COMPENSATION OF HEAD MOVEMENTS IN THE DATA REGISTERED WITH A HEADSET EYE TRACKER USING THE EVM* SOFTWARE PACKAGE

Bibianna BAŁAJ1, Piotr FRANCUZ2, Marek STERNAL3, Jacek MATULEWSKI3
1 Nicolaus Copernicus University in Torun, Faculty of Humanities, Institute of Philosophy, Torun, Poland
2 John Paul II Catholic University of Lublin, Faculty of Social Sciences, Institute of Psychology, Department of Experimental Psychology, Lublin, Poland
3 Nicolaus Copernicus University in Torun, Faculty of Physics, Astronomy and Applied Informatics, Institute of Physics, Torun, Poland

Source of support: Study was financed with the National Science Center grant titled "Oculomotor, electroencephalographic, and behavioral activity during execution of perceptive and cognitive tasks" (UMO-2013/09/B0HS6/03266).

Author's address: J. Matulewski, Nicolaus Copernicus University in Torun, Faculty of Physics, Astronomy and Applied Informatics, Institute of Physics, Grudziacka 5 Street, 87-100 Torun, Poland, e-mail: jacek@phys.uni.torun.pl

Introduction: Comparative analysis of the data obtained with a headset eye tracker while driving a motor vehicle (e.g. plane or car) is subject to numerous errors and approximations. It is due to the movements of pilot’s or driver’s head in relation to fixed elements of a cockpit of a motor vehicle, and - in case of a simulator – to fixed elements located outside the cockpit (e.g. the same cloud pattern or objects along the street). Movement of the head evokes change to image captured by the headset camera. As a result, the positions of fixation points within the areas of interest (e.g. the speedometer) are stored in different coordinate systems relative to the edges of the current image.

Methods: Our solution is based on an automatic image analysis; it allows mapping of gaze positions recorded in the flight simulator, while the pilots are performing the tests, to the fixed system of coordinates of the simulator monitor. This solution has been implemented in the original Eye View Mapping (EVM) software package. During software testing we exported 200 random samples of flight profiles and conducted a comparative analysis of visual fixation in the regions of interest before and after the transformation.

Results: We presented in detail the steps of analysis and image transformation using EVM and the process of comparing gaze position before and after transformation. Gaze most often fixated on the artificial horizon indicator, followed by the external view and the altimeter. Positions before and after transformation were 100% congruent.
INTRODUCTION

Examination of machine operators in natural or near-natural environment is a source of numerous problems related to the assessment of the quality of this examination due to the limitations of measuring equipment. These problems largely concern the tasks performed during flight control [13] or flying an aircraft. Piloting an aircraft requires a number of complex and demanding tasks [4]. Pilots must simultaneously monitor the position of the plane based on navigation parameters and visual cues from the outside the cabin [2]. Nonetheless, it is important to control the equipment registering flight parameters. These tasks engage primarily visual attention indicated by position, time, or sequence of visual fixation [16].

Using a headset eye tracker is a non-invasive method of examination of pilot’s visual attention in the cockpit or flight simulator [3]. It does not reduce the visual field and does not constrain pilot’s movements, while allowing the researchers to register pilot’s gaze directed at any point inside or outside the cabin. Establishing regions of interest on images registered using a headset camera enabled analysis of visual fixation time in selected regions of interest (inside and outside the cockpit) as well as the frequency of shifting visual attention between those regions. Moreover, registration of eye movements allows for estimating the indexes of situational awareness [7,20] and changes in pilot’s cognitive load in various phases of flight [5, 6]. Measuring the cognitive load during takeoff and landing maneuvers is particularly important due to increased requirements and reduced cognitive reserve of a pilot [21]. On the other hand, boredom and distraction may occur during flight phases characterized by lesser load, resulting in decreased sensitivity to crisis situations [17]. Therefore, regardless of the flight phase, tracking of pilot’s eye movement is an excellent measure of his effectiveness during execution of flight-related tasks. It may be also an indicator of pilot’s competence. Generally, expert pilots are characterized by shorter fixation time and greater fixation frequency than amateurs [10].

