



A score level fusion method for eye movement biometrics[☆]



Anjith George*, Aurobinda Routray

Department of Electrical Engineering, IIT Kharagpur, Kharagpur, West Bengal 721302, India

ARTICLE INFO

Article history:

Available online 2 December 2015

Keywords:

Eye tracking Biometrics
Gaze analysis
Eye movements
Decision level fusion
Radial basis functions
Gaze tracking

ABSTRACT

This paper proposes a novel framework for the use of eye movement patterns for biometric applications. Eye movements contain abundant information about cognitive brain functions, neural pathways, etc. In the proposed method, eye movement data is classified into fixations and saccades. Features extracted from fixations and saccades are used by a Gaussian Radial Basis Function Network (GRBFN) based method for biometric authentication. A score fusion approach is adopted to classify the data in the output layer. In the evaluation stage, the algorithm has been tested using two types of stimuli: random dot following on a screen and text reading. The results indicate the strength of eye movement pattern as a biometric modality. The algorithm has been evaluated on BioEye 2015 database and found to outperform all the other methods. Eye movements are generated by a complex oculomotor plant which is very hard to spoof by mechanical replicas. Use of eye movement dynamics along with iris recognition technology may lead to a robust counterfeit-resistant person identification system.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

Biometrics is an active area of research in pattern recognition and machine learning community. Potential applications of biometrics include forensics, law enforcement, surveillance, personalized interaction, access control [1], etc. Physiological features like fingerprint, DNA, earlobe geometry, iris pattern, facial recognition, [2] are widely used in biometrics. Recently, several behavioral biometric modalities have been proposed including gait, eye movement patterns, keystroke dynamics [3] signature, etc. Even though many such parameters like brain signals [4] (using electroencephalogram) and heart beats [5] have been proposed as biometric modalities, their invasive nature limits their practical applications.

An effective biometric should have the following characteristics [1]: (1) the features should be unique for each individual, (2) they should not change with time (template aging effects), (3) acquisition of parameters should be easy (low computational complexity and noninvasive), (4) accurate and automated algorithms should be available for classification, (5) counterfeit resistance, (6) low cost, and (7) ease of implementation. Other characteristics that might make the system more robust are portability and the ability to extract features from non-co-operative subjects.

Out of many biometric modalities, iris recognition has shown the most promising results [6] obtaining Equal Error Rates (EER) close to 0.0011%. However, it can only be used when the user is co-operative. Such systems can be spoofed by contact lenses with printed patterns. Even though most of the biometric modalities perform well on evaluation databases, one may be able to spoof such systems with mechanical replicas or artificially fabricated models [7]. In this regard, several approaches have been presented [8] to detect the liveliness of tissues or body parts presented to the biometric system. However, such methods are also vulnerable to spoofing.

Biometrics using patterns obtained from eye movements is a relatively new field of research. Most of the conventional biometrics use physiological characteristics of the human body. Eye movement-based biometrics tries to identify the behavioral patterns as well as information regarding physiological properties of tissues and muscles generating eye movements [9]. They provide abundant information about cognitive brain functions and neural signals controlling eye movements. Saccadic eye movement is the fastest movement (peak angular velocities up to 900° per second) in the human body. Mechanically replicating such a complex oculomotor plant model is extremely difficult. These properties make eye movement patterns a suitable candidate for biometric applications. The dynamics of eye movement along with these properties can give inbuilt liveliness detection capability.

Initially, eye movement biometrics has been proposed as a soft biometric. However, with the high level of accuracy achieved, it seems there are more opportunities regarding its application as an independent biometric modality. Eye movement detection can be

[☆] This paper has been recommended for acceptance by Maria De Marsico.

* Corresponding author. Tel.: +91 7501549613.

E-mail address: anjith2006@gmail.com (A. George).

integrated easily into already existing iris recognition systems. A combination of iris recognition and eye movement pattern recognition may lead to a robust counterfeit-resistant biometric modality with embedded liveness detection and continuous authentication properties. Eye movement biometrics can also be made task-independent [10] so that the movements can be captured even for non-co-operative subjects.

The rest of the paper is organized as follows. Section 2 describes previous works related to the use of eye movement as a biometric. Section 3 presents the proposed algorithm. Evaluation of the algorithm along with the results are outlined in Section 4. Conclusions regarding eye movement biometrics and possible extensions are detailed in Section 5.

