Simple and Precise Commercial Camera based Eye Tracking Methodology

Vladislava Bobić and Stevica Graovac

Abstract — In this paper, an eye tracking methodology is presented. It represents a commercial camera based system for detection of eye pupil and its trajectory and classification of eye movements that led to the formation of assessed eye path. Methodology was examined on 10 subjects, who performed specified series of eye movements in two recording conditions: with and without simultaneous head motions. Image segmentation was performed and coordinates of eye pupil were found for each consecutive video frame. Modified velocity threshold method for eye movement classification was performed. Classification accuracy was evaluated with two parameters: Sensitivity (S) and Positive predictive value (PP), and compared for both recording conditions. In the absence of head motions, a high accuracy result was obtained (S=96.25%; PP=100%). A less accurate result was obtained in the second case (S=87.50%; PP=88.61%). However, it was shown that proposed methodology can provide good results for commercial eye tracking regardless of the recording conditions.

Keywords — eye tracking, image gradient, pupil detection, velocity threshold.

I. INTRODUCTION

Eye Tracking Methodology

YE tracking represents detection of an eye position and its movements and has found the application in many academic and commercial research areas, including diagnosis of patients in ophthalmology, neurology and psychology, as well as market, usability, and gaming research, human computer interaction, etc. Because of that, it is of high importance to provide a system that will allow accurate monitoring of the eye and provide information about its trajectory, but also about the movements that led to the formation of such an eye path. The eye movements, which are commonly used for assessing subjects’ attention and abilities and that are considered in this paper, are fixations and saccades.

Paper received May 7, 2017; revised June 27, 2107; accepted July 25, 2017. Date of publication July 31, 2017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Branimir Reljin.

This paper is a revised and expanded version of the paper presented at the 24th Telecommunications Forum TELFOR 2016 [29].

This work was supported in part by the Serbian Ministry of Education, Science and Technological Development under Grant No. 175016.

Vladislava Bobić is PhD student with the School of Electrical Engineering, University of Belgrade, Bul. kralja Aleksandra 73, 11120 Belgrade, Serbia, and Innovation Center of School of Electrical Engineering, University of Belgrade, Bul. kralja Aleksandra 73, 11120 Belgrade, Serbia, (e-mail: vladislava.bobic@yahoo.com).

Stevica Graovac is with the School of Electrical Engineering, University of Belgrade, Bul. kralja Aleksandra 73, 11120 Belgrade, Serbia (e-mail: graovac@etf.rs).

Fixation represents the ability of eye to inhibit ocular shift by focusing to one point, i.e. the target. Gathering of visual information about the external environment is possible only during a stable fixation [1].

Saccade is a rapid change between two fixations within the field of view [2]. It represents a fast movement that allows eye to change a fixation from one point to another. They are characterized with a large initial acceleration and final slowdown. The amplitude of saccadic movements is usually expressed in degrees of visual angle [3]. A visual angle is formed between two lines connecting the center of the eye and two consecutive eye targets. In general, this eye movement has an amplitude within the range of 1 - 40°, whereby the tracking of moving targets exceeding 30° must include corresponding head movements. The duration of saccade may vary from 10 to 120ms [4]. Velocity of saccadic movement is characterized by the peak whose maximum value depends on the amplitude of saccades and it can reach values up to 400-600°/s.

The absence of fixation can indicate some diseases, primarily the occurrence of autism [5]. Babies at the age of 2-6 months, disabled to perform fixation were later diagnosed with autism [6]. Some studies have shown that patients with an atypical form of Parkinsonism have reduced velocity of saccades [7]. Additionally, eye tracking is used in the diagnosis of several other diseases such as: traumatic brain injury [8], schizophrenia [9], Alzheimer's disease [10], a functional loss associated with macular degeneration [11], and Meniere's disease [12]. Because of that, monitoring of eye condition can be very useful in the diagnostics of the most common and most problematic neurological or psychological disorders nowadays.

Until now, several systems have been created for eye tracking in clinical settings. The simplest of them is certainly a visual inspection by the physician. More sophisticated methods are also applied, such as video nystagmography [13], electrooculography [14] and computerized eye tracking devices that work in the infrared or visible part of the spectrum [15], [16]. All clinical systems are characterized by a high sampling frequency and high image resolution as well as by the use of specific stimulating systems. Therefore, they are unavailable for home assessment (primarily because of their price).