Using an eye-tracking headset during piloting is also burdened with a significant drawback. It is related to subject’s head movements, where the eye-tracking device and camera for registration of the surroundings are mounted. Freedom of movement is the reason why image captured by the headset camera is not a stable representation of a single fragment of the scene (the monitor, the cockpit). As a result, current gaze position reported by the oculographic device in a reference system relative to the edges of the image (frame) is not consistent with its actual position relative to the fixed elements of the scene, such as e.g. the screen in the control panel of the cockpit. Therefore, it is impossible to define fixed areas of interest within captured image; it is also not possible to conduct any kind of comparative analyses of the eye movement data. Similar problems related to conducting research as well as data analysis also appear in other studies utilizing portable oculographs, e.g. while registering open spaces [8,11], in virtual reality [14], as well as in marketing research [9,19].

In order to solve this problem it is necessary to transform the collected data in such a way to depict time-dependent head position and orientation in a fixed reference system, e.g. relative to flight simulator monitor. A head-mounted accelerometer is used for that purpose, providing information regarding orientation of the head relative to all variables registered by the oculograph. However, such solution requires additional equipment and may be a source of errors related to synchronization of both measurement devices in time.

Analysis of an image captured by the headset camera using Semantic Gaze Mapping (SGM) method is an alternative solution. It requires determination of fixed reference frames for all film frames (it may be for example a monitor outline on the control panel) followed by manual marking of gaze position (based on registered gaze superimposed on a scene image in the film) to its corresponding place in the reference frame. It is a more time-consuming and, above all, less precise method (Fig. 1).

Data processing involves examining the entire record, fixation after fixation. Gaze fixation positions obtained from subsequent film frames (image on the right) are manually applied onto
B. Bałaj et al. - Compensation of head...

Software package contains a library for NET 4.0 framework, which enables loading, editing and saving changes to IDF files following their conversion into a text format (SMI Co., manufacturer of the device does not provide specifications for the binary format). Editing involves transformation of gaze coordinates both in the constant as well as time-dependent transformation matrix. It is worth noting that coordinates written into the IDF file by the headset eye tracker pertain to the gaze location on a “virtual screen”, i.e. on a plane placed away from the oculography (on the Z axis) at such a distance that one pixel on a movie frame corresponds to a unit of measurement for gaze position.

METHODS

EVM software

Our goal was to automate the processing of the eye movement data for gaze analysis based on the video recording from headset eye tracking device. We obtained it by analyzing the location of simulator monitor in the frame images of the movie recorded by the oculography, calculating time-dependent transformation matrix in homogene-ous coordinates, which describes relative orientation of pilot’s head and the cockpit. The EVM (Eye View Mapping) software enables transformation of oculomotor data acquired by the headset eye tracker into a reference frame of flight simulator monitor. In this publication we present the method itself, the software package, and the results of its preliminary verification.

Software package contains a library for NET 4.0 framework, which enables loading, editing and saving changes to IDF files following their conversion into a text format (SMI Co., manufacturer of the device does not provide specifications for the binary format). Editing involves transformation of gaze coordinates both in the constant as well as time-dependent transformation matrix. It is worth noting that coordinates written into the IDF file by the headset eye tracker pertain to the gaze location on a “virtual screen”, i.e. on a plane placed away from the oculography (on the Z axis) at such a distance that one pixel on a movie frame corresponds to a unit of measurement for gaze position.

The essential part of the package consists of a program with graphic user interface, which conducts image analysis using the algorithms implemented in the OpenCV library (http://opencv.org/). As a result, affine transformation in homogene-ous coordinates (3x3 matrix) is determined...
for each frame. Such transformation converts the coordinates of a given point from the reference frame of eye tracker’s image to the frame of reference of flight simulator monitor visible on this recording. Individual stages of this analysis are presented in Table 1. Such transformation is performed in order to convert subsequent frames in such a way that flight simulator’s monitor would keep still and maintain size.