2. Related works

Initial attempts to use eye movements as a biometric modality were carried out by Kasprowski and Ober [11]. They recorded the eye movements of subjects following a jumping dot on a screen. Several frequency domain and Cepstral features were extracted from this data. They applied different classification methods like naive Bayes, C45 decision trees, SVM and KNN methods. The results obtained further motivated research in eye movement-based biometrics. Bednarik et al. [12] conducted experiments on several tasks including text reading, moving cross stimulus tracking and free viewing of images. They used FFT and PCA on the eye movement data. Several combinations of such features were tried. However, the best results were obtained using the distance between eyes, which is not related to eye dynamics. Komogortsev et al. [13] used an Oculomotor Plant Mathematical Model (OPMM) to model the complex dynamics of the oculomotor plant. The plant parameters were identified from the eye movement data. This approach was further extended in [14]. Holland and Komogortsev [15] evaluated the applicability of eye movement biometrics with different spatial and temporal accuracies and various types of stimuli. Several parameters of eye movements were extracted from fixations and saccades. Weighted components were used to compare different samples for biometric identification. A temporal resolution of 250 Hz and spatial accuracy of 0.5° were identified as the minimum requirements for accurate gaze-based biometric systems. Kinnunen et al. [10] presented a task-independent user authentication system based on eye movements. Gaussian mixture modeling of short-term gaze data was used in their approach. Even though the accuracy rates were fairly low, the study opened up possibilities for the development of task-independent eye movement-based verification systems. Rigas et al. [16] explored variations in individual gaze patterns while observing human face images. Eye movements resulted were analyzed using a graph-based approach. The Multivariate Wald–Wolfowitz runs test was used to classify the eye movement data. This method achieved 70% rank-1 IR and 30% EER on a database of 15 subjects. Rigas et al. [17] extended this method using features of velocity and acceleration calculated from fixations. The feature distributions were compared using Wald–Wolfowitz test.

Zhang et al. [18] used saccadic eye movements with machine learning algorithms for biometric verification. They used multilayer perceptron networks, support vector machines, radial basis function networks and logistic discriminant for the classification of eye movement data. Recently Cantoni et al. [19] proposed a gaze analysis technique called GANT in which fixation patterns were denoted by a graph-based representation. For each user, a fixation model was constructed using the duration and number of visits at various points. Frobenius norm of the density maps was used to find the similarity between two recordings. Holland and Komogortsev presented an approach (CEM) [20] using several scan path features including saccade amplitudes, average saccade velocities, average saccade peak velocities, velocity waveform, fixation counts, average duration of fixation,

length of scan path, area of scan path, regions of interest, number of inflections, main sequence relationship, pairwise distances between fixations, amplitude duration relationship, etc. A comparison metric of the features was computed using Gaussian cumulative density function. Another similarity metric was obtained by comparing the scan paths. A weighted fusion of these parameters obtained the best case EER of 27%. Holland and Komogortsev proposed a method (CEM-B) [21], in which the fixation and saccade features were compared using statistical methods like Ansari–Bradley test, two-sample t-test, two-sample Kolmogorov–Smirnov test, and the two-sample Cramer–von Mises test. Their approach achieved 83% rank-1 IR and 16.5% EER on a dataset of 32 subjects.

To the best knowledge of the authors, the best case EER obtained is 16.5% [21]. Most of the works presented in the literature were evaluated on smaller databases. The effect of template aging was not considered in these works. For the application of eye movement as a reliable biometric, the patterns should remain consistent with time. In this paper, we try to improve upon the existing methods. The proposed algorithm can reach an EER up to 2.59% and rank-1 accuracy of 89.54% in RAN_30min dataset of BioEye 2015 database [22] containing 153 subjects. Template aging effect has also been studied using data taken after an interval of 1 year. The average EER obtained is 10.96% with a rank-1 accuracy of 81.08% with 37 subjects.

3. Proposed method

In the proposed approach, eye movement data from the experiment are classified into fixations and saccades, and their statistical features are used to characterize each individual. For each individual, the properties of saccades of same durations have been reported to be similar [23]. We use this knowledge and extract the statistical properties of the eye movements for biometric identification. Different stages of the algorithm are described below.

