User Identification Based on Eye Gaze Data

Mohammad Reza Mahmoudian Motlagh
Gjøvik University College
Gjøvik, Norway
mohammad.mothlagh@hig.no

Patrick Bours
Gjøvik University College
Gjøvik, Norway
patrick.bours@hig.no

Abstract

In this paper, we are going to introduce two new definitions of fixations and saccades in eye gaze. Then we will investigate their usability in eye movement identification. Fixations and saccades are the two important features of eye gaze. Most researches in eye movement identification and authentication are performed based on the features extracted from them. We shall investigate to see whether these definitions of fixations and saccades can improve the previously obtained results in identification and authentication. The performance results obtained showed that the fixations are better in distinguishing different eye movement behavior than the defined saccades.

Keywords-Eye Movement, Identification, Fixations, Saccades, Biometrics

1 Introduction

Biometric recognition is a fast growing trend in authentication and identification. Biometrics' popularity among different security providers and other organizations is growing rapidly. Biometric recognition by eye movements is one of the biometric modalities that is becoming more and more popular due to the fact that this biometric modality has very good fraud resistant capabilities since it is difficult to forge. Eye movement recognition is simply the recognition of identities by the way their eyes move. It can be classified as both behavioral i.e. unique ways people move their eyes and physiological i.e. the structure of muscles in oculomotor system[11]. There are two features, that are most commonly used in eye movement recognition methods i.e. fixations and saccades. A fixation is defined as the moment when the gaze is almost still and focused at some spot and saccade is defined as the moment when the eyes are moving between the fixation points. In most research, the raw eye signals are either categorized as fixations or saccades.

This paper was presented at the NIK-2014 conference; see http://www.nik.no/.

In our paper we have come with another definitions for fixations and saccades which are derived from the fixations and saccades respectively. Compared to the above definitions fixations and saccades, here fixations are defined as aggregation of points where boundaries of this aggregation are obtained based on the deviation from a center of the fixation. The center is the middle point of this aggregation. We will try to identify users through the new derived definitions.

As a quick guide to the next sections of this paper, in section 2 we shall have a review on current state of the art. Next, in section 3 we will explain the dataset we used for the analysis. Section 4 will describe the analysis. In this section, we will explain the features extraction process and the features we used, the template creation process and the distance metrics we used. In section 5, we discuss the result of the analysis and in the end we will discuss the conclusion and future work.

2 Related Work

The research on eye movement Biometrics has produced significant results, indicating a great future for further development. One of the earliest publications on application of eye movement for identification was a paper by Kasprowski and Ober [6]. They examined eye movement signals in order to find out if it is possible to use eye movement as a biometric identification method. Eye movement data signals were collected from participants where they had to follow a jumping point on the screen. From the data signals they extracted cepstrum features and applied classification algorithms such as k-NN, SVM, C4.5 decision trees, and Naive Bayes theorem [6]. They obtained the best performance with 3-NN with FAR ≈ 1.5 and FRR ≈ 22.

Most researches in eye movement biometric identification focus on fixation based [10] or saccade based features [8, 14] or both [5, 7, 13].

Rigas *et al.*[10] used only fixation data for identification. They collected the data where participants observed static face images. Features extracted were actually the (x, y) coordinates of the eye gaze. They however used only (x, y) coordinates related to fixation points. They created minimum spanning trees out of the fixation data and used Wald-Wolfowitz [3] method to compare these spanning trees. Support Vector Machine (SVM) [2] and k-Nearest Neighbors (k-NN) classifiers were used to obtain the best performance with 70% identification rate and an EER of 30%. Komogortsev *et al.*[8] considered only saccade features obtained from the raw data signal. They tried to model the eye muscles structure called oculomotor plant. For data collection they used the jumping point stimulation. Using the extracted saccade feature information, they found the oculomotor plant properties of each subject for identification. They obtained an FMR of 5.4% and an FNMR of 56.6% with 5-NN and an FMR of 80% and zero FNMR with C4.5. Similarly, Zhang and Juhola [14] used only saccade based features such as maximum velocity, acceleration, and amplitude and applied data mining algorithms. Like in [8], they also used jumping point stimulation. From the raw eye signals, high amplitude saccades were selected. Then their beginning and ending points were identified and their features were extracted. Using multiple classification methods such as SVM, Multilayer Perceptron (MLP) networks [4] and Logistic Discriminant Analysis (LogDA), the test data was classified into two classes, users and non-users and comparison was performed. The best result was accuracy rate of 89.9%.