Two masks for selecting the part of the screen that is particularly interesting for the researcher are produced during stages II and III presented in Table 1. The first mask (see: cell II.5, Table 1) is meant for removing static and dynamic elements of the image that are unrelated to the experiment itself (that might include artifacts caused by damage or imperfection of eye tracker camera or a logo added by eye tracker’s driver). The second mask (see: cell III.3, Table 1) is meant for designating the area occupied by the simulator monitor for each film frame separately, which is easier on the assumption that the simulator is dark during the flight and the screen image stands out. As a result, after stage III using the Suzuki & Abe algorithm [18] we find the most fitting convex hull (bounding polygon) that assumes a shape of an oriented bounding box, i.e. at any position and orientation (see: red contour on the figure in cell IV.1, Table 1), within a black-and-white film frame with clearly visible monitor screen. The content of this rectangle, the position of which changes in every film frame, should be transformed in such a way as to constitute the full content of the frame after the transformations (see: cell IV.2). Affine transformation of the plane depends on (with some simplifications) the orientation and position of the head of a person carrying the headset eye tracker camera. Such transformation, since it must also include translation, must be expressed through homogeneous coordinates, where a so-called scaling coordinate $w$ is added to $x$ and $y$ gaze position coordinates. Derivation of the transformation matrix is presented below.

In general, the transformation of interest is an affine transformation represented as a 2x2 multiplication by matrix $A$ (with the following elements: $a_{xx}, a_{xy}, a_{yx}$, and $a_{yy}$) and translation with the use of $\mathbf{t}$ vector (with the following elements: $t_x$ and $t_y$), which can be written as:

$$
\begin{pmatrix}
  x' \\
  y' \\
  w'
\end{pmatrix}
= A
\begin{pmatrix}
  x \\
  y \\
  w
\end{pmatrix}
+ \mathbf{t}
$$

(1)

and in homogeneous coordinates may be written as:

$$
\begin{pmatrix}
  x' \\
  y' \\
  w'
\end{pmatrix}
= A
\begin{pmatrix}
  x \\
  y \\
  1
\end{pmatrix}
+ \mathbf{t}
$$

(2)

The coordinates refer to a transformed coordinate system and the $x'y'$ coordinates to the original one. Transformation matrix contains six unknowns (the elements of matrix $A$ and $\mathbf{t}$ vector), which means that in order to determine it unequivocally we must know the coordinates of three points before and after transformation. It is most convenient if they are vertices of a rectangle surrounding the screen on the original picture and the vertices of the full picture for coordinates after transformation. We can write the equations for those points indexed from 1 to 3:

$$
\begin{aligned}
  x'_1 &= a_{xx}x_1 + a_{xy}y_1 + t_x \\
  y'_1 &= a_{yy}y_1 + t_y \\
  x'_2 &= a_{xx}x_2 + a_{xy}y_2 + t_x \\
  y'_2 &= a_{yy}y_2 + t_y \\
  x'_3 &= a_{xx}x_3 + a_{xy}y_3 + t_x \\
  y'_3 &= a_{yy}y_3 + t_y \\
\end{aligned}
$$

(3)

Sticking to the matrix notation, we could write the above set of equations as follows:

$$
\mathbf{X}' = B \mathbf{X}
$$

(4)

thus, in homogeneous coordinates:

$$
\begin{pmatrix}
  x'_1 \\
  y'_1 \\
  x'_2 \\
  y'_2 \\
  x'_3 \\
  y'_3 \\
  w'
\end{pmatrix}
= \begin{pmatrix}
  a_{xx} & a_{xy} & 0 & 0 & 0 & 0 & t_x \\
  a_{yy} & 0 & 0 & 0 & 0 & 0 & t_y \\
  0 & 0 & a_{xx} & a_{xy} & 0 & 0 & t_x \\
  0 & 0 & a_{yy} & 0 & 0 & 0 & t_y \\
  0 & 0 & 0 & 0 & a_{xx} & a_{xy} & t_x \\
  0 & 0 & 0 & 0 & a_{yy} & 0 & t_y \\
  0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
  x \\
  y \\
  x \\
  y \\
  x \\
  y \\
  w
\end{pmatrix}
$$

where six-element $\mathbf{X}$ and $\mathbf{X}'$ vectors contain the coordinates of points in the original and the transformed coordinate system, respectively. Matrix $B$ coefficients are unknown in this set of equations. However, it is worth looking at the set of equations (3) a little differently. Let us rewrite them as if it was a transformation of a six-element $\mathbf{Y}$ vector consisting of the sought elements of $\mathbf{B}$ matrix transformation. Transformation matrix with $\mathbf{M(\mathbf{X})}$ matrix contains the $\mathbf{X}$ vector coordinates, thus the coordinates of points in the original coordinate system. As a result, we obtain the following equation:
In this case, we are searching for vector $C$, which may be easily found by inverting the matrix $M$:

$$C = M^{-1}(X)X'$$

(7)