In order to improve their visual marketing and make it more effective, companies want to track and obtain information about consumers’ partiality, and affection. Because of that, eye tracking has become a significant part of visual marketing in USA and Europe where obtained spatiotemporal measures indicate attention processes of prime interest [17]. Gaming, as one of the fastest growing industries, has effectively included eye tracking as one of
the interactive control methods [18]. Different systems are used for such purposes, including infrared corneal reflection methodology [19], modern table-mounted devices, stereo imaging [20], combinations of wide-angle and zoom cameras [21], as well as a variety of other eye-tracking systems, with different calibration times, and natural exposure conditions.

In the field of eye pupil detection, researchers have presented different approaches that are based on the analysis of: only one eye [16], both eyes independently [22, 23] and both eyes simultaneously [24]. Additionally, different methods for classification of eye movements are demonstrated as well [16], [25], [26]. Tools used for this application can be divided into several groups. One approach includes determination of the eye position distribution at certain points, i.e. Dispersion-based Algorithms. The most famous of them are: Dispersion-Threshold Identification (E-DT) and identification based on minimum spanning trees (I-MST). Classification of eye movements can be performed in certain regions of the eye as well, i.e. Area-based Algorithms [25].

However, the most commonly used method uses eye velocity (i.e. Velocity-based algorithms), calculated relatively to the stimulus of the eye, if the distance between the subjects and the stimulus is known, (e.g. Threshold Identification - I-VT [25]; Hidden Markov model identification - I-HMM [16]; Binocular-Individual Threshold – BJT [26]; Kalman Filter Identification - I-KF [27]). Depending on the chosen method, the classification criterion is selected and each image sample is classified as part of the fixation or saccade. The combined distribution function is assigned to successive samples that are arranged in the same group of movements, i.e. they are grouped into the fixation or saccades blocks. I-VT method is the simplest velocity based algorithm for classification of eye movements, where given criterion for comparison represents the velocity threshold [16].

Selecting an appropriate threshold is one of the most difficult tasks set in such research. It is shown that the accurate detection of saccades can be achieved with a threshold varying from 5°/s [28] to 300°/s [25].

In this paper, a new methodology for detection and tracking of eye pupil coordinates in offline recording conditions is presented. Video data was recorded with a commercial camera. By using modified I-VT algorithm, two types of eye movements are defined and detected. This paper represents expanded version of paper presented at the 24th Telecommunications Forum TELFOR 2016 [29]. The proposed methodology was examined for one additional recording condition - with simultaneous head motions of small range, which was not considered before. Since head motions can introduce both false positive and false negative saccades, the accuracy was evaluated with both Sensitivity and Positive predictive value. In this way, the methodology was examined in a more practical sense, since in real applications recordings without head motions are hardly achievable.

The proposed methodology provides eye tracking with good accuracy regardless of the presence of the head motions. It can be useful for some simple applications (such as marketing, gaming) which don’t require advanced eye tracking systems.

II. METHOD AND MATERIALS

A. Instrumentation

A commercial camera Nikon D3200 (Nikon Corporation, Tokyo, Japan) was used in this experiment. Utilized camera has a maximum video recording frequency of f_s= 60 fps (frames per seconds) or 60 Hz [25], and in that case image resolution of its frames is 1280x720. Applied camera operates in the visible part of the spectrum, so it was necessary to perform this experiment in a bright light. All recordings were performed in the same lightening conditions.

B. Experiment

In this study, ten healthy subjects (5 male/5 female) were included. Participants were chosen to be at various age (37.2 ± 19.1 years) and with different eye colour (5 with dark brown and 5 with light blue eye colour), so better examination of system performance could be achieved.

During the experiment, subjects were sitting comfortably in the chair and looking directly to the camera at the beginning of the recordings. The camera was placed at the distance of about 20 cm from the respondent’s face (optical zoom was set to 55 mm) on a still surface. In this way, video recording that shows both respondents’ eyes in focus was obtained. Participants were asked to perform a specified sequence of rapid eye movements with short fixations in between (eight saccadic movements with nine fixations in between, as it is shown in Fig. 1.). The recordings were repeated for two different conditions. Firstly, subjects were instructed not to move their head relatively to the camera. However, this is not natural for respondents and it is very difficult to achieve in practice, so it was necessary to repeat the procedure and examine effectiveness of designed system for video recordings with simultaneous head motions of small range (head is moving relatively to the camera but those motions do not remove eyes from video frames). For both recording conditions, the procedure was repeated until the subject has done the exact series of the given motions.

C. Signal processing

Signal processing was performed in custom-made Matlab code (Matlab 7.6.0., R2008). A new methodology for eye tracking was performed. It consists of two segments: 1) Pupil detection, 2) Classification of eye movements. For detection of eye centre coordinates, image processing tools were used and applied on a series of images (frames). Classification of eye movements was performed with modified I-VT method.