Holland and Komogortsev [5] used scan path i.e. a sequence of fixations and saccades for identification while reading text. Based on the features specific to the fixations and saccades, i.e. fixation count and average duration, average saccade amplitudes and velocities etc. they obtained features such as scan path length, scan path area, and inflection points. By fusion of all scan path features they obtained an EER of 27%. Recently, with a similar methodology Tripathy *et al.*[13] used scan path features to obtain oculomotion characteristics of the eyes. From the scan path data they obtained the oculomotion matrices and by their novel hybrid Intelligent model(HIM) they achieved a 63% identification accuracy rate. Both fixation and saccade data were employed by Kinnunen *et al.*[7] with the purpose of "task independent authentication" where the training and the test data were "arbitrary". They collected eye movement tracking data when subjects were watching a movie and used different timing periods of collected data for training and testing. The lowest EER obtained was approximately 29% using an adapted Universal Background Model [9].

Another work that utilized both fixation and saccade features was by Silver and Biggs [12]. They introduced a multimodal biometric system by combining keystroke dynamics and eye movement Biometrics. The eye movement features used were mostly fixation based such as fixation duration and specific fixations but also two saccade based features, i.e average duration and velocity. Their experiment consisted of reading a text on the screen and typing it on the keyboard. Using probabilistic neural networks classification algorithm they obtained a true positive proportion of 66% when considering only eye movement and 70% when combining both keystroke and eye movements.

3 Data Set

The dataset is the training set of EMVIC 2014 (Second Eye Movements Verification and Identification Competition[1]), made available as a competition related to the International Joint Conference on Biometrics[2]. The dataset contains recorded eye movements of 34 participants while they observed images of faces. For each image of a face, a participant had to decide if he knew this face or not and the decisions are part of the data set. The dataset contained 837 observations where the number of observations per participant was either 24 or 27.

A participant was not restricted in time when observing a face and therefore do observations vary in length, ranging between 0.891 and 22.012 seconds. For each observation the identity of the participant was recorded ($s01$ to $s34$) and the decision of the participant if he knew the face or not was recorded as well (either false or true). The remainder of data related to each observation is alternating x_i and y_i values where (x_i, y_i) represents the position of the i^{th} observation. Because the middle point of the image was chosen as the origin, can both x_i and y_i values be positive and negative. The used equipment has a sampling frequency of 1000 samples per second, i.e. the number of recorded samples per observation range between 891 and 22012.

An example of an observation is given in Figure 1. In this example one can almost clearly see the fixations of the participant(the concentrated points), as well as the saccades between fixations.The fixations can have various shapes, like round, but also long stretched. Also it is clear that in some cases the participant moves more or less in a straight line from one fixation to the next, while in other cases he moves his eyes in a more undetermined manner until he reaches the next fixation.