We may find the analytical form of the matrix inverse to $M$; however, its size makes it impossible to present clearly. The OpenCV library was used in the EVM software to find the inverse matrix and multiply it by the $X'$ vector. Such calculations must therefore:

$$X' = M(X)C$$

(6)

---

Tab. 1. Stages of image processing from eye tracker camera scenes.

<table>
<thead>
<tr>
<th>Result of image processing</th>
<th>Sample frame</th>
<th>Obtained effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>I. Scene capture</td>
<td></td>
<td>Crude headset camera image with marked coordinates of left and right eye position obtained from the IDF file (red and green circles, respectively) as well as gaze position obtained using BeGaze (orange circle).</td>
</tr>
<tr>
<td>II. Generating a mask for static and dynamic artifacts</td>
<td><img src="image1.png" alt="Frame of a static image present throughout the film. In our case it is the logo added by the manufacturer of SMI equipment, but it could also be an artifact related to e.g. camera damage. EVM user can individually generate a schematic frame using any recording." /></td>
<td>Frame of a static image present throughout the film. In our case it is the logo added by the manufacturer of SMI equipment, but it could also be an artifact related to e.g. camera damage. EVM user can individually generate a schematic frame using any recording.</td>
</tr>
<tr>
<td></td>
<td><img src="image2.png" alt="A frame that is selected based on a unique color is a mask for a dynamic graphical element. In order to test this mechanism we used a circle superimposed on the recording using the gaze marking BeGaze software." /></td>
<td>A frame that is selected based on a unique color is a mask for a dynamic graphical element. In order to test this mechanism we used a circle superimposed on the recording using the gaze marking BeGaze software.</td>
</tr>
<tr>
<td></td>
<td><img src="image3.png" alt="Summarized static (for the entire recording or a set records acquired from a single source) and dynamic mask (prepared for the current frame)." /></td>
<td>Summarized static (for the entire recording or a set records acquired from a single source) and dynamic mask (prepared for the current frame).</td>
</tr>
<tr>
<td></td>
<td><img src="image4.png" alt="A mask with superimposed dilation filter [15]. Use of this filter prevents from projecting the mask onto the original picture, which is why removed elements remain visible." /></td>
<td>A mask with superimposed dilation filter [15]. Use of this filter prevents from projecting the mask onto the original picture, which is why removed elements remain visible.</td>
</tr>
<tr>
<td></td>
<td><img src="image5.png" alt="Final mask intended for removal (after binarization)." /></td>
<td>Final mask intended for removal (after binarization).</td>
</tr>
</tbody>
</table>
be repeated for each film frame since the $X$ vector is different for each one of them (image of the simulator monitor changes position on the recording). Subsequently, using a set of time-dependent transformation matrices obtained in such a manner, based on linear interpolation of determined matrices we transform the gaze coordinates written in the IDF file into gaze coordinates in the flight simulator monitor coordinate system. That enables gaze analysis relative to a flight simulator monitor, which stays still on the recording.

The above considerations are somewhat broader than those necessary to achieve transformation in the particular case we are dealing with. In fact, we assume that the transformed rectangle containing the simulator monitor visible on the recording maintains its shape – it therefore assumes rotation, scaling and translation. And in such a case, matrix $A$ elements are equal to:

$$A = \begin{pmatrix} a_{xx} & a_{xy} \\ a_{yx} & a_{yy} \end{pmatrix} = \begin{pmatrix} \cos(\alpha) & \sin(\alpha) \\ -\sin(\alpha) & \cos(\alpha) \end{pmatrix} \begin{pmatrix} s \cos(\alpha) & s \sin(\alpha) \\ -s \sin(\alpha) & s \cos(\alpha) \end{pmatrix}$$

(8)

Where $s$ is a scaling coefficient and $\alpha$ is the angle of rotation. That means in particular that $a_{xx} = a_{yy}$ and $a_{yx} = a_{xy}$, which lowers the number of sought parameters down to four ($a_{xx}$, $a_{xy}$, $t_x$ and $t_y$).

