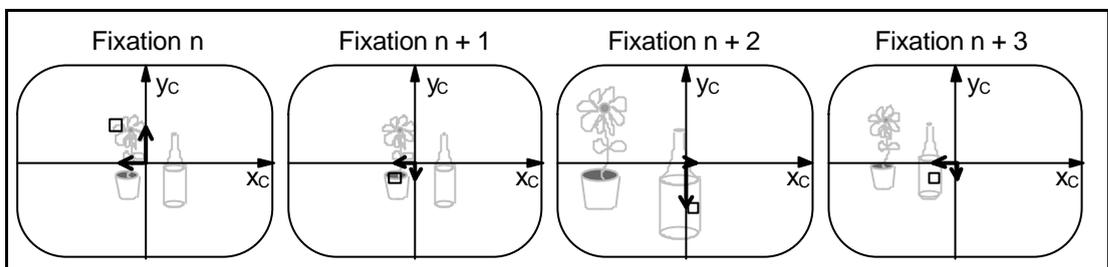


Figure 4: Illustration of a sequence of fixations. Although the position of the fixation $n+1$ and $n+3$ are identical in camera coordinates, different objects are looked at due to a movement of the head.

5. Determine fixation position in "world" coordinates. Use the position of the fixations in "camera" coordinates and the position of the field camera during that fixation to determine the direction of the gaze in world coordinates.

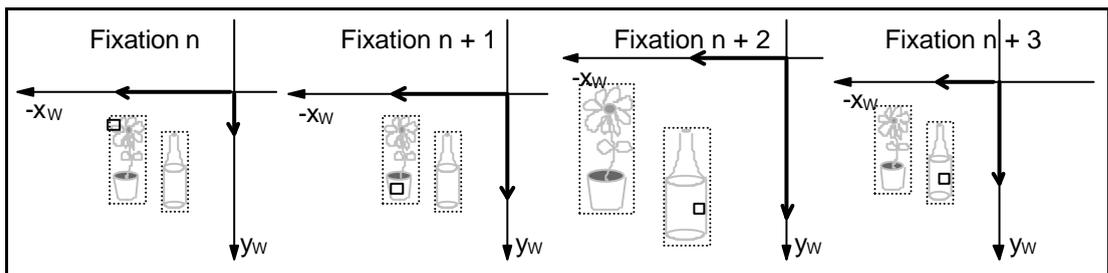


Figure 5: The gaze direction in "world" coordinates can be determined (see text for description).

6. Determine fixated object. The object list build in step 1 is used to determine the object the person fixates. The name of the object is added to the description of the fixation. Mean values of all the parameters in the fixation list can now be calculated for every object. Additionally, transition frequencies and typical fixation sequences can now be analyzed.

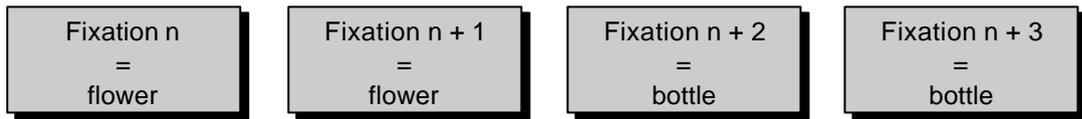


Figure 6: The fixation coordinates in "world" coordinates can now be related to an object.

4. Semi-automatic identification of dynamic objects

As mentioned above, an automatic object identification is not suited for moving objects. The traditional procedure to identify the objects is a manual frame by frame analysis of the video with the recorded eye movements. The new semi-automatic process uses a computer controlled video recorder and custom software. A script file must be written that describes all possible objects and which may contain task activities as well. Based on this data the program generates an input mask, where objects and activities can be selected with the mouse. Figure 7 shows a screen used for the analysis of a tram drivers' eye movement and activities.