[1]http://www.kasprowski.pl/emvic/index.php

[2]http://ijcb2014.org/index.html

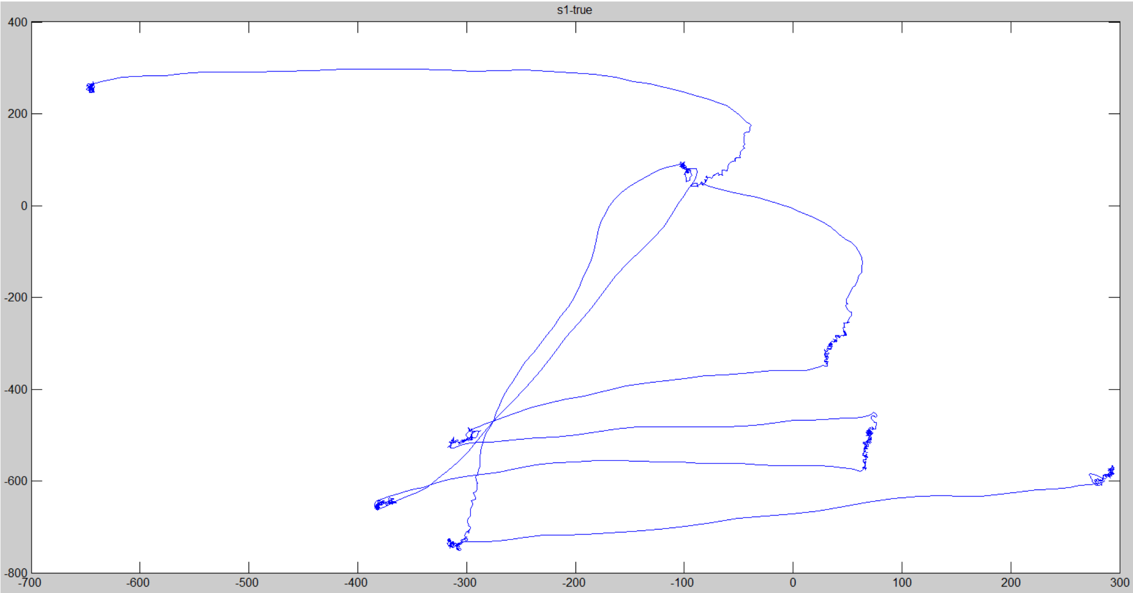

Figure 1: Example of observations

4 Analysis

Feature Selection

From the raw eye movement (x,u) data we extracted information on fixations and saccade paths. Fixations are defined as set of points that are very close to each other based on a fixed deviation. The movement paths are defined as the set of points which do not satisfy the fixation deviation criteria but they are also close to each other with based on a defined distance threshold. To find fixations, we considered m consecutive points and took their mean(μ) as the center of the fixation and also their and standard deviation(σ). If σ was less than a threshold t the set of m points would be considered a fixation. Then, the next successive points were considered to be 'inside' the fixations if the distance between them was below $\mu \pm C \times \sigma$, where C is a constant coefficient for σ. All such points would form a fixation along with the first m points. Thus, the minimum number of points in a fixation was m.

We extracted different features related to a fixation. They were, starting and ending coordinates, the time spent at a fixation which can be denoted by the number of points since the sampling rate used for recording the data was 1 kHz. Other fixation features were, center coordinates and maximum deviation from the center in order to estimate the fixation boundaries, distance between the farthest point to the center and total length of micro movements within a fixation and finally, average of distances of all points within a fixation to its center.

The different features related to saccade paths that we used were, number of points in a saccade which again denotes the time spent at saccade, actual displacement. There were 4 features related to μ and σ of two types of deviations from the shortest distance i.e. a straight line between the beginning and ending of the saccade. One type of deviation was simply the average of shortest distances of all the points in that particular saccade from the straight line joining the two end points of saccade. The other type was the average shortest distance of each point from the straight line connecting the previous point to the ending of the saccade.

Template Creation

From the extracted features, we first created a signature for each image data (see Table 1) by averaging fixation and saccade path features and adding some global features related to the whole image data. The 2 global features we used were the total time spent looking at an image and the starting zone. We divided the 2-D graph of gaze data into 4 zones. For zone 1, both *x* and *y* value were positive, while zone 2 indicated a positive *y* and a negative *x* value. In zone 3 both *x* and *y* values are negative and finally in zone 4 the *x* values are positive and *y* values negative.

For each user we created 2 templates: one related to the false decisions (not a known face) and the other related to the true decisions (recognition of a known face).

