ARTICLE IN PRESS
JID: PATREC [m5G;July 15, 2015;21:22]
Pattern Recognition Letters 000 (2015) 1–6
Contents lists available at ScienceDirect
Pattern Recognition Letters
journal homepage: www.elsevier.com/locate/patrec

Eye movements during scene understanding for biometric identification ☆

Usman Saeed*

University of Jeddah, Jeddah 21589, Saudi Arabia

ARTICLE INFO

Article history:
Available online xxx

Keywords:
Biometric identification
Eye movement
Scene understanding

ABSTRACT

The human eye is rich in physical and behavioral attributes that can be used for biometric identification. Eye movement is a behavioral attribute that can be collected non-intrusively for biometric identification. Usually a task oriented visual stimulus is presented to the subject and his eyes are tracked using a video camera, which are then used for biometric identification. The most common visual stimulus employed includes the moving object and free viewing. In this paper I have experimented with a novel task oriented visual stimulus i.e. scene understanding. In scene understanding the observers are instructed beforehand that they must perform a task based on the contents of the image/video that will be presented. A biometric identification system has been developed based on the eye movements extracted during scene understanding. A compact and easy to extract feature vector based on clustering of eye movements has been proposed and tested using several publicly available databases and two classification schemes. The results presented in this paper with a correct identification rate of 85.72% are quite promising. Furthermore, I also provide comparative results by implementing three commonly used feature vectors for eye movements.

1. Introduction

The human eye provides several measures that can be used for biometric identification, the most performant amongst them being physiological such as iris and retina patterns. However, the eye also exhibits a strong behavioral component visible in the form of eye movements. Eye movements depict the viewing behavior of a person i.e. what and how a person chooses to see a visual scene. Although the structure of the eye has an influence on the eye movements, but the features of the visual scene the person fixates on and the skills and techniques one uses to gather information is mainly a behavioral attribute. Eye movements for biometric identification are typically extracted by presenting a visual stimulus and tracking the eye using a video camera. This setup is contactless and non-intrusive, thus ideal for continuous authentication.

The movements of the eye are mainly classified into four categories: saccades, smooth pursuit, vergence, and vestibulo–ocular movements [21]. Saccades are rapid motion of the eyes that are used to quickly change the fixation point (the point of interest at which the eye is relatively stable). Saccades range in scale from small movements used during reading, to much larger movements required for finding an object in a scene. Smooth pursuit movements are slower than saccades and are used to keep the fovea aligned with the moving stimulus. Vergence movements align the fovea of each eye with targets located at different distances from the observer. Vestibulo–ocular movements are used to compensate for head movements, thus stabilize the eyes relative to the world. The nature of the task being performed determines the type of eye movements produced. In eye movement research the tasks are generally classified as free viewing, reading, object pursuit and scene understanding. Therefore I shall present the literature review of biometric identification techniques categorized using the task employed to elicit the eye movements.

During free viewing eye movements are highly dependent on the nature of the stimulus. Free viewing of image and video has been employed to gather eye movements which are then used for biometric identification. In case of image based stimulus [6,16,23], fixations are the most important whereas in video stimulus [14], fixations, saccades and smooth pursuit movements have been found significant.

Human eye movements during reading [22], exhibit certain distinctive characteristics such as the average fixation duration is about 225–250 ms and the average saccade is 8–9 character spaces. Another important characteristic called regression is that 10–15% of the time readers voluntarily move their eyes to material that they have already read before. For biometric identification the subjects are required to read a specific text and eye movements are tracked as in Holland and Komogortsev [9].

Human eyes mainly exhibit smooth pursuit movements while following moving objects, but for biometric identification [6,11,12,15]

☆ This paper has been recommended for acceptance by Gabriella Sanniti di Baja.
* Tel.: +966 599566058.
E-mail address: usaeed@kau.edu.sa, usman.m.saeed@gmail.com

http://dx.doi.org/10.1016/j.patrec.2015.06.019
0167-8655/

Please cite this article as: U. Saeed, Eye movements during scene understanding for biometric identification, Pattern Recognition Letters (2015), http://dx.doi.org/10.1016/j.patrec.2015.06.019

and [18] two different moving objects have been employed. The first type is moving smoothly on the screen while the second one is in the form of an object jumping from point to point in a uniform or random pattern. The eye movements that are most prominent in this category are smooth pursuit movements when the object is moving smoothly and saccades in the jumping object scenario.

Scene understanding is defined as gazing upon a visual stimulus to develop a basic understanding of the scene or to memorize the important features. The average fixation duration in scene understanding is longer than in reading, and the average saccade distance is also larger [22]. It has been found that the task also has a strong impact on eye movements during scene understanding e.g. memorizing the scene for a later test.

A comprehensive survey on the use of eye movements for biometric identification has been compiled by me in Saeed [24]. During the survey I was unable to find any study utilizing eye movements extracted from scene understanding based task for biometric identification.

Therefore in this paper I have employed scene understanding based visual stimulus and extracted the observer's eye movements for biometric identification. The observers in the databases utilized [8,20], were instructed beforehand that they must perform a task based on the contents of the image/video that will be presented to them. A biometric identification system has been developed, including pre-processing, feature extraction and classification. Several feature vectors based on clustering of eye movements have been extracted and tested using two different classification methods. The proposed system has been tested on three publicly available databases and the results presented in this paper with a correct identification rate of 85.72% are quite promising. Furthermore, three feature vectors commonly used for eye movements have also been implemented for comparison. Lastly I have also presented a comparative analysis of our results with some recent studies which have employed different task oriented visual stimuli.