3.1. Data pre-processing and noise removal

The data contains visual angles in both x and y directions along with stimulus angles. Information about the validity of samples is also available. Eye movement data has been captured at a sampling frequency of 1000 Hz. The data obtained is decimated to 250 Hz using an anti-aliasing filter. In the proposed feature extraction method, most of the parameters are computed with reference to the screen coordinate system. Hence, in the pre-processing stage, the data obtained is converted to screen coordinates based on head distance and geometry of the acquisition system as

$$x_{screen} = \left(\frac{d * w_{pix}}{w} \right) \tan(\theta_x) + \frac{w_{pix}}{2} \quad (1)$$

$$y_{screen} = \left(\frac{d * h_{pix}}{h} \right) \tan(\theta_y) + \frac{h_{pix}}{2} \quad (2)$$

where, d , θ_x and θ_y denote distance from the screen and visual angles in x and y direction (in radian) respectively. x_{screen} and y_{screen} denote the position of gaze on the screen. w_{pix} , h_{pix} , w , h denote resolution and physical size of the screen in horizontal and vertical directions respectively.

Raw eye gaze positions may contain noise. Most of the features used in this work are extracted from velocity and acceleration profiles. The presence of noise makes it difficult to estimate the velocity and acceleration parameters using differentiation operation. Eye movement signals contain high-frequency components, especially during saccades. High-frequency components would be more prominent in velocity and acceleration profiles [24]. Savitzky–Golay filters are useful for filtering out the noise when the frequency span of the signal is large [25]. They are reported to be optimal [26] for minimizing the least-square error in fitting a polynomial to frames of the

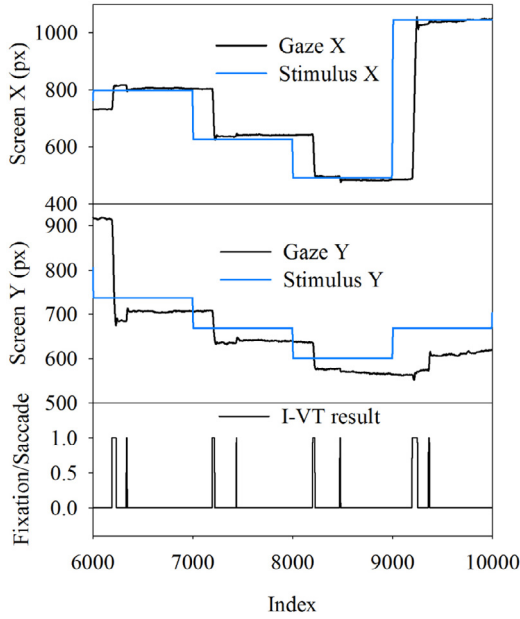


Fig. 1. Gaze data and stimulus for RAN_30min sequence.

noisy data. We use this filter with polynomial order of 6 and frame size of 15 in our approach (Fig. 1).

3.2. Eye movement classification and feature extraction

3.2.1. Eye movement classification

The I-VT (velocity threshold) algorithm [27,28] is used to classify the filtered eye movement data into a sequence of fixations and saccades (Algorithm 1). Most of the earlier works specify the velocity threshold for angular velocity. The angular velocity computed from the filtered data is used to classify the eye movements. A velocity of 50° per second is used as the threshold in I-VT algorithm.

A minimum duration threshold of 100 ms has been chosen to reduce the false positives in fixation identification. Algorithm 1 returns the classification results for each data point as either fixation or saccade. Points that are not a part of fixations are considered as saccades in this stage. In the proposed approach, we consider saccades with their durations more than a specified threshold to minimize the effect of spurious saccade segments. From the results of Algorithm 1, a list containing starting index and duration of all fixations and saccades is created. A post-processing stage is carried out to remove small-duration saccades. Saccades with duration less than 12 ms are removed in this stage.

3.2.2. Feature extraction

After the removal of small-duration saccades, each eye movement data is arranged into a sequence of fixations and saccades. The sequence of gaze locations and corresponding visual angles are also available for each fixation and saccade. Several statistical features are extracted from the position, velocity and acceleration profiles of the gaze sequence. Other features like duration, dispersion, path length and co-occurrence features are also extracted for both fixations and saccades. Earlier works [13] suggested that saccades provide a rich amount of information about the dynamics of oculomotor plant. Hence, we extract several other parameters including the saccadic ratio, main sequence, angle, etc. Saccades in horizontal and vertical directions are generated by different areas of the brain [29]. We use the statistical properties of the gaze data in x and y directions to incorporate this information. The distance and angle with the previous fixation/saccade are also used as features to leverage the

Algorithm 1: Fixation and Saccade classification algorithm.