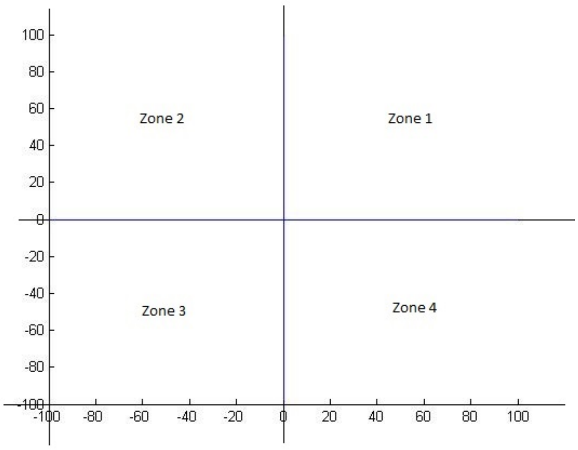

Figure 2: 2-D graph divided into 4 zones

For each user we created 2 templates: one related to the false decisions (not a known face) and the other related to the true decisions (recognition of a known face).

	Feature	Description
1	Zone	Starting zone
2	Total Time	Total gaze time
3	Total fixations	Total number of fixations
4	Time at fixation	Avg. time at an fixation
5	Smallest fixation	Quickest fixation
6	Largest fixation	Longest fixation
7	fixation length	Avg. fixation length
8	fixation distance	Avg. center to center distance
9	Micro moves	Avg. length of micro movements
10	saccade length	Avg. saccade length
11	saccade speed	Avg. saccade speed
12	Deviation 1	Deviation type 1
13	Deviation 2	Deviation type 2

Table 1: A picture signature

Distance Calculation

To calculate the distance scores, we used 3 metrics: (1) Manhattan distance; (2) Euclidean distance; and (3) Minkowski distance with $p = 3$. Let

$$Temp = (f_1, f_2, f_3..., f_n)$$

denote the template and

$$Test = (f_1', f_2', f_3'..., f_n')$$

denote test data where n is the number of features per user, then the Minkowski distance between $Temp$ and $Test$ is :

$$dist_{Minkowski}(Temp, Test) = \sqrt[p]{(\sum_{i=1}^{n} |f_i - f_i'|^p}$$

In fact, Manhattan and Euclidean distance are special cases of the Minkowski distance for $p = 1$ and $p = 2$.

In case of starting zone the distance is based on the difference between the zone (see Table 2). If the zones of the test and template matched then the distance was 0 and if the zones difference was 1 or 3 i.e. adjacent zones then the distance was set to 0.5. Finally if the difference was 2 i.e. opposite zones then the distance would be 1.

Starting Zones	Distance
Same	0
Adjacent	0.5
Opposite	1

Table 2: Distance calculation based on the starting zones

We split all data in two sets, those related to not recognizing the face on the image and those related to recognizing the face. Let $d_{i,j}^k$ denote the feature set of the i^{th} dataset if user j where the decision was k. Here k is either 0 or 1, j ranges from 1 to 34 and i ranges between 1 and n_j^k which denotes the number of datasets of user j with decision k. We compare each $d_{i,j}^k$ to all other feature sets with the same decision, i.e. the same value of k.

Before we applied the distance metric we first normalized all features, except the starting zone feature, to the $[0,1]$ range. The normalization is based on all feature sets of that specific user. For example if we compare $d_{i,j}^k$ to $d_{p,q}^k$ then both feature sets are first normalized based on the values of $d_{r,q}^k$ where r is between 1 and n_q^k.

Each feature set $d_{i,j}^k$ is compared to all other feature sets $d_{p,q}^k$ using the various distance metrics mentioned above [1]. We denote the distance between $d_{i,j}^k$ and $d_{p,q}^k$ by $dist_{i,j}(p,q)$. In our analysis we applied 4 different classification techniques. For a given test input $d_{i,j}^k$ we applied k-NN with both $k = 3$ and $k = 7$, when comparing the test input with all other feature sets. Besides those we also looked at the minimal distance, i.e. $minarg_q(dist_{i,j}(p,q))$. Finally for the last classification technique we averaged the distances over all distances of a user, i.e.

$$d_q = \frac{1}{n_q^k} \sum_{p=1}^{n_q^k} dist_{i,j}(p,q),$$

and the classification was based on the least average distance, i.e. $minarg_q(dist_q)$.

