CrossMark

THEORETICAL ADVANCES

Fusion of eye movement and mouse dynamics for reliable behavioral biometrics

Pawel Kasprowski[1] · Katarzyna Harezlak[1]

Abstract This paper presents the first attempt to fuse two different kinds of behavioral biometrics: mouse dynamics and eye movement biometrics. Mouse dynamics were collected without any special equipment, while an affordable The Eye Tribe eye tracker was used to gather eye movement data at a frequency of 30 Hz, which is also potentially possible using a common web camera. We showed that a fusion of these techniques is quite natural and it is easy to prepare an experiment that collects both traits simultaneously. Moreover, the fusion of information from both signals gave 6.8 % equal error rate and 92.9 % accuracy for relatively short registration time (20 s on average). Achieving such results were possible using dissimilarity matrices based on dynamic time warping distance.

Keywords Eye movement · Mouse dynamics · Biometric fusion

1 Introduction

There have been many solutions developed for user identification including passwords, PINs, access tokens, ID badges and PC cards, yet they are often inconvenient or even insufficient due to technological development. People are provided access to so many secured resources that they are not able to memorize all the necessary PIN codes and passwords. That is why so-called biometric identification that uses human body characteristics (like face, iris or fingerprint recognition) has gained interest. The most popular methods utilize mostly physiological patterns of a human body; however, this makes them vulnerable.

The aforementioned inconveniences led to a search for new solutions. Biometric identification based on human behavioral features may solve these problems. There are various human characteristics to be considered and explored for the purposes of biometric identification. Among them voice, gait, keystroke, signature [1] as well as eye movement and mouse dynamics should be mentioned.

The aim of the paper is to provide a new approach to biometric identification using a combined feature analysis based on eye movement and mouse dynamics signals. The main contribution of the paper is the first attempt to build an identification model based on a fusion of these two different biometric traits. For this purpose, a novel experiment that had not previously been studied was designed. Additionally, the usage of a dissimilarity matrix [2] to prepare samples for the classification purpose was introduced.

The paper is organized as follows. The state of the art of both mouse and eye-movement-based identification is presented in the second section. The third section describes the scenario of the experiments, the group of participants and the experimental setup. Section 4 contains details of the methods used to preprocess and extract features. This is followed by a description of the evaluation procedure. Section 5 contains results of the experiments. The discussion of these results is presented in Sect. 6. Finally, conclusions and future work are provided in Sect. 7.

2 State of the art

Both mouse dynamics and eye-movement-based biometrics have been studied previously; hence, this section provides some comparative analyses of previous achievements.

✉ Pawel Kasprowski
  pawel.kasprowski@polsl.pl

[1] Institute of Informatics, Silesian University of Technology, ul. Akademicka 16, 44-100 Gliwice, Poland

2.1 Information fusion in biometrics

Information fusion is a very popular tool for improving biometric identification system performance. According to [3], fusion may combine multiple representations of the same biometric trait, multiple matchers using the same representation and, finally, multiple biometric modalities. Multimodal fusion may be done on various levels: (1) a feature extraction level, in which multiple traits are used together to form one feature vector; (2) a matching score level, in which results (typically similarity scores) obtained from different biometric systems are fused; and (3) a decision level, in which only output decisions (accept/reject) from different biometric systems are used in a majority vote scheme.

There are a lot of examples of multimodal biometric fusions. The most popular are fusions of physiological modalities like face and iris [4, 5] or fingerprint and iris [6, 7]. There are also works that present a fusion of the same modality measured by different sensors [8]. Finally, fusions of different algorithms processing the same data on matching score or decision levels have improved biometric identification results significantly [9, 10].

2.2 Mouse dynamics

Analyzing the research regarding mouse event-based biometric identification, we find various approaches and many features of mouse movement that have been studied. Data obtained as a dynamic mouse signal consist of recordings including low-level mouse events such as raw movement and pressing or releasing mouse buttons. These are typically the timestamps and coordinates of an action and can be grouped in higher-level events such as move and click, highlight a text, or a drag and drop task. Based on these aggregated actions, a number of mouse-related features have been developed and applied for user identification.

Experiments available in the literature may be differentiated by various aspects. The first of them is the type of experiment, which includes edit text tasks [11], browser tasks [11, 12] and game scenarios [11, 13]. Ahmed and Traore [14] collected data during users' daily activities. Similarly, online forum tasks for gathering mouse movement signal were utilized in the studies presented in [15]. A different type of experiment was proposed in the research presented in [16], in which a user had to use a mouse to follow a sequence of dots presented on a screen.

Studies may also be analyzed in terms of the environments used. In one group of experiments, participants worked on computers without any specially prepared environment [11, 12, 14]. Another approach was to use a controlled environment to prevent unintended events influencing the quality of samples [16–18]. Zheng and el.

[15] conducted tests in a self-prepared environment involving routine, continuous mouse activities as well as using an online forum.

Research can also be classified by the time in which an authentication takes place. There are studies that collected such data only at the beginning of the session [16] or continuously during the whole session [11, 13, 14, 18]. Since data gathered during experiments have to be processed to be useful in further analysis, each registered mouse movement signal is divided into small elements representing various mouse actions. Among such elements, several features can be distinguished, forming two types of vectors: spatial and temporal. The first describes changes in mouse position and includes mouse position coordinates; mouse trajectory; angle of the path in various directions; and curvature and its derivative. The second type of vectors depicts quantities related to mouse movement like horizontal, vertical, tangential and angular velocities, tangential acceleration and jerk.

The mouse movement dynamic has also been used in research applying various fusion methods. For example, in [19] a fusion of keystroke dynamics, mouse movement and stylometry was studied. Keyboard and mouse dynamics were also used in [20], yet this time were fused with interface (GUI) interactions. Two types of fusion were utilized: feature level fusion and decision level fusion.

We have also found studies in which: (1) two multimodal systems that combine pen/speech and mouse/keyboard modalities were evaluated [21]; and (2) fingerprint technology and mouse dynamics were used [22]. A different type of mouse dynamic-related fusion was utilized in [23]. This fusion considered only mouse movement, yet divided it into independently classified feature clusters. Subsequently, a score level fusion scheme was used to make the final decision.

2.3 Eye movement biometrics

Eye movement biometrics have been studied for over 10 years [24, 25] on the assumption that the way in which people move their eyes is individual and may be used to distinguish them from each other. Two aspects of eye movement may be analyzed: the physiological, concerning the way that a so-called oculomotor plant works, and the behavioral, which focuses on the brain activity that forces eye movement. Therefore, plenty of possible experiments may be utilized.

The most popular experiments focus just on forcing eye movements, as the physiological aspect seems easier to analyze and more repeatable. The simplest example of such an experiment is a so-called jumping point stimulus. During such a scenario, users must follow with their eyes a point displayed on a screen periodically changing position

[24, 26, 27]. Studies with this kind of stimulus mostly measure physiological aspects, as subjects are instructed where to look and cannot make this decision autonomously.