Data: [Time Gazex Gazey]
Result: Res
Constants: VT=Velocity threshold, MDF=Minimum duration for fixation;
States=[FIXATION,SACCAD];
 fixationStart=1;
 Velocity=smoothDiff(data);
 N ← Number of samples of data;
for index ← 1 **to** N **do**
 if Velocity[index] < VT **then**
 currentState=FIXATION;
 if lastState ≠ currentState **then**
 fixationStart = index;
 end
 else
 if lastState= FIXATION **then**
 duration = data(index,1) - data(fixationStart,1);
 if duration < MDF **then**
 for i ← fixationStart **to** index **do**
 res[i]= SACCAD;
 end
 end
 end
 currentState=SACCAD;
 end
 lastState=currentState;
 res[index]=currentState;
end
 Res←res;

temporal properties. The method used for computation of features is described below.

Let $X = \{x_1, x_2, x_3, \dots, x_N\}$ and $Y = \{y_1, y_2, y_3, \dots, y_N\}$ denote the set of coordinate positions of gaze in each fixation/saccade and let N denotes the number of data points in any fixation or saccade. (x_i, y_i) denotes gaze location on the screen coordinate system and (θ_i^x, θ_i^y) denotes the corresponding horizontal and vertical visual angles.

A large number of features are extracted from the gaze sequence in each fixation and saccade. Some features are derived from the angular velocity. The differentiation operation for finding velocity and acceleration is carried out using forward difference method on the smoothed data. List of features extracted from fixations and saccades along with the methods of computation are shown in Tables 1 and 2. The features are extracted independently for each fixation and saccade.

The control mechanisms generating fixations and saccades are different. The number of fixations and saccades is also different in each recording. There is a total of 12 and 46 features extracted from fixations and saccades respectively. A feature normalization scheme is used to scale each feature into a common range to ensure equal contribution in the final classification stage.

3.2.3. Feature selection

The large number of features extracted may contain redundancy and correlation. A backward feature selection algorithm, as shown in Algorithm 2 is used to retain a minimal set of discriminant features. We use the wrapper-based approach [30] for selecting the features. An RBFN classifier is used for finding the Equal Error Rate (EER) in each iteration. Cross-validation has been carried out in the training set to avoid overfitting. We used a random 50% subset of the development dataset for the feature selection algorithm. Feature selection algorithm starts with a set of all the features. Now in each iteration,

Table 1
List of features extracted from fixations.

Used in TEX	RAN	Fixation features	Description
N	Y	Fixation duration	Obtained from I-VT result
N	N	Standard deviation (X)	From the screen coordinates during fixation
Y	N	Standard deviation (Y)	"
Y	Y	Path length	Length of path traveled in screen $path\ length = \sum_{i=1}^{N-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$
Y	Y	Angle with previous fixation	Angle with centroid of previous fixation
Y	Y	Distance from the last fixation	Euclidean distance from the previous fixation
Y	Y	Skewness (X)	From screen coordinates
Y	Y	Skewness (Y)	"
N	N	Kurtosis (X)	"
Y	Y	Kurtosis (Y)	"
Y	Y	Dispersion	Spatial spread during a fixation, computed as $D = (\max(X) - \min(X)) + (\max(Y) - \min(Y))$
Y	Y	Average velocity	$AV = path\ length / duration$

Y and N denote inclusion or exclusion of the feature in the particular stimulus after feature selection.

Table 2
List of features extracted from saccades.

Used in TEX	RAN	Saccade features	Description
N	N	Saccadic duration	Obtained from I-VT result
Y	Y	Dispersion	$D = (\max(X) - \min(X)) + (\max(Y) - \min(Y))$, during saccade
NNNNYY	NNNNYY	M3S2K (Angular velocity)	Features from angular velocity
YYYYYN	YYYYYY	M3S2K (Angular acceleration)	Features from angular acceleration
Y	Y	Standard deviation (X)	Obtained from screen positions
Y	Y	Standard deviation (Y)	"
Y	Y	Path length	Distance traveled in screen, $\sum_{i=1}^{N-1} \sqrt{(x_{i+1} - x_i)^2 + (y_{i+1} - y_i)^2}$
Y	Y	Angle with previous saccade	Difference in saccadic angle with previous saccade
Y	Y	Distance from the previous saccade	Euclidean distance between the centroid of the previous saccade
Y	Y	Saccadic ratio	$SR = \max(angular\ velocity) / saccade\ duration$
Y	Y	Saccade angle	Obtained from first and last points as, $saccade\ angle = \tan^{-1} \left(\frac{y_N - y_1}{x_N - x_1} \right)$
Y	Y	Saccade amplitude	Obtained as: $\sqrt{(x_N - x_1)^2 + (y_N - y_1)^2}$
YYYYYY	YYYYYY	M3S2K (Velocity_X_direction)	Features from screen positions
YYYYYY	YYYYNY	M3S2K (Velocity_Y_direction)	"
YYYYYY	YYYYYY	M3S2K (Acceleration_X_direction)	"
YYYYYY	YNNYYY	M3S2K (Acceleration_Y_direction)	"