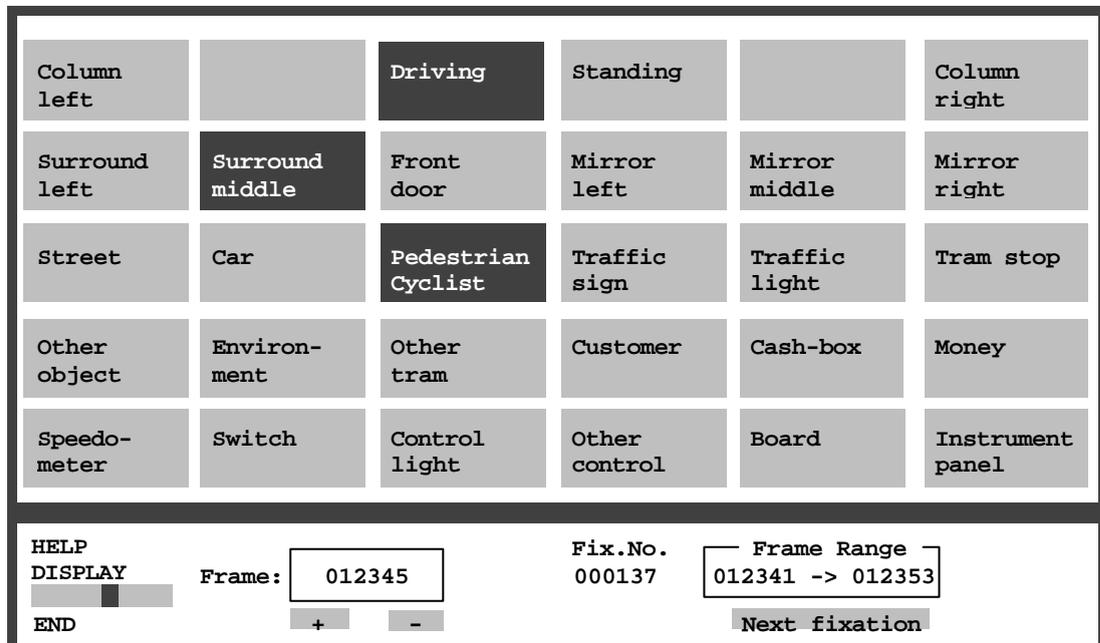


Figure 7: Screen layout for the semi-automatic analysis of eye movements and activities, here for a tram driver. The upper part of the screen shows the different activities (e.g. driving or standing) and objects. All can be selected with a mouse-click. Displayed at the bottom is the number of the currently displayed video frame, the number of the current fixation, and the numbers of the frames where the fixation begins and ends. The "+" and "-" under the frame number are used to display the previous and next frame. A press on the "Next fixation" button ends the input for the current fixation and displays the next one.

The interactive object identification starts with the (automatically generated) list of fixations. The computer sends a control sequence to the video recorder and the video is wound to a frame within the first fixation. The frame number is coded with the eye data in the blanking period of the video signal and read by the custom converter card.

After the user selected the respective object (and activity) from the input mask, "Next fixation" is pressed to continue. Output of the semi-automatic object identification is a time sequence of fixated objects correlated with the task activities.

Compared with a manual analysis, the semi-automatic analysis saves time because not every frame, but ideally only one frame within every fixation must be coded. Typically, a manual analysis takes about 30 to 60 times the acquisition duration, for the semi-automatic analysis this can be shortened to 5 to 10 times the acquisition duration.

5. Evaluation

Fixation data and saccade data can be superimposed on the video taken by the EMR field camera. A "static" mode can be used to display fixations and saccades for constant stimulus material, like computer user interfaces, control room panels or advertisements.

A "dynamic" mode enables the "replay" of the eye movements. Not only the fixation for the currently displayed video frame can be shown, but also past and even future fixations. This mode is suited for dynamic stimulus material and can enhance the understanding of the eye movements.

Both overlay modes are especially valuable for training purposes and presentations of results.