The other popular type of experiment is recording eye movement while users are looking at a static image [25, 28, 29]. The content of the image may differ, but the most popular content so far is images with human faces. This results from the conviction that the way in which faces are observed is different for everyone [28, 30, 31]. A changing scene (movie) is the other possible stimulus [32, 33].

Another kind of experiment is recording eye movement while users fulfill some specific visual tasks. This seems to be a promising scenario; however, there are only a few research papers published so far including text reading [34], following with eyes more complex patterns [35] and normal activity like reading and sending emails [36].

When eye movement recordings are gathered, the next problem is how to extract attributes that may be usable for human identification. Various approaches have been proposed, one of the most popular of which involves the extraction of fixations (moments when an eye is relatively still to enable the brain to acquire a part of an image) and saccades (rapid movement from one fixation to another) and performing different statistical analyses on them. Simple statistics may be applied [37–39] or more sophisticated, like comparisons of distributions used [40]. In ref. [26], an interesting attempt to use eye movement data to build a mathematical model of the oculomotor plant has also been presented. Other approaches analyze the eye movement signal using well-known transformations like Fourier, wavelet or cepstrum [24, 41, 42]. There are also some methods that take spatial positions of gaze data into account to build and then analyze heat maps or scan paths [28, 30].

The results obtained in all the aforementioned experiments are far from ideal. Additionally, it is difficult to compare results of various experiments because scenarios, hardware (i.e., eye tracker) and participants vary between them all. Unfortunately, authors are reluctant to publish their data, which would enable future comparisons. A notable exception is the EMBD database (http://cs.txstate.edu/~ok11/embd_v2.html) published by Texas State University and databases used in publicly accessible Eye Movement Verification and Identification Competitions: EMVIC 2012 [27] and EMVIC 2014 [31].

Although it seems natural that the eye movement modality may be combined with other modalities, to the best of our knowledge there have been only two attempts to provide eye movement biometrics in fusion with another modality. In ref. [43], eye movements were combined with keystroke dynamics, but the results showed that errors for eye movements were very high and the improvement when fusing both keystroke and eye movements was not significant. In ref. [44], eye movement biometrics were fused with iris recognition using low-quality images recorded with a cheap web camera.

2.4 Paper's contribution

The analysis of the existing methods used for biometric identification in both previously described areas encouraged the authors to undertake studies aimed at compounding signals of eye and mouse movement in a user authentication process. There are several reasons that such studies are worth undertaking. Both signals stem from human behavioral features, which are difficult to forge. Their collection is easy and convenient for users, who naturally use their eyes and a mouse to perform computer-related tasks. Furthermore, the devices that acquire these signals are simple and cheap, especially when built-in web cameras are used, and can be easily incorporated in any environment by installing the appropriate software. The important feature of the considered solution is also the fact that both signals can be registered simultaneously, which makes data collection quicker. Additionally, if necessary, the method may also be used for covert authentication.

A novel type of experiment that was based on entering a PIN was designed for this purpose.

Data obtained from both eye and mouse movements were processed to construct dissimilarity matrices [2] that would provide a set of samples for training and testing phases of a classification process. A similar approach was used in [17] for mouse dynamics; however, it has never been applied for eye movement data. Taking the above into consideration, the research contribution may be listed as follows:

- Introduces a new idea for biometric identification based on fusion of eye and mouse movements that reduces identity verification time and improves security.
- Elaborates a new experiment type which can be easily applied in many environments.
- Applies a dissimilarity space using dynamic time warping for extraction of features from eye movement and mouse dynamics.

3 Experiment

This section describes the environment used for conducting experiments. The test scenario and some quantitative information about the data analyzed are presented.

3.1 Scenario

All data were gathered with one experimental setup consisting of a workstation system equipped with an optical mouse and the Eye Tribe (www.theeyetribe.com) system for recording eye movement signal at sampling rate of 30 Hz and an accuracy error of less than $1°$. It is worth mentioning that this eye tracker is affordable ($100) and convenient to use, unlike most of the eye trackers used in the previous research of eye movement biometrics. The eye tracker was placed below a screen of size 30×50 cm. The users sat centrally at a distance of 60 cm. Three such systems were used simultaneously during the data collection phase. The low frequency usage was motivated by the idea of checking whether valuable data may be obtained even for frequencies available to commonly used web cameras. Additionally, mouse movements were recorded with the same frequency.

All tests were conducted in the same room. At the beginning of each session, participants signed a consent form and were informed about the purpose of the experiment. Each session for each participant started with a calibration process ensuring adjustment of an eye tracker to the eye movement of the particular user. Users were asked to follow a point on the screen with their eyes. After nine locations, the eye tracker system was able to build a calibration function and measure a calibration error. Only users obtaining a calibration error value below $1°$ were allowed to continue the experiment.

In the next step, circles with 10 digits (0–9) were evenly distributed over the screen, displayed (Fig. 1). The participant's task was to click these circles with the mouse to enter a PIN number. The PIN was defined as a four-digit sequence, for which every two consecutive digits were always different. Both mouse positions and eye gaze positions were recorded during this activity. It was assumed that people look where they click with the mouse; therefore, eye and mouse positions should follow more or less the same path. One such recording of a PIN being
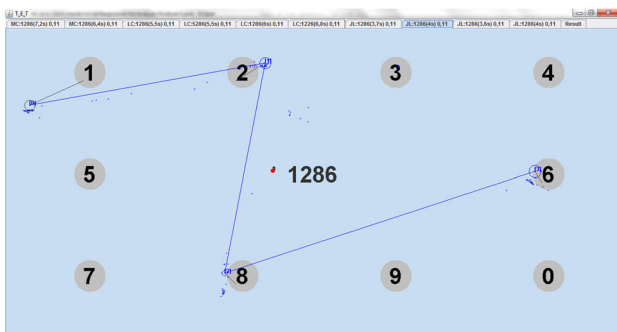
Fig. 1 Example view of a screen with eye movement fixations mapped to the chosen digits

entered is called a trial in subsequent sections. A trial is a completed task of entering one PIN, during which eye and mouse movements were registered. To make simulation of a genuine–impostor behavior possible, all participants entered the same PIN sequence: 1–2–8–6.

There were several sessions with at least a 1-week interval between sessions. During each session, the task was to enter the same PIN three times in a row.

3.2 Collections used

A total of 32 participants took part in the experiments, and 387 trials were collected. As each user entered the PIN three times during one experiment, the trials were grouped into sessions. Each user's session consisted of three subsequent trials. The gathered trials were used to prepare three collections differing in the number of sessions registered for one user:

- C4—24 users, four sessions per user, each containing three trials,
- C3—28 users, three sessions per user, each containing three trials,
- C2—32 users, two sessions per user, each containing three trials.