*M3S2K – Statistical features:

Mean, Median, Max, Std, Skewness, Kurtosis Y and N denote inclusion or exclusion of the feature in the particular stimulus after feature selection.

Algorithm 2: Backward feature selection.

```

Data: Feature matrix
Result: featureList[1: Included,0:Excluded]
N ← Number of features;
featureList ← ones(N);
for i ← 1 to N do
    W ← featureList;
    E ← +Inf;
    for j ← 0 to 1 do
        W[i] ← j;
        T ← EER with included features using RBFN;
        if T < E then
            featureList[i] ← j;
            E ← T;
        end
    end
end

```

the EER with inclusion and exclusion of a particular feature is found. The feature is retained if the EER with the use of the feature is better than EER with exclusion. The procedure is repeated for all the features in a sequential manner. The feature selection algorithm is iterated ten times each time on a random 50% subset for cross-validation. After these iterations, a set of important features is retained. To evaluate the generalization ability of the selected features, we have tested

the algorithm (with the selected features) on an entirely disjoint set that was not used in the feature selection process. The results with the evaluation set [22] (as shown in the public results of BioEye 2015 competition) show the stability and generalization capability of the selected features. The subset of features selected were different for different stimuli (TEX and RAN sets). The list of features selected for TEX and RAN stimuli is shown in Table 1 (Fixation features) and Table 2 (Saccade features). The features thus selected are used as inputs to the classification algorithm.

After obtaining the set of features from fixations and saccades, we develop a model to represent the data. It has been empirically observed that the performance of classification approaches with Kernel-based methods is better than linear classifiers. It has also been reported that the parameters like amplitude-duration and amplitude-peak velocity may vary with the angle of saccade [31]. The nature of saccade dynamics may be different in different directions as the stimulus is changing randomly at various points on the screen. For each person, saccades of different amplitudes and directions form clusters in the feature space. In order to use the multi-mode nature of the data, we represent them by clustering them in the feature space. Representative vectors from each cluster are used to characterize each person. We use Gaussian Radial Basis Function Network (GRBFN) to model these data. The multiple cluster centers in the feature space are used as representative vectors in this approach. These vectors are selected using the K-means algorithm. Two different RBFNs are trained separately for fixation and saccade. Details about the structure of network and score fusion stage are described in the following section.

3.3. RBF network

Radial Basis Function Network (RBFN) is a class of neural networks initially proposed by Broomhead and Lowe [32]. Classification in RBFN is done by calculating the similarity between training and test vectors. Multiple prototype vectors corresponding to each class are stored in each neuron. The Euclidean distance between the input vector and the prototype vector is used to calculate neuron activations.

In the RBF network, input layer is made of feature vectors. $\varphi(x)$ is a radial basis function that finds the Euclidean distance between the input vector and the prototype vector. A weighted combination of scores from the RBF layer is used to classify the input into different categories.

The number of prototypes per class can be defined by the user, and these vectors can be found from the data using different algorithms like K-means, Linde–Buzo–Gray (LBG) algorithm, etc.

The Gaussian activation function of each neuron is chosen as

$$\varphi(x) = e^{-\beta \|x - \mu\|^2} \quad (3)$$

where, μ is the mean of the distribution. The parameter β can be found from the data.

In this work, we have used K-means algorithm for selecting the representative vectors. For each person, 32 clusters for fixations and 32 cluster centers for saccades are kept, resulting in $32N$ clusters for each RBFN (where N is the number of persons in the dataset). The number of clusters to keep is obtained empirically. We have clustered the fixations/saccades of each individual separately to obtain a fixed number of representative vectors for each person. A maximum of 100 iterations is used to form the clusters. A standard K-means algorithm is used with squared Euclidean distance, and the centers are updated in each iteration. Each data point is assigned to the closest cluster center obtained from the K-means algorithm. For a particular neuron, the value of β is computed from the distance of all points belonging to that particular cluster as

$$\beta = \frac{1}{2\sigma^2} \quad (4)$$

where σ is the mean Euclidean distance of the points (assigned to the specific neuron) from the centroid of the corresponding cluster.