6. Applications

The system was used for various research applications, including bus and tram driving (Göbel, Scherff, Springer & Luczak, 1994)) and the design of a 3D radar display for air traffic control tasks (Göbel, Stallkamp & Springer, 1994).

The registration of eye movements during bus and tram driving was part of exploratory research projects. Common aim of both projects is the redesign of the drivers' workplace.

The percentage of time both bus and tram drivers spent looking at certain objects is quite similar (figure 8). A relatively large difference is only seen for the columns next to the front window. The bus drivers seem to use them for navigation.

The mirrors are looked at for about ten per cent of the whole driving time. Not only the traffic has to be watched, but the passengers inside as well. Analysing the mirrors further (figure 9) reveals that about every fourth gaze shift terminates in a mirror (23.8% bus, 26.6% tram). The only difference between bus and tram is the relative importance of the different mirrors: Whereas the left mirror has the highest relevance for bus drivers, the trams are not even equipped with a left mirror.

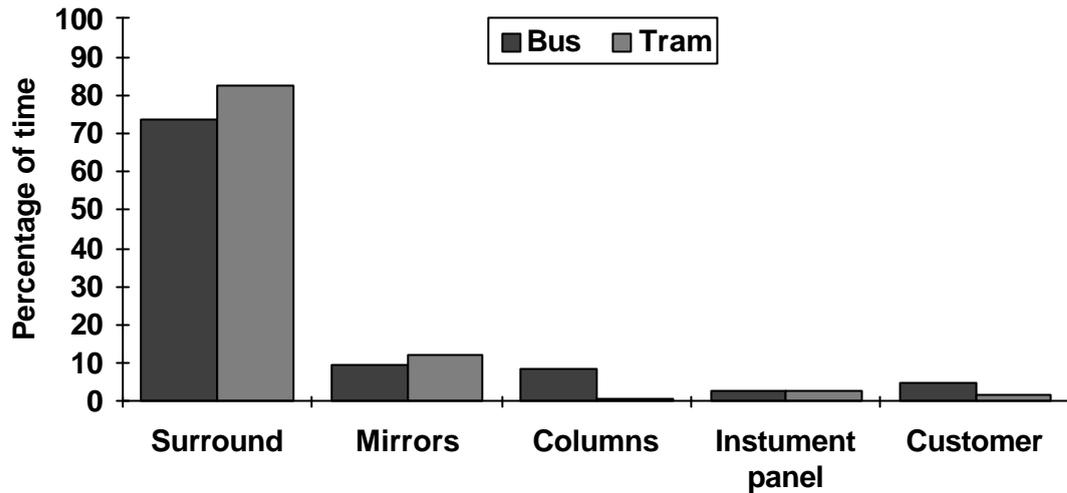


Figure 8: Percentage of time bus and tram drivers look at different objects.



Figure 9: Number of gaze changes that terminate in one of the mirrors. The total number of gaze changes is 1883 per hour for the bus drivers and 1270 per hour for the tram drivers.

Compared to former analysis procedures the time for analysis could be shortened and simultaneously the meaningfulness of the resulting data was improved.

7. References

- Göbel, M., Scherff, J., Springer, J. & Luczak, H. (1994). *Bus driving task and stress analysis during inner-city-driving*. Proceedings IEA'94, Vol. 6, Part 2, p. 422.
- Göbel, M., Stallkamp, J. & Springer, J. (1994). *Three-dimensional radar displays for air traffic control (in German: Dreidimensionale Radarbilddarstellung bei der Flugüberwachung)*. ASIM Arbeitsgemeinschaft Simulation in der Gesellschaft für Informatik, Mitteilungen aus den Arbeitskreisen Nr. 42. Magdeburg: Otto von Guericke Universität.
- Unema, P. & Rötting, M. (1990). Differences in eye movements and mental workload between experienced and inexperienced motor vehicle drivers. In D. Brogan (Ed.), *Visual Search*. London: Taylor & Francis, pp. 193-202.