4 Methods

The data gathered in the described experiment were then processed to obtain information about people's identity. The process was divided into several phases:

- Preparation phase—when every trial was processed to extract different signals,
- Feature extraction phase—when a sample was built on the basis of features derived from signals (there are three different approaches presented below),
- Training phase—when samples with known identity were used to build a classification model,
- Testing phase—when the model was used to classify samples with unknown identity,
- Evaluation phase—when the results of the testing phase were analyzed.

This section describes all these steps in detail.

4.1 Preparation phase

The aim of the preparation phase was to separate different signals from eye and mouse movements recorded during the experiments. A signal is defined as a characteristic feature that can be extracted from each trial. This analysis concerned only parts of recordings collected between the first and fourth mouse click.

Table 1 Set of signals extracted from eye and mouse movements

Signal	Formula	Description
x, y	X and Y	The raw coordinates
vx, vy	$V_x = \frac{\partial x}{\partial t}, V_y = \frac{\partial y}{\partial t}$	The first derivative of X and Y (i.e., vertical and horizontal velocities)
vxy	$V = \sqrt{V_x^2 + V_y^2}$	The first derivative for absolute velocity
ax, ay	$V'_x = \frac{\partial V_x}{\partial t}, V'_y = \frac{\partial V_y}{\partial t}$	The second derivative of X and Y (i.e., vertical and horizontal accelerations)
axy	$V' = \sqrt{V_x'^2 + V_y'^2}$	The derivative of vxy
jx, jy	$V''_x = \frac{\partial V'_x}{\partial t}, V''_y = \frac{\partial V'_y}{\partial t}$	The third derivative of X and Y (jerk)
jxy	$V'' = \sqrt{V_x''^2 + V_y''^2}$	The derivative of axy
Diffmgx	$x_{\text{mouse}} - x_{\text{gaze}}$	The difference between mouse and gaze positions—axis x
Diffmgy	$y_{\text{mouse}} - y_{\text{gaze}}$	The difference between mouse and gaze positions—axis y

Fig. 2 Diagram of the statistic-based feature extraction algorithm
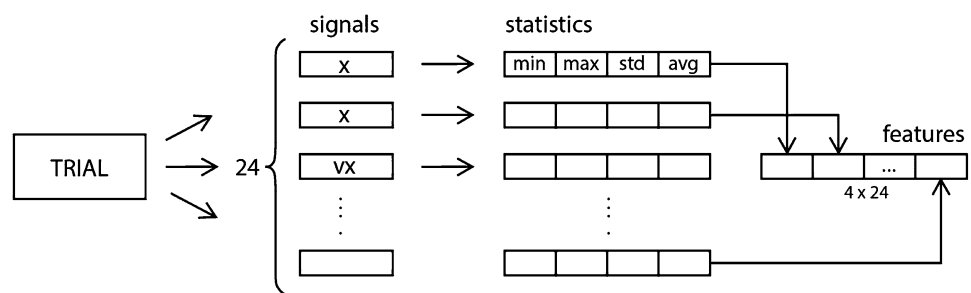
As a result, 24 separate signals were calculated: 11 signals for mouse, 11 signals for gaze and two additional signals representing mouse and eye position differences (Table 1). Depending on the length of the recording, each signal consisted of 105–428 values (from 5 to 21 s).

4.2 Feature extraction phase

The second step in the authentication process was to define a set of samples that could be used as input for a classifier. The input for this phase was the fusion of 24 mouse and eye signals prepared for each trial earlier.

Three different feature extraction algorithms were used:

- Statistic values
- Histograms
- Distance matrix

The detailed description of each is presented in the following sections.

4.2.1 Features based on statistic values

The first of the applied methods is commonly used in many studies [13, 16, 18]. It is based on statistical calculations relating to previously extracted signals. For each, four

statistics were calculated independently for each trial: min, max, avg, stdev. A sample in this method was defined as a vector including statistics for all signals from one trial. As the total number of signals was 24, a vector consisted of $24 \times 4 = 96$ attributes (Fig. 2).

4.2.2 Histograms

In the second of the feature extraction methods, a sample is represented by histograms built for each signal and evaluated for each trial separately. The frequencies of values occurring in histogram bins were stored as sample attributes. Because various numbers of bins (B) were considered—$B \in (10, 20, 30, 40, 50)$—a sample for one trial consisted of $24 * B$ attributes.

4.2.3 Distance matrix

In the last of the developed methods, the feature extraction process was based on an evaluation of distances between all training trials. While constructing relevant data structures, the signal-based description of a trial was taken into account. Therefore, each signal (for instance x, vx, y, vy) was treated individually and was used to build an independent distance matrix. Let us recall that 24 signals were

Fig. 3 Diagram of the feature extraction algorithm based on a distance matrix
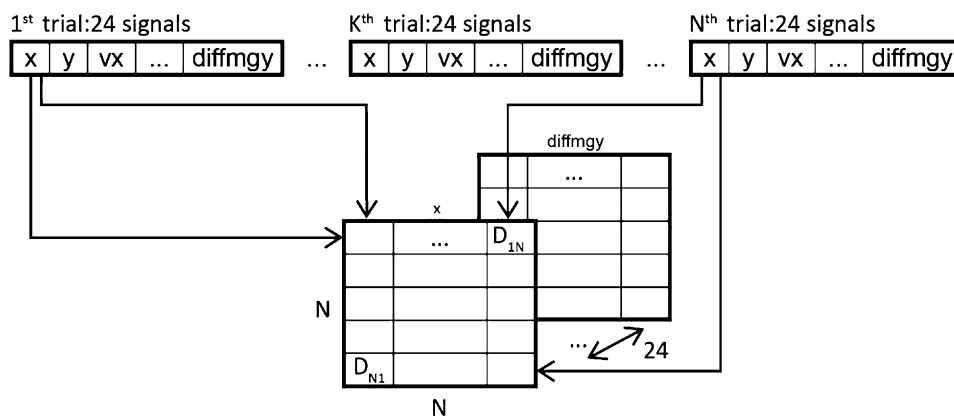
determined in the preparation phase; thus, 24 distance matrices were built. Further, for N training trials, a matrix consisting of N rows and N columns ($N \times N$ cells) was obtained to define distances for all training trials (Fig. 3).