5 Results

We first evaluated the performance on the baseline configuration. We used all the features, the three different distance metrics and $m = 20$. Evaluated performance shows that among the three distance metrics, Manhattan performs better than Minkowski and has a slightly better performance than Euclidean distance (see Table 3). We decided to use Manhattan for the remainder of our analysis as it is easy to calculate and the performance was at least as good as the other distance metrics. Results based on the classification methods shows that both Nearest Neighbor classifiers performed similar to each other. We see from the table that both minimum and average distance did not perform as well as k-NN. For both minimum and average distance we see that the Manhattan distance outperformed the other two distance metrics. The best overall performance was achieved by 3-NN with an accuracy of 24% using Manhattan distance.

	AvgDist	MinDist	3-NN	7-NN
Manhattan	20.22	20.57	24.03	23.03
Euclidean	17.46	16.27	22.37	21.05
Minkowski	17.82	16.63	21.05	19.62

Table 3: Baseline with all features

Extended Analysis

The analysis was extended to different minimum fixation sizes and different feature combinations. We first tested on $m = 40$ and $m = 50$ to see if changes in the minimum of fixation size improves the performance. Again all the features were taken into account (see Table 4). Increasing the size would cause larger fixations. Consequently, change in size of m results in changing the initial characteristics of some fixations and saccades compared to when $m = 20$. This however, didn't affect the previously identified fixations where the number of points was more than 40. The performance increased significantly when changing the value of m from 20 to 40 and but the results for $m = 40$ and $m = 50$ are close together.

As mentioned before, both minimum distance and average distance methods improved when increasing the size of m but still there was a slight difference compared to k-NN. The best performance was achieved using $m = 40$ and 7-NN with an accuracy rate of 33.3%.

m	AvgDist	MinDist	3-NN	7-NN
20	20.22	20.57	24.03	23.03
40	27.75	32.54	32.54	33.25
50	26.94	31.90	32.85	32.04

Table 4: Results based on different minimum fixation sizes

One of the reasons we get a better performance for $m = 40$ compared to $m = 20$ might be that fixations play a more important role in distinguishing between different users. To investigate this, have we varied the features used in the comparison. We calculated the performance using different combinations of features when $m = 40$. Among those combinations, five are presented here. We first tried to test the performance using only fixations or only saccades. The performance rate dropped to approximately 10% when using only saccade features. On the other hand the accuracy rate was between 22% and 25% when using only fixation related features. A reason for the difference in performance could be that the number of saccade related features is almost half of the number of fixation related features. Another reason could be the similarity of saccade trajectories between users or the high variability in saccade of a single user.

Next, we combined fixations and saccade features with global features. It was observed that the performance improved when global features were included. From the two global features i.e. total time of gaze and the starting zone, the latter was more promising since, we also tested with the starting zone alone and the difference was not large. This shows that as a 'habit', the starting position of a gaze could be a stable feature for a person. As it was expected, fixations performed better than saccades when combined with global features.

Finally, we also combined the fixations and saccade related features. There was not a slight difference between the performance of fixation related features combined with saccade related features and fixation related features combined with global features. The results (see Table 5) show that although, saccade related features by themselves do not perform very well, they do improve the performance when combined with other features. Overall we saw that reducing the number of features did not raise the accuracy rate and the best performance remained to 33.3%.