1) Pupil detection

In order to avoid potential errors that may occur as a result of a simultaneous analysis of both eyes, each consecutive frame was divided into two equal parts (left and right). In this way, the area of one eye with surrounding face regions was separated for further analysis. If the subject has a specific problem in which the movements of the left and right eye are not compliant, it is possible to perform an independent analysis of both image parts (both eyes).
Fig. 1. Visualization of performed sequence of eye fixations.

Fig. 2. Representation of image processing procedure for detection of eye pupil. Example is given for two subjects: one with light eye colour (upper row) and one with dark eye colour (lower row).

R component was extracted from the original RGB image (as shown in Fig. 2, the first column) and used in further analysis because within R component areas that correspond to the skin and eyeball regions have higher values (they are brighter) than the regions of eye pupils (and iris, if subject has a dark eye colour), eyelashes and eyebrows (colour also depends on subjects’ eye hair colour). Because of that, contrast between low and high intensity values was increased, and magnitude gradient of the adjusted image was found [30], by using the “Sobel” gradient operator, which gave good results for all included respondents.

Consequently, the magnitude of the gradient image displayed boundaries of pupil and other dark regions (eyebrows, eyelashes) with other image areas (the second column in Fig. 2). Afterwards, morphological processing (erosion and dilation) was applied: larger regions obtained filled holes and smoother borders, whereas some white regions with a smaller area were completely removed from the gradient image (the third column in Fig. 2). In this way, the unique white regions were obtained and later described with the following features:

- Area – an area of the region defined by the number of pixels within the region,
- Centroid – center of mass of the region, defined by the ordered pair of x and y coordinates of the point in which the center of mass is located.

For the first image (frame) of the video recording, in which the subject is looking directly to the camera, eye center or eye pupil is defined as the center of mass of the region with the largest area (as shown in the third and fourth columns in Fig. 2). Each succeeding frame is corrected according to the location of pupil from the previous frame that displays subject’s eyes as open (Fig. 3).

Specifically, for each frame a circle of specified diameter \( d \) was determined. Center of the defined circle is located at the point which represents the eye center in the first preceding frame that shows subject’s eyes as open (the procedure is shown in Fig. 3). This way, all regions that lie outside its boundaries and that can introduce errors are eliminated. For example, in Fig. 3, a string of hair fell into the frame (area with blue boundary shown in lower panel), so without the correction, errors may occur. From the regions isolated by the aforementioned criteria (located
within the circle area), the one with the largest area is chosen, and its center of mass is defined as the eye pupil center.

Simultaneously, the area of selected region is examined: if the region surface is smaller than a specified value $P$, the ordered pair of eye center coordinates obtained value $(x, y) = (0, 0)$. This way, the frames on which the eyelid is down, and eye center is not visible in its entirety (the eyes are closed or almost closed) are discarded from the analysis.

The values of parameters $d$ and $P$ can be adjusted according to the distance between the camera and the subject’s face and focus of his eyes. As a final result, ordered pairs of the pupil coordinates were obtained for each successive frame in a form of two vectors providing information about eye pupil trajectory during the video recording.

![Fig. 3. Correction of an eye center for two consecutive frames that show subject’s eyes as open. Smaller red circle marks an eye center detected in the frame shown on the upper panel, whereas the larger red circle represents correction circle. Region with blue boundary is a string of hair that fell into the frame and that could introduce error in the absence of the corresponding correction. Example is given for one subject.](image)

2) Classification of eye movements

Classification of eye movements was performed with somewhat modified I-VT algorithm. Since eye stimulation wasn’t used and distance between the face and the camera is not precisely known, previously mentioned method cannot be fully applied. Instead, the ordered pairs of eye center coordinates were placed in a coordinate system with origin located in the upper left corner of the image. The Cartesian coordinates of the eye pupil were transformed into polar coordinates $(x, y) \rightarrow (\rho, \theta)$. Therefore, the rates were determined by $\rho$ and $\theta$ coordinates, that is $\dot{\rho}$ and $\dot{\theta}$, respectively. Small variations of eye center positions can be obtained (due to image processing). Because of that, obtained velocities were filtered using a 5-point moving average filter. In this paper, thresholds were defined empirically: $p_\theta = 10 \, ^\circ/s$ for $\dot{\theta}$, and $p_\rho = 1 \, \text{cm/s}$ for $\dot{\rho}$. Each sample of calculated velocity was compared with the threshold: if a sample value was higher than the threshold value, the sample was defined as part of saccades, and vice versa (a sample value below the threshold was marked as a fixation part). The procedure was repeated for both velocities and positive sample results for saccadic movements were merged into one sequence. Successive samples, assigned to the same group of movements, were grouped into a fixation or saccades blocks. Fixations lasting less than 100 ms were discarded from the analysis [16].