Various metrics may be used when comparing distances of two signals. Euclidean is most common, based on the sum of all differences for every value registered for a signal. However, the Euclidean metric is not robust when comparing shapes of signals, which are shifted in time. Therefore, it was decided to use a nonlinear dynamic time warping distance metric for signal comparisons [45]. The DTW algorithm first calculates distances between all values in both signals and then searches for a sequence of point pairs (called the warping path) that minimizes the warping cost (sum of all distances) and satisfies boundary, continuity and monotonicity conditions [46]. The distance for each signal was calculated as the sum of distances between point pairs on the warping path (see Eq. 1).

$$\mathrm{DTW}\left(T_{\mathrm{a}}^{\mathrm{signal}}, T_{\mathrm{b}}^{\mathrm{signal}}\right) = \sqrt{\sum_{k=0}^{K}(w_k)/K} \qquad (1)$$

where $w_0 - w_K$ is a warping path consisting of K points with (i, j) coordinates and

$$w_k = \left(T_{\mathrm{a}}^{\mathrm{signal}}[i] - T_{\mathrm{b}}^{\mathrm{signal}}[j]\right)^2 \qquad (2)$$

The DTW algorithm applied for two signals from two different trials T_i and T_j provided one value representing their distance D_{ij}^{signal}. This value became an element of a distance vector forming a sample of the analyzed signal. A similar attempt limited to mouse dynamics signal was used in [17].

$$D^{\mathrm{signal}} = \begin{vmatrix} D_{11} & \cdots & D_{1N} \\ \vdots & \ddots & \vdots \\ D_{N1} & \cdots & D_{NN} \end{vmatrix}, \quad \mathrm{signal} \in 1 \cdots 24 \qquad (3)$$

For classification purposes, every column of such a matrix was treated as one feature. The rows of the matrices were

then used as training samples to train classifiers. The same procedure was then repeated for every testing sample, whose distances to all N training samples were calculated and used as N features of that sample. The distances were calculated for each of 24 signals forming 24 matrices.

4.3 Training and testing phase

At the end of the feature extraction phase, several sets of samples were collected:

1. One set with statistic values as features—*stat*,
2. Five sets with histograms for 10, 20, 30, 40 and 50 bins as features—$\mathrm{hist}_{\mathrm{bin}}$,
3. 24 sets with DTW distances as features—one for each signal type—$\mathrm{matrix}_{\mathrm{signal}}$.

All these sets were built separately for all collections of trials (C2, C3 and C4) described in Sect. 3.2. Each set, divided into N training and M testing samples, was then evaluated using the cross-validation method (Table 2). It is very important to emphasize that the division into training and testing sets was not random. Consecutively collected trials tend to be more similar to each other than trials collected after longer intervals; therefore, due to the short-term learning effect [47], including them in both training and testing sets may produce improperly obtained better accuracy results. Hence, the general rule was not to use trials of the same user gathered in the same session for both

Table 2 Number of training and testing samples for each collection

Collection	Samples per user	Training samples (N)	Testing samples (M)
C4	12	216	72
C3	9	168	84
C2	6	96	96

training and testing purpose. Detailed analysis of this phenomenon can be found in Sect. 5.2.

Building a rule according to which a fold was related to one session was a motivating factor. Therefore, collection C4 was divided into fourfold representing four sessions. As a result, all samples of one user from the same session were always in the same fold and were used together as either training or testing samples. A similar procedure was applied for C3 and C2 collections, dividing them into three and twofold, respectively. For such a folding strategy, a testing set always contained three trials of each user recorded during the same session (one by one).

A classification model was built based on N training samples, with usage of an SVM classifier [48]. Using data of a similar structure utilized in our previous research [49] and a grid search algorithm, we obtained the best results for the RBF kernel with gamma $= 2^{-9}$ and $C = 2^{15}$. Therefore, these values were used in the current research. The sequential minimal optimization algorithm was used [50] with the multiclass problem solved using pairwise coupling [51]. The classification model was then used for classification of M testing samples. For each of them, the classifier returned a vector of probability values that a given sample belongs to a particular user. If the number of users is denoted by U, for every testing sample we obtain a U element vector representing distribution of probabilities for each of U possible classes. A set of such M vectors (for all testing samples) forms a matrix of size $M \times U$.

Initially, during the testing phase, all trials in a testing set were classified separately giving independent distributions for each trial a: Ptrial$_a$. These distributions were subsequently summed up and normalized for trials related to the same session (let us recall that there were three trials for one session). Having probability vectors of three trials (a, b and c) of the same user gathered during the same session, the probability vector for the session was calculated as:

$$\text{Psession}_i^{\text{set}} = \frac{(\text{Ptrial}_a^{\text{set}} + \text{Ptrial}_b^{\text{set}} + \text{Ptrial}_c^{\text{set}})}{3} \qquad (4)$$

where *set* represents the set of samples used. Such a probability vector was the outcome of the method using the statistic features. However, an additional step was designed for *hist*$_{\text{bin}}$ and matrix$_{\text{signal}}$ types as both corresponding methods for the feature extraction define more than one set. The histogram method provided different sets for a particular number of bins (10, 20, 30, 40 and 50)—altogether five sets—whereas in the distance matrix approach we obtained 24 sets, each for one signal. Hence, the result in these cases was determined as a sum calculated for all bins or signals sets. The vector of probability distribution, after the last step, included values as those presented in Eq. 5, where X represented the number of sets used (a number of bins or a number of signals, 5 or 24, respectively).

$$p_i = \frac{\sum_{j=1}^{X} \text{Psession}_i^{\text{set}_j}}{X} \qquad (5)$$

The result of this step was three probability distributions:

- One for statistic values.
- One for histogram values (normalized sum of results for five histograms).
- One for distance matrix values (normalized sum of results for matrices built for 24 signals).

These three distributions were then used in the subsequent evaluation step to check their correctness. It should be emphasized that in the process of the probability distribution evaluation, a fusion of features characterizing eye movement and mouse dynamic was applied.

4.4 Evaluation phase

The last step of the classification process was to assess the quality of models developed in the previous phases. The result of the testing phase was probability distributions for every possible class U (user identity). As was explained in the previous section, distributions were calculated using three trials from one session so the number of distributions was $S = M/3$, where M was the number of testing trials. The result was a matrix $P : [S \times U]$, where each element $p_{i,j}$ represented the probability that the *ith* testing sample belongs to user *j*.

In the evaluation phase, this matrix was used to calculate accuracy (ACC), false acceptance rate (FAR) and false rejection rate (FRR) for different rejection threshold *th* values and finally to estimate equal error rate (EER) for every collection and feature extraction method.

At first, the correctness of the classification $c(i)$ for every ith distribution on the basis of its correct class $u(i)$ was calculated as:

$$c(i) = \begin{cases} 1 & p_{i,u(i)} = \max(p_{i,1} \ldots p_{i,u}) \\ 0 & \text{otherwise} \end{cases} \qquad (6)$$

Then, the accuracy of the classification for the whole testing set was calculated:

$$\text{accuracy} = \frac{\sum_{i=1}^{S} c(i)}{S} \qquad (7)$$

The next step was calculation of acceptance $a_{i,j}$ for different thresholds *th*. The value of thresholds ranged from 0 to 1.

$$a_{i,j}(\text{th}) = \begin{cases} 1 & p_{i,j} > \text{th} \\ 0 & \text{otherwise} \end{cases} \qquad (8)$$