	Avgdist	Mindist	3-NN	7-NN
All Features	27.75	32.54	32.54	33.25
Fixations	21.77	23.44	23.92	25
Saccades	10.17	9.09	8.97	10.29
Fixations+Global	23.21	26.08	27.27	29.19
Saccades+Global	12.18	16.27	15.67	14.71
Fixations+Saccades	25.24	29.90	30.26	29.19

Table 5: Results based on different feature settings

6 Conclusions and Further Research

We introduced two definitions of fixations and saccades. We evaluated the performance using various combinations of features, extracted from fixations and saccades. Best performance achieved was 33.3% with all features under consideration and minimum time spent at a fixation was 40. The result obtained is not significant since this is probably the first time these features are introduced. There are however, several ways that possibly could improve the performance which are explained below along with the identified issues. One way to improve the performance is to not create any image signature and compare all the features related to the fixations and saccades in both test and template. Because the signature is created by averaging related features of fixations and saccades. When taking average, some aspects and details are missed which may cause the missing of some properties of the features. Another way is to consider standard deviation along with the mean and use scaled distance metrics. Standard deviation shows that how much a user deviates from his/her estimated normal behavior i.e. the mean value. One way was to remove outliers in both fixations and saccades. We observed that the results obtained by inclusion of fixations is significantly better than those including saccades and not fixations. This would indicate that the most important part of fixations is the area around the borders with saccades where the x or y directions of the points is more similar to those of saccade points compared to the points closed to the center of fixations. However, the best performance also contained the saccades . This shows that although the saccades have a less determining role in identification compared to the boundary parts of the saccades, they can not be neglected. Another possible reason for the low performance could be the choice of global features that may not be appropriate to be used on the sequences that characterize eye movements. A possible solution could be employing more global features.

References

[1] R. Bolle, J. Connell, S. Pankanti, N. Ratha, and A. Senior. The relation between the roc curve and the cmc. In *Automatic Identification Advanced Technologies, 2005. Fourth IEEE Workshop on*, pages 15–20, Oct 2005.

[2] C.-C. Chang and C.-J. Lin. Libsvm: A library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, 2(3):27:1–27:27, May 2011.

[3] J. H. Friedman and L. C. Rafsky. Multivariate generalizations of the wald-wolfowitz and smirnov two-sample tests. *The Annals of Statistics*, 7(4):697–717, 07 1979.

[4] S. Haykin. *Neural Networks: A Comprehensive Foundation*. International edition. Prentice Hall International, 1999.

[5] C. Holland and O. Komogortsev. Biometric identification via eye movement scanpaths in reading. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–8, Oct 2011.

[6] P. Kasprowski and J. Ober. Eye movements in biometrics. In D. Maltoni and A. Jain, editors, *Biometric Authentication*, volume 3087 of *Lecture Notes in Computer Science*, pages 248–258. Springer Berlin Heidelberg, 2004.

[7] T. Kinnunen, F. Sedlak, and R. Bednarik. Towards task-independent person authentication using eye movement signals. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 187–190, New York, NY, USA, 2010. ACM.

[8] O. V. Komogortsev, S. Jayarathna, C. R. Aragon, and M. Mahmoud. Biometric identification via an oculomotor plant mathematical model. In *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ETRA '10, pages 57–60, New York, NY, USA, 2010. ACM.

[9] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn. Speaker verification using adapted gaussian mixture models. *Digital signal processing*, 10(1):19–41, 2000.

[10] I. Rigas, G. Economou, and S. Fotopoulos. Biometric identification based on the eye movements and graph matching techniques. *Pattern Recognition Letters*, 33(6):786 – 792, 2012.

[11] I. Rigas, G. Economou, and S. Fotopoulos. Human eye movements as a trait for biometrical identification. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pages 217–222, Sept 2012.

[12] D. L. Silver and A. Biggs. Keystroke and eye-tracking biometrics for user identification. In H. R. Arabnia, editor, *IC-AI*, pages 344–348. CSREA Press, 2006.

[13] B. Tripathi, V. Srivastava, and V. Pathak. Human recognition based on oculo-motion characteristics. In *AFRICON, 2013*, pages 1–5, Sept 2013.

[14] Y. Zhang and M. Juhola. On biometric verification of a user by means of eye movement data mining. In *The proceeding of the Second International Conference on Advances in Information Mining and Management*, pages 85–90. Xpert Publishing Services, 2012.