Since no stimulation was applied, evaluation of classification that is usually applied (comparison between the number of fixation/saccades samples encoded in the stimulation, and the number of fixations/saccades samples that are detected [16]) cannot be used as a part of this methodology. In this paper, Sensitivity $S$ and Positive predictive value $(PP)$ were calculated. Sensitivity was obtained as the ratio of the number of correctly detected saccades (number of true positives) and the number of actual saccades that were asked from subject to perform, expressed as percentage. Positive predictive value was calculated as:

$$PP = \frac{TP}{TP + FP} \times 100.$$

Parameter TP represents the number of correctly detected saccades or number of true positives, whereas TF represents the number of saccades that were detected by the system but were not performed by subject (or the number of false positives). TP and TF were determined by visual inspection.

Parameters were calculated and compared for both recording conditions: 1) without simultaneous head motions (no relative moving between camera and head); 2) with simultaneous head motions (head is moving relatively to the camera).

This method of evaluation is rough, but it does not depend on subject’s answer to the stimulus and it gives overall evaluation of system performance and movement classification.

III. RESULTS

Two calculated velocities with defined thresholds are presented in Fig. 4. Example is given for one subject.

In Table 1, the number of actual and the number of detected saccades are presented, as well as calculated Sensitivity of the methodology. The number of false positive saccadic movements and calculated Positive predictive value are presented in Table 1, as well. Results are summarized and presented for 10 subjects that were included in the experiment and for both recording conditions (with and without simultaneous head motions).
A new methodology for eye tracking is presented in this paper. The first part of the methodology is dedicated to detection of eye pupil coordinates. For this purpose, a custom-made algorithm that uses image processing tools has been implemented on the recorded video recordings. Based on the calculated eye pupil coordinates, classification of eye pupil movements on fixations or saccades was performed with modified I-VT algorithm.

Based on the calculated eye pupil coordinates, classification of eye pupil movements on fixations or saccades was performed with modified I-VT algorithm. For this purpose, a paper. The first part of the methodology is dedicated to detection of eye pupil coordinates. For this purpose, a custom-made algorithm that uses image processing tools has been implemented on the recorded video recordings. Based on the calculated eye pupil coordinates, classification of eye pupil movements on fixations or saccades was performed with modified I-VT algorithm.

Saccadic and fixation movements were detected based on the universal threshold, empirically determined for all included subjects \((p_r=1 \text{ cm/s} \text{ and } p_\theta=10 \text{ °/s})\). Evaluation of defined methodology and its accuracy was performed with two parameters (Sensitivity and Positive predictive value).

Results enclosed in Table 1 show that in the first case (without head motions) a high accuracy result was obtained (Sensitivity was 96.25% and Positive predictive value was 100%). However, for the second recording condition results were poorer but also high enough for some simple practical implementation (Sensitivity was 87.5% and Positive predictive value was 88.61%). Head motions that follow eye movement can decrease its velocity. Also, natural head movements between two saccades can change the position of eye pupil relatively to the camera and therefore introduce error.

The question that arises is whether it is really necessary to observe both velocities to detect saccadic movements? The question is hidden in Fig. 4 that displays velocity profiles for one subject. If only angular velocity was considered, then by adopting a universal threshold for classification of eye movements, the first two saccadic movements wouldn’t be detected (red rectangle in Fig. 4.). Also, some movements wouldn’t be detected by analysing only the changes per distance from the origin (green rectangle in Fig. 4.). By using this method, the influence of the chosen origin is eliminated, so changes of the eye position can be examined in terms of its angle and distance from the origin. In addition, the method does not depend on the choice and the presence of the stimulus. For some practical application, calibration should be performed in advance.

The main imperfection of this system is a small recording frequency. With this sampling frequency, microsaccades of duration less than 17ms wouldn’t be detected, and for such use it is necessary to apply a camera with a higher sampling frequency. Besides that, deficiencies of this system include required bright light and necessity of fixating head relatively to the camera for a high accuracy result.

Despite those imperfections, this system allows touchless recording of eye trajectory and classification of its movements. It is a simple and accurate system which doesn’t require expensive equipment and it can be used in offline, but also online recording conditions. The used camera can be replaced with any other camera with similar recording options and characteristics. Also, defined system can be integrated with some specific stimulators.

Future work includes methodology enhancements for detection of eye movements with simultaneous head motions, including elimination of head motions effects and consequently achievement of better classification accuracy and a smaller number of false positive results. The defined methodology will be additionally tested on a larger group of subjects of various age and eye colour, by using the same parameters for detection of eye pupil and thresholds for classification of eye movements, with and without simultaneous head motions.
REFERENCES