Based on this acceptance, it was possible to calculate FAR and FRR for different thresholds.
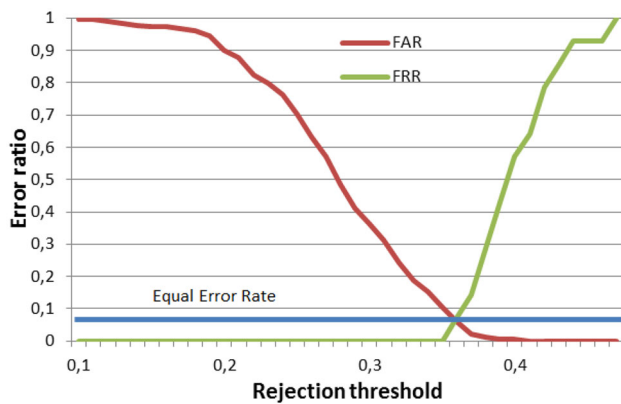
Fig. 4 Chart showing how FRR and FAR depend on the value of the rejection threshold

$$\text{FRR(th)} = \frac{S - \sum_{i=1}^{S} a_{i,u(i)}}{S} \tag{9}$$

$$\text{FAR(th)} = \frac{\sum_{i=1}^{S} \sum_{j=1, j \neq u(i)}^{U} a_{i,j}}{(U-1) * S} \tag{10}$$

It can be easily predicted that all samples were accepted for a rejection threshold th = 0; thus, FRR = 0 and FAR = 1. When increasing the threshold, fewer samples were accepted, hence FRR increased and FAR decreased. For th = 1, no samples were accepted, consequently FRR = 1 and FAR = 0. FAR and FRR dependency on rejection threshold value is presented in Fig. 4.

Equal error rate (EER) was calculated for the rejection threshold value for which FAR and FRR were equal (as visible in Fig. 4).

5 Results

Feature extraction methods used in training and testing phases and as presented in Sect. 4.2 were independently evaluated for each collection of trials: C4, C3 and C2. As was described earlier, they differed in the number of recorded sessions, which amounted 4, 3 and 2 sessions accordingly, whereas one session consisted of 3 trials. At the end of the classification process, two values were reported for each collection and each type of features (stat, hist, matrix). These were Accuracy and ERR, calculated according to methods described in evaluation phase section. The results are presented in Table 3.

The best result was obtained for collection C4, when the matrix type that was based on the fusion of distances of eye and mouse features was applied. In this case, 4 different sessions were available for each subject and the classification model was trained using three of them each time (12

Table 3 Results of identification (Accuracy) and verification (EER) for different collections and sets

Collection	Set	Accuracy (%)	EER (%)
C2	Stat	25.00	31.15
C2	Hist	21.88	34.78
C2	Matrix	15.62	34.59
C3	Stat	32.14	21.28
C3	Hist	32.14	20.68
C3	Matrix	46.43	16.78
C4	Stat	28.57	20.30
C4	Hist	57.14	10.32
C4	Matrix	92.86	6.82

trials compared to 9 in C3 and 6 in C2). The hist type was the best option also for collection C3, while the statistic method gave the lowest errors for C2. However, the results for collections C3 and C2 were significantly worse. The ERR value was 31.15 % (C2 collection and a stat set), which cannot be treated as a good outcome, especially as it was not significantly better than other ERR values for this collection. The probable reason of such findings was the fact that to build a training model for each user, less data were available (only two and one session accordingly).

The DET curves presenting the dependency of FRR and FAR ratios are shown in Fig. 5.

5.1 Comparison of mouse and gaze

The next research question was to check whether a fusion of gaze and mouse biometrics gives results better than a single modality. For this purpose, two additional experiments for the C4 dataset were performed: one using only mouse-related signals and one using only gaze-related signals. Both concerned only the matrix method, which yielded the best outcomes in the previous tests. Table 4 presents a comparison of these results to the fusion of both modalities.

The row denoted by "Gaze" corresponds to the efficiency of the algorithm when only 11 signals derived from eye movement were taken into account. The same regards the "Mouse" row, which shows results for 11 signals derived from mouse-related signals. The results presented in the "Fusion" row are calculated on the basis of all 24 signals (11 mouse + 11 gaze related + 2 based on mouse–gaze differences). All these outcomes revealed that mouse dynamics gave better accuracy and lower errors than eye movements. Most importantly, the fusion of mouse and gaze gave results significantly better than both modalities alone.
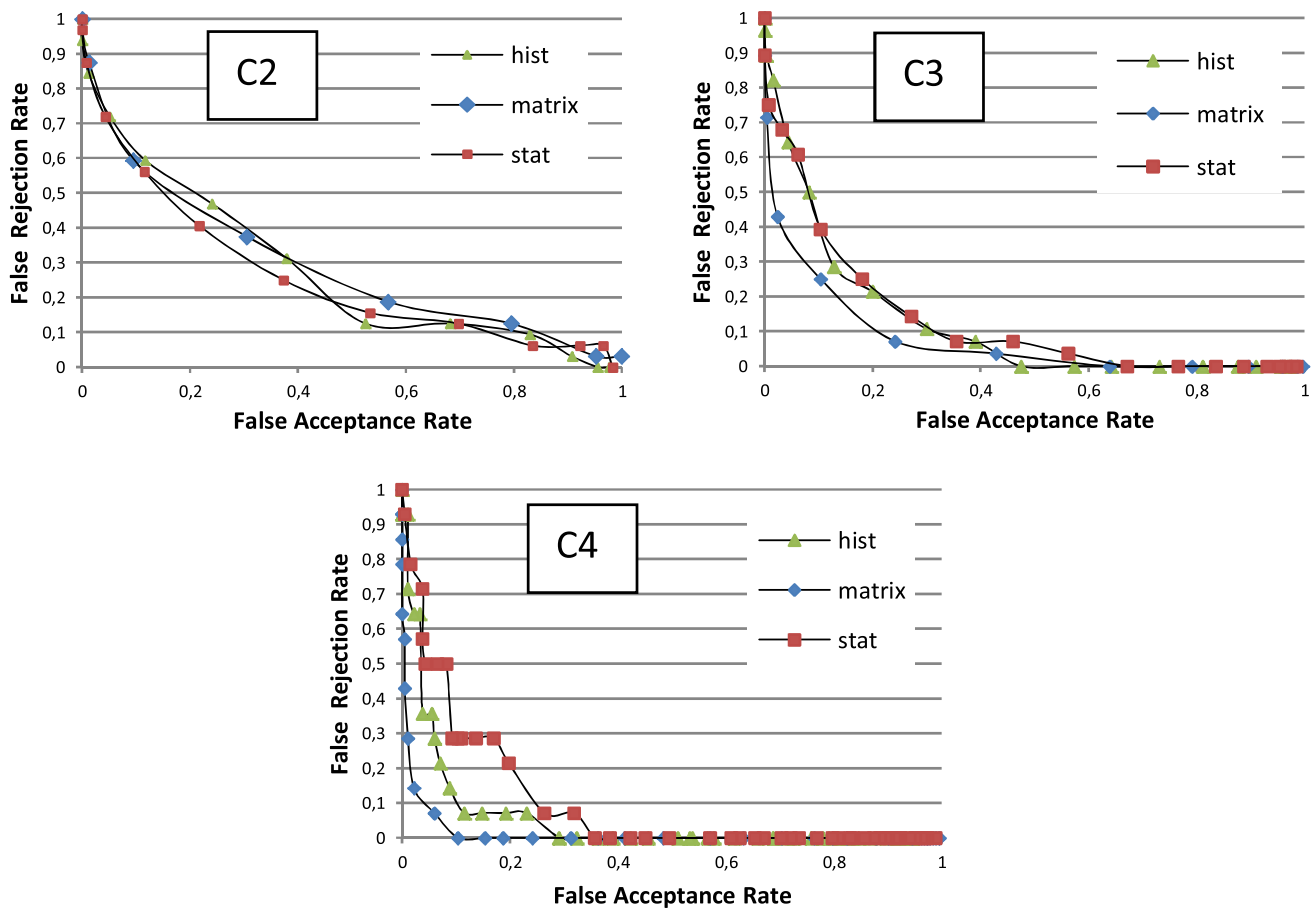
Fig. 5 DET curves for different feature extraction methods and collections C2, C3 and C4, respectively