Therefore, in order to determine the relevant affine transformation we only need the coordinates of two points before and after transformation.

A simple test that checks program functioning, particularly the consistency of transformation of
Comparisons were conducted through a zero-one assessment of congruency of gaze location in the region of interest (navigation indicators, plane status indicators, outside view) before and after transformation (Table 3).

Assessment of the congruency of gaze positioning in the areas of interest before and after transformation using EVM software was confirmed in 100% of analyzed cases. Full congruency was due to the preciseness of mathematical transformations and did not require statistical analysis. Using the EVM software we obtained results comparable with those acquired from SGM in BeGaze in a shorter period of time. Transformation of the data (text files containing gaze positions as well as video material with gaze position superimposed on the simulator monitor view) lasts about an hour for a recording of a several dozen minutes long flight, but it’s an automated process. Manual transfer of fixation points using the SGM method for an analogous film takes about four hours for a 44-minute recording and requires full focus when adjusting gaze position from the video recording to the reference frames.

Software tests

Eye movements were registered using a head-mounted eye tracker (SMI glasses) by SMI. As a result, we obtained a recording of gaze position (.idf file) and a recording of gaze position superimposed on the image from the camera at the scene (.avi).

RESULTS

Analyses

The purpose of validation of the method was to compare the gaze positions (from the recordings obtained during simulated flight) from 200 randomly chosen samples before and after transformation using EVM software. Ten areas of interest were created (left column, Table 3). Every gaze position fell onto a particular area. Comparisons were conducted through a zero-one assessment of congruency of gaze location in the region of interest (navigation indicators, plane status indicators, outside view) before and after transformation (Table 3).

Assessment of the congruency of gaze positioning in the areas of interest before and after transformation using EVM software was confirmed in 100% of analyzed cases. Full congruency was due to the preciseness of mathematical transformations and did not require statistical analysis. Using the EVM software we obtained results comparable with those acquired from SGM in BeGaze in a shorter period of time. Transformation of the data (text files containing gaze positions as well as video material with gaze position superimposed on the simulator monitor view) lasts about an hour for a recording of a several dozen minutes long flight, but it’s an automated process. Manual transfer of fixation points using the SGM method for an analogous film takes about four hours for a 44-minute recording and requires full focus when adjusting gaze position from the video recording to the reference frames.

Our software, as well as SGM by SMI, is meant for preparing oculomotor data for further statistical analyses. After stabilizing the position and size of the screen image it is possible to conduct...
automated analyses on unmoving regions of interest (times of fixation in the regions of interest, number of passes between the regions, number of returns) and subsequently export the data prepared in such manner to the statistical software.

**CONCLUSIONS**

Use of the EVM software produces a recording with stable position of flight simulator monitor with superimposed position of gaze fixation as well as a file containing the gaze data transformed to new positions. Stable regions of interest (navigation, plane status indicators and the outside view) may be superimposed on such data sets and further comparisons may be conducted on the collected data.

The proposed solution implemented in the EVM software enables automation the transformation of oculomotor data registered by a head-set eye tracker. It allows faster data processing for further analyses using regions of interest.

It is worth noting that the program may be used not only in flight simulators, but also in other kinds of simulators (road vehicles, forklifts, but also air traffic controllers) as well as operator’s natural or near-natural working environment.

**AUTHORS’ DECLARATION:**

*Study Design:* Balaj Bibianna, Piotr Francuz; *Data Collection:* Balaj Bibianna; *Data Analysis:* Balaj Bibianna, Sternal Marek, Matulewski Jacek; *Manuscript Preparation:* Balaj Bibianna, Matulewski Jacek, Sternal Marek, Piotr Francuz; *Funds Collection:* Piotr Francuz, Balaj Bibianna; *Implementation in the original software package:* Sternal Marek, Matulewski Jacek. The Authors declare that there is no conflict of interest.

**REFERENCES**


ACKNOWLEDGEMENTS

The views, opinions, and findings contained in this article are our own and should not be construed as an official Polish Air Force position, policy, or decision, unless so designated by other official documentation.