The rest of the paper is divided as follows. In Section 2 the proposed method is elaborated. In Section 3 the experiments and the results obtained are described. Finally the paper concludes with the remarks and future works in Section 4.

2. Proposed method

In this section I shall present the steps of the proposed system. First, depending on the database employed various pre-processing steps are applied. Next, various features pertinent to biometric identification are extracted and finally classification methods are applied for biometric identification.

2.1. Pre-processing

The databases used in our study have been designed for human eye movement analysis under cognitive load, whereas the objective of this study is biometric identification using eye movements. Therefore certain pre-processing steps are required, which are described below.

- Some observers in the databases were not instructed to perform the scene understanding task and were free to view as they please. These observers were removed as they were not relevant to our study.
- The databases were arranged according to the objective of their designers, which was mostly inclined towards eye movement analysis. Therefore the databases were rearranged to separate the videos according to the observer's identity.
- Next, data such as time stamp, pupil diameter etc. which is not employed for biometric identification in this study was removed. Leaving behind only *XY* coordinates of eye movements and parameters identifying dominant eye, fixations/saccades.
- Furthermore, only the *XY* coordinates of the dominant eye are used to reduce data complexity.
- In some of the databases employed the parameter identifying the eye movement as a fixation or saccade is present, this parameter is retained and used in the feature extraction phase. However in other databases, this parameter is not available, therefore the algorithm developed by Krassanakis et al. [17] was employed to identify fixations.

2.2. Proposed features

The pre-processing steps results in the *XY* coordinates of the dominant eye movements and a parameter signifying whether the current eye coordinate is a fixation or a saccade. Using this information each eye tracking recording is represented by the raw *XY* coordinates of the eye movements, including both fixation and saccades, given by V_{Rfs} in Eq. (1).

$$V_{Rfs} = [x_1, y_1, x_2, y_2, \ldots, x_n, y_n] \quad (1)$$

Next another vector V_{Rf} given by Eq. (2) was created which contains only the raw *XY* coordinates of fixation points.

$$V_{Rf} = \left[x_{f1}, y_{f1}, x_{f2}, y_{f2}, \ldots, x_{fn}, y_{fn}\right] \quad (2)$$

Next several feature vectors were extracted which are described below.

(1) Clustered *XY* coordinates

It was observed that the raw *XY* coordinates contain a large amount of redundant information due to the fixations, when the eye does not move. Therefore k-mean clustering [19] was applied to reduce this redundancy by using only cluster centroids.

The first feature vector V_{Cfs} was created by applying clustering to the raw *XY* coordinates vector V_{Rfs} consisting of both fixations and saccades. Thus V_{Cfs} given by Eq. (3) contains the *XY* coordinates of the cluster centroids consisting of both fixation and saccades.

$$V_{Cfs} = [x_{c1}, y_{c1}, x_{c2}, y_{c2}, \ldots, x_{cn}, y_{cn}] \quad (3)$$

The second feature vector V_{Cf} was created by applying clustering to the raw *XY* coordinates vector V_{Rf} consisting of only fixations. Thus V_{Cf} given by Eq. (4) contains the *XY* coordinates of the cluster centroids consisting of only fixation.

$$V_{Cf} = \left[x_{cf1}, y_{cf1}, x_{cf2}, y_{cf2}, \ldots, x_{cfn}, y_{cfn}\right] \quad (4)$$

(2) Clustered *XY* coordinates with time parameter

One of the side effects of clustering which was observed was that repeat fixations (when the observer returns to view the same point after some time) were combined into a single cluster. To avoid this situation and retain the independence of repeat fixations a time parameter t was added to the raw *XY* coordinates vector V_{Rfs} and V_{Rf} to create new vectors V_{Rfst} and V_{Rft} given by Eqs. (5) and (6) respectively.

$$V_{Rfst} = [x_1, y_1, t_1, x_2, y_2, t_2, \ldots, x_n, y_n, t_n] \quad (5)$$

where $\{t \in \mathbb{Z}; t > 0\}$

$$V_{Rft} = \left[x_{f1}, y_{f1}, t_{f1}, x_{f2}, y_{f2}, t_{f2}, \ldots, x_{fn}, y_{fn}, t_{fn}\right] \quad (6)$$

where $\{t \in \mathbb{Z}; t > 0\}$.

Next clustering was applied and two feature vectors were extracted. The first feature vector V_{Cfst} as in Eq. (7), was created by applying clustering to V_{Rfst} and it contains the *XY* coordinates of the cluster centroids containing both fixations and saccades and the time parameter.

$$V_{Cfst} = [x_{c1}, y_{c1}, t_{c1}, x_{c2}, y_{c2}, t_{c2}, \ldots, x_{cn}, y_{cn}, t_{cn}] \quad (7)$$

The second feature vector V_{Cft} as in Eq. (8), was created by applying clustering to V_{Rft} and it contains the *XY* coordinates of the cluster centroids containing only fixations and the time parameter.

$$V_{Cft} = \left[x_{cf1}, y_{cf1}, t_{cf1}, x_{cf2}, y_{cf2}, t_{cf2}, \ldots, x_{cfn}, y_{cfn}, t_{cfn}\right] \quad (8)$$