Table 4 Results achieved for the matrix method for collection C4 for different subsets of signals

Set	Accuracy (%)	EER (%)
Fusion	92.86	6.82
Gaze	64.29	16.79
Mouse	78.57	9.05

5.2 Examining the learning effect

The learning effect is a phenomenon characteristic of biometric modalities that measures changes of human behavior over time [47]. It is sometimes treated as a kind of well-known template aging problem, but its nature is slightly different. While template aging is related to biometric template changes over a long time (e.g., a face gets older), the learning effect addresses short time changes in human behavior. It is obvious that a tired or sad person reacts differently than a rested and relaxed one. Various beverages and food such as coffee or alcohol may also influence people's behavior. For this reason, it is very important to register behavioral biometric templates with some considerable time interval to avoid short-term similarities and extract truly repeatable features. This phenomenon has already been studied for eye movement, and the results showed that eye movement samples collected at intervals of less than 10 min are much more similar to each other than samples collected at 1-week interval [52].

During the tests described in Sect. 4, we tried to avoid this problem by the appropriate preparation of training and testing folds of samples. We ensured that during the cross-validation, samples related to a user's session were never split into two folds (see Sect. 4.3) and the time interval between two sessions of the same user was never shorter than 1 week. We called this folding strategy "session-based folding," as data for the whole session was always in either a training or testing set.

However, we decided to raise the research question to check whether mixing samples derived from one session in training and testing sets did indeed result in better classification performance. Therefore, the additional cross-validation experiment was performed with a different fold preparation strategy. As there were always three trials in

Table 5 Results achieved for the matrix method for collection C4 for mixed session folding

Set	Accuracy (%)	EER (%)
Fusion	100	2.94
Gaze	92.86	9.37
Mouse	85.71	5.04

each session, this time every set was divided into three folds: The first trial of the session was in fold 1, the second attempt in fold 2 and the third one in fold 3. We called this folding strategy "mixed sessions folding," as this time trials from the same session were always divided into separate folds.

Using such folds for cross-validation ensured that there was always a sample of the same user from the same session in both training and testing sets. The classification results are compared to the previous ones and presented in Table 5.

As could be expected, the accuracy for modified folds was higher and errors were lower because it was easier for the classifier to classify a trial with two other trials from the same session (i.e., very similar). The errors were lower for both modalities, but the difference for gaze-based biometrics was more significant. As given in Table 5, accuracy for the gaze was even better than for the mouse. Accuracy for the fusion reached 100 % because the correct class had the highest probability for every sample, but EER was not 0 % because it was not possible to find one threshold that worked perfectly for every sample distribution. If a threshold perfectly separated probabilities of genuine and impostor classes for one sample, the same threshold did not work perfectly for other samples.

6 Discussion

At the beginning of our research, we raised some research questions that were answered one by one during consecutive experiments. Our primary objective was to examine the possibility of fusing eye and mouse characteristics to define a robust authentication model. Accuracy of 92.86 % and EER of 6.82 % seem to be very good results compared to previous studies concerning both modalities independently. Other advantage of our approach is the development of an identification/verification scenario that is very convenient for users and—very importantly compared to other research in this field—it takes on average only 20 s to collect biometric data. It must be mentioned that some authors of mouse-related research reported lower error rates, but these results were achieved for longer mouse recordings, e.g., 2.46 % EER for 17 min of a signal registration in [14]. Recordings with comparable time yielded results worse or comparable to ours, yet usually much more training data were required. An extended comparison of our method to others found in the literature is presented in Table 6.

A similar analysis may be provided that considers the second modality. The results obtained in our studies for eye-movement-related biometrics are comparable in performance to recent achievements. Yet, it is once again important to emphasize that our experiments required significantly shorter registration time. Another advantage of our method is that results were achieved for a very low frequency of eye movement recordings. Obviously, a frequency of 30 Hz gives less data for analysis; however, its advantage is that it can register eye movements with classic low frequency web cameras, which are built-in components of many computer systems.

Broader summary of results published since 2012 is found in Table 7.

On the basis of these comparisons, we may deduce that our feature extraction method based on the fusion of distance matrices gives very good results, even when much less data are available compared to previous research. On the other hand, fusing eye movement with mouse dynamics allows for further improvement of the overall results of the whole biometric system. Deeper analysis of the results reveals other important findings.

Table 6 Comparison of outcomes of different mouse-related research and the results presented in this paper

References	Testing sample duration (s)	Equal error rate	Training samples duration (s)
Gamboa et al. [13]	50 s (200 s)	2 % (0.2 %)	200 s
Hashiaa et al. [16]	20 s	15–20 % [HTER[a]]	400 s
Zheng et al. [15]	100 s–37 min	1.3 %	166 min–60 h
Feher et al. [18]	42 s (139 s)	10 % (7.5 %)	n/a (15 h per user)
Shen et al. [17]	12 s	8.35 % [HTER[a]]	885 s
Our result (mouse)	20 s	9.05 %	60 s
Our result (fusion)	20 s	6.82 %	60 s

[a] HTER—half total error rate—(FAR+FRR)/2 for some threshold

Table 7 Comparison of different gaze-related research with the results presented in this paper

References	Testing sample duration (s)	Equal error rate (%)	Recording frequency
Komogortsev et al. [53]	100 s	16 %	1000 Hz
Holland et al. [40]	60 s	16.5 %	1000 Hz
Holland et al. [40]	60 s	25.6 %	75 Hz
Rigas et al. [33]	60 s	12.1 %	1000 Hz
Cantoni et al. [30]	160 s	22.4 %	50 Hz
Tripathi et al [38]	60 s	37 %	1000 Hz
Our result (gaze)	20 s	16.79 %	30 Hz
Our result (fusion)	20 s	6.82 %	30 Hz

(1) We discovered that a modality based on mouse dynamics outperforms one based on eye movement; yet, more importantly, a fusion of both characteristics gives the best results.

(2) The conducted experiments were based on three different feature extraction strategies. The distance matrix-based feature extraction method outperforms traditional methods based on statistics and histograms with ERR of 6.82, 10.32, 20.30 %, respectively.

(3) Tests considering several collections with different numbers of trials, with the best results for those consisting of 3 training and 1 testing sessions (C4), showed that slightly increasing the number of training samples influences performance significantly.

(4) Last but not least of the findings, related to the learning effect, confirmed the importance of correct evaluation phase planning, which is especially remarkable when cross-validation is used, as an incorrect and unfair folding strategy may easily lead to a model overfitting.

7 Summary

The research presented in this paper aimed to find a new method for behavioral biometrics. The main objective of the studies was to find a solution characterized with a relatively short identity verification time and a low level of classification errors. The results obtained during experiments confirmed that the objective was achieved. The paper showed that the fusion of the mouse dynamics and eye movement modalities may be used for this purpose. Furthermore, it proved that such a fusion may be achieved in one experiment that is both short and convenient for participants.

The novel feature extraction method, which was based on fusion of distance matrices, yielded results comparable or better than those previously published for both single modalities. The algorithm applied in the method makes it useful for any kind of modality fusion.

It is also worth mentioning that despite the 6 % error rate, our method may be used in practical applications as a part of a verification system. Participants of our experiment entered a 4-digit PIN by clicking digits in the correct order with a mouse. Because we were interested in the comparison of eye and mouse movements only, all participants entered the same PIN (namely the sequence 1–2–6–8). However, in a real-life environment knowledge of a PIN could be the first stage of verification. If a participant entered the proper PIN, our algorithm would be activated to check whether the participant's identity claim was genuine. The proper setting of the rejection threshold could lower false rejections, as it is unlikely that an impostor knows the PIN number and has similar mouse and eye movement dynamics that characterize a genuine user.

To conclude the presented studies, we will summarize the most important contributions of the paper:

1. The proposed feature extraction method using the fusion of distance matrices gave results (92.86 % accuracy and 6.82 % Equal Error Rate) which are competitive compared to those already published in this field, while less data were used for both training and testing phases (about 60 and 20 s, accordingly). This is the case for both eye movement and mouse dynamics.

2. The paper showed that the fusion of the mouse dynamics and eye movement modalities can be done in one experiment which is both short and convenient for participants.

3. We showed that the fusion of these two modalities may lead to better results than for each single modality.

4. It was shown that eye movement data recorded with a low frequency (30 Hz) may give information sufficient to achieve equal error rates (16.79 %) comparable to the state-of-the-art results.

Additionally, it should be noticed that the setup of the experiment is not complicated and may be reconstructed

easily. The only hardware requirements are a computer equipped with a mouse and an eye tracker. The research described in the paper showed that the frequency of commonly used webcams may provide satisfactory results. The appropriate software (e.g., ITU Gaze Tracker) could be used in this case. Another affordable solution is a low-cost remote eye tracer, like that used in the experiments (i.e., Eye Tribe).

7.1 Future work

When designing our research, we decided to involve the fusion technique on the decision level for the distance matrix method and on the feature level for the statistic one [3]. The next planned step is to extend all methods to involve fusion on various levels. For this purpose, various feature selection methods are also planned to be taken into consideration.

Additionally, we plan to conduct the same experiments for more participants. Data were collected for 32 participants used during the experiment. Such a pool of data seem to be enough to draw some meaningful conclusions; however, a much larger pool is necessary to confirm our findings. Moreover, our experiments showed that a higher number of training samples guarantees better classification performance. Therefore, it may be expected that more than three training samples (as was for our best collection) should improve the results. Five to six sessions are planned for each participant. With more data to analyze, it would be possible to calculate weights for each of the elements of the fusion. Weighted fusion would probably give even better results.

References

1. Porwik P, Doroz R, Wrobel K (2009) A new signature similarity measure. In: IEEE world congress on nature and biologically inspired computing, 2009. NaBIC 2009, pp 1022–1027. doi:10. 1109/NABIC.2009.5393858
2. Duin RP, Pekalska E (2012) The dissimilarity space: bridging structural and statistical pattern recognition. Pattern Recognit Lett 33(7):826–832
3. Ross A, Jain A (2003) Information fusion in biometrics. Pattern Recognit. lett. 24(13):2115–2125
4. Wang Y, Tan T, Jain AK (2003) Combining face and iris biometrics for identity verification. In: Kittler J, Nixon MS (eds) Audio-and video-based biometric person authentication. Springer, Heidelberg, pp 805–813
5. Connaughton R, Bowyer KW, Flynn PJ (2013) Fusion of face and iris biometrics. In: Handbook of Iris Recognition. Springer, pp 219–237
6. Conti V, Militello C, Sorbello F, Vitabile S (2010) A frequency-based approach for features fusion in fingerprint and iris multimodal biometric identification systems. Syst Man Cybern Part C Appl Rev Trans 40(4):384–395
7. Mehrotra H, Rattani A, Gupta P (2006) Fusion of iris and fingerprint biometric for recognition. In: Proceedings of international conference on signal and image processing, pp 1–6
8. Marcialis GL, Roli F (2004) Fingerprint verification by fusion of optical and capacitive sensors. Pattern Recognit Lett 25(11):1315–1322
9. Prabhakar S, Jain AK (2002) Decision-level fusion in fingerprint verification. Pattern Recognit 35(4):861–874
10. Vatsa M, Singh R, Noore A (2008) Improving iris recognition performance using segmentation, quality enhancement, match score fusion, and indexing. Syst Man Cybern Part B Cybern IEEE Trans 38(4):1021–1035
11. de Oliveira PX, Channarayappa V, ODonnel E, Sinha B, Vadakkencherry A, Londhe T, Gatkal U, Bakelman N, Monaco JV, Tappert CC (2013) Mouse movement biometric system. In: Proceedings of CSIS Research Day
12. Jorgensen Z, Yu T (2011) On mouse dynamics as a behavioral biometric for authentication In: Proceedings of the 6th ACM symposium on information, computer and communications security. ACM, pp 476–482
13. Gamboa H, Fred A (2004) A behavioral biometric system based on human–computer interaction In: Defense and security. International society for optics and photonics, pp 381–392
14. Ahmed AAE, Traore I (2007) A new biometric technology based on mouse dynamics. Depend Secure Comput IEEE Trans 4(3):165–179
15. Zheng N, Paloski A, Wang H (2011) An efficient user verification system via mouse movements. In: Proceedings of the 18th ACM conference on computer and communications security. ACM, pp 139–150
16. Hashiaa S, Pollettb C, Stampc M, Hall M (2005) On using mouse movements as a biometric. In: Proceeding of the international conference on computer science and its applications, vol 1
17. Shen C, Cai Z, Guan X, Du Y, Maxion RA (2013) User authentication through mouse dynamics. Inf Forens Secur IEEE Trans 8(1):16–30
18. Feher C, Elovici Y, Moskovitch R, Rokach L, Schclar A (2012) User identity verification via mouse dynamics. Inf Sci 201:19–36
19. Calix K, Connors M, Levy D, Manzar H, MCabe G, Westcott S (2008) Stylometry for e-mail author identification and authentication. In: Proceedings of CSIS research day, Pace University, pp 1048–1054
20. Bailey KO, Okolica JS, Peterson GL (2014) User identification and authentication using multimodal behavioral biometrics. Comput Sec 43:77–89
21. Perakakis M, Potamianos A (2008) Multimodal system evaluation using modality efficiency and synergy metrics. In: Proceedings of the 10th international conference on multimodal interfaces. ACM, pp 9–16
22. Asha S, Chellappan C (2008) Authentication of e-learners using multimodal biometric technology. In: Biometrics and Security Technologies, ISBAST 2008. International Symposium on IEEE, vol 2008, pp 1–6
23. Nakkabi Y, Traoré I, Ahmed AAE (2010) Improving mouse dynamics biometric performance using variance reduction via extractors with separate features. Syst Man Cybern Part A Syst Hum IEEE Trans 40(6):1345–1353

24. Kasprowski P, Ober J (2004) Eye movements in biometrics. In: International workshop on biometric authentication. Springer, Heidelberg, pp 248–258

25. Maeder AJ, Fookes CB (2003) A visual attention approach to personal identification. In: Eighth australian and new zealand intelligent information systems conference, pp 10–12

26. Komogortsev OV, Jayarathna S, Aragon CR, Mahmoud M (2010) Biometric identification via an oculomotor plant mathematical model. In: Proceedings of the 2010 symposium on eye-tracking research and applications. ACM, pp 57–60

27. Kasprowski P, Komogortsev OV, Karpov A (2012) First eye movement verification and identification competition at btas 2012. In: biometrics: theory, applications and systems (BTAS), 2012 IEEE fifth international conference on IEEE, pp 195–202

28. Rigas I, Economou G, Fotopoulos S (2012) Biometric identification based on the eye movements and graph matching techniques. Pattern Recognit Lett 33(6):786–792

29. Deravi F, Guness SP (2011) Gaze trajectory as a biometric modality. In: Proceedings of the BIOSIGNALS conference, Rome, Italy, pp 335–341

30. Cantoni V, Galdi C, Nappi M, Porta M, Riccio D (2015) Gant: gaze analysis technique for human identification. Pattern Recognit 48(4):1027–1038

31. Kasprowski P, Harezlak K (2014) The second eye movements verification and identification competition. In: Biometrics (IJCB), 2014 IEEE international joint conference on IEEE, pp 1–6

32. Kinnunen T, Sedlak F, Bednarik R (2010) Towards task-independent person authentication using eye movement signals. In: Proceedings of the 2010 symposium on eye-tracking research and applications. ACM, pp 187–190

33. Rigas I, Komogortsev OV (2014) Biometric recognition via probabilistic spatial projection of eye movement trajectories in dynamic visual environments. Inf Forensics Sec IEEE Trans 9(10):1743–1754

34. Holland C, Komogortsev OV (2011) Biometric identification via eye movement scanpaths in reading. In: Biometrics (IJCB), 2011 international joint conference on IEEE, pp 1–8

35. Darwish A, Pasquier M (2013) Biometric identification using the dynamic features of the eyes. In: Biometrics: theory, applications and systems (BTAS), 2013 IEEE sixth international conference on IEEE, pp 1–6

36. Biedert R, Frank M, Martinovic I, Song D (2012) Stimuli for gaze based intrusion detection. In: Future information technology, application, and service, ser. Lecture notes in electrical engineering. Springer, the Netherlands, vol. 164, pp 757–763

37. Holland CD, Komogortsev OV (2012) Biometric verification via complex eye movements: the effects of environment and stimulus. In: Biometrics: theory, applications and systems (BTAS), 2012 IEEE fifth international conference on IEEE, pp 39–46

38. Tripathi B, Srivastava V, Pathak V (2013) Human recognition based on oculo-motion characteristics. In: AFRICON IEEE 2013, pp 1–5

39. Zhang Y, Juhola M (2012) On biometric verification of a user by means of eye movement data mining. In: Proceedings of the 2nd international conference on advances in information mining and management

40. Holland CD, Komogortsev OV (2013) Complex eye movement pattern biometrics: analyzing fixations and saccades. In: Biometrics (ICB), 2013 International conference on IEEE, pp 1–8

41. Cuong NV, Dinh V, Ho LST (2012) Mel-frequency cepstral coefficients for eye movement identification. In: Tools with artificial intelligence (ICTAI), 2012 IEEE 24th international conference on IEEE, vol. 1, pp 253–260

42. Bednarik R, Kinnunen T, Mihaila A, Fränti P (2005) Eye-movements as a biometric. In: Image analysis. Springer, pp 780–789

43. Silver DL, Biggs A (2006) Keystroke and eye-tracking biometrics for user identification. In: International conference on artificial intelligence (ICAI), pp 344–348

44. Komogortsev OV, Karpov A, Holland CD, Proença HP (2012) Multimodal ocular biometrics approach: a feasibility study. In: Biometrics: theory, applications and systems (BTAS), 2012 IEEE fifth international conference on IEEE, pp 209–216

45. Berndt DJ, Clifford J (1994) Using dynamic time warping to find patterns in time series. In: KDD workshop, vol. 10, no. 16. Seattle, WA, pp 359–370

46. Keogh EJ, Pazzani MJ (2000) Scaling up dynamic time warping for data mining applications In: Proceedings of the sixth ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 285–289

47. Kasprowski P, Rigas I (2013) The influence of dataset quality on the results of behavioral biometric experiments. In: Biometrics special interest group (BIOSIG), 2013 international conference of the IEEE, pp 1–8

48. Hearst MA, Dumais ST, Osman E, Platt J, Scholkopf B (1998) Support vector machines. Intell Syst Appl IEEE 13(4):18–28

49. Kasprowski P and Harezlak K (2015) Using non-calibrated eye movement data to enhance human computer interfaces. In: Intelligent decision technologies. Springer, pp 347–356

50. Platt J et al (1999) Fast training of support vector machines using sequential minimal optimization. In: Advances kernel methods support vector learning, vo. 3

51. Hastie T, Tibshirani R et al (1998) Classification by pairwise coupling. Ann Stat 26(2):451–471

52. Kasprowski P (2013) The impact of temporal proximity between samples on eye movement biometric identification. In: Computer information systems and industrial management. Springer, pp 77–87

53. Komogortsev OV, Karpov A, Price LR, Aragon C (2012) Biometric authentication via oculomotor plant characteristics. In: Biometrics (ICB), 2012 5th IAPR international conference on IEEE, pp 413–420