3.3.1. Notations

The biometric identification problem is similar to a multiclass classification problem. Let there be n samples of a p dimensional data. Assume there are m classes (corresponding to m different individuals) with c samples per class ($n = mc$). Let y_i be the label corresponding to i th sample. Let K be the number of representative vectors from each class. The value of K is chosen empirically ($K = 32$).

3.3.2. Network learning

The activations can be obtained as: $A = \varphi_{i,j}(x_k)$, $i = 1, \dots, K$, $j = 1, \dots, m$, $k = 1, \dots, n$

The output of the network can be represented as a linear combination of the RBF activations as

$$f(x) = \sum_{j=1}^m w_j \varphi_j(x) \quad (5)$$

where, $f(x)$ contains the class membership in vector form. Given the activations and output labels, the objective of the training stage is to find the weight parameters of the output layer. The weights are obtained by minimizing the sum of squared errors.

The output layer is represented by a linear system as

$$A\hat{w} = \hat{y} \quad (6)$$

The optimal set of weights can be found using the Moore–Penrose pseudoinverse. Alternatively, these weights can be learned through

gradient descent method. In the learning phase, features extracted from each fixation and saccade are used to train the model. Each fixation/saccade is treated as a sample in the training process.

The method described here uses two-phase learning. RBF layer and weight layer trainings are carried out separately. However a joint training similar to back-propagation is also possible [33].

3.3.3. Training stage

Only the session 1 data from the datasets are used in the training stage. Cluster centers and corresponding β values are computed separately for each person (resulting in $32N$ neurons for both fixation and saccade RBFNs). The output weights (\hat{w}_{fix} and \hat{w}_{sacc}) are found using all fixations and saccades from all the subjects in the dataset.

3.3.4. Testing stage

Session 2 data is used in the testing stage. Parameters of RBFN are computed separately for fixations and saccades in the training session. The scores from both RBFNs are combined to obtain the final result. The overall configuration of the scheme is shown in Fig. 2.

For an unlabeled probe, the activations for each fixation and saccade (A_{fix} and A_{sacc}) are found separately using the cluster centers obtained in the training stage. The final classification is carried out using the combined score obtained from all saccades and fixations. Let n_{fix} and n_{sacc} be the number of fixations and saccades in an unlabeled gaze sequence. The combined score can be obtained as

$$score = \lambda \frac{1}{n_{fix}} \sum_{i=1}^{n_{fix}} A_{fix}^i \hat{w}_{fix} + (1 - \lambda) \frac{1}{n_{sacc}} \sum_{i=1}^{n_{sacc}} A_{sacc}^i \hat{w}_{sacc} \quad (7)$$

where, $\lambda \in [0, 1]$ is the weight used in the score fusion. The parameter λ decides the contribution of fixations and saccades in the final decision stage. This value can be obtained empirically. In the present work, λ value of 0.5 is used.

The label of the unknown sample can be obtained as

$$label = \arg \max_m (score) \quad (8)$$

4. Experiments and results

4.1. Datasets

The data used in this paper are part of the development phase of BioEye 2015 [22] competition. Data recorded in three different sessions are available. First two sessions are separated by a time interval of 30 min containing recordings of 153 subjects (ages 18–43). A third session, conducted after 1 year, (37 subjects) is also available to evaluate the robustness against template aging. The database contains gaze sequences obtained using two distinct types of visual stimuli. In one set (RAN), a white dot moving in a dark background was used as the stimulus. The subjects were asked to follow the dot. Text excerpt shown on the screen was used as the stimulus in the other set (TEX). The samples were recorded with an EyeLink eye-tracker (with a reported spatial accuracy of 0.5°) at 1000 Hz and down-sampled to 250 Hz with anti-aliasing filtering. The development dataset contains the ground truth about the identity of the persons. An additional evaluation set is also available without ground truth.

In each recording, visual angles in x and y direction, stimulus angle in x and y direction and information regarding the validity of the samples are available. Details about the stimulus types in BioEye2015 database are given below.

4.1.1. Random dot stimulus (RAN_30min & RAN_1year)

The stimulus used was a white dot appearing at random locations on a black computer screen. The position of the stimulus would change every second. The subjects were asked to follow the dot on the screen and recording was carried out for 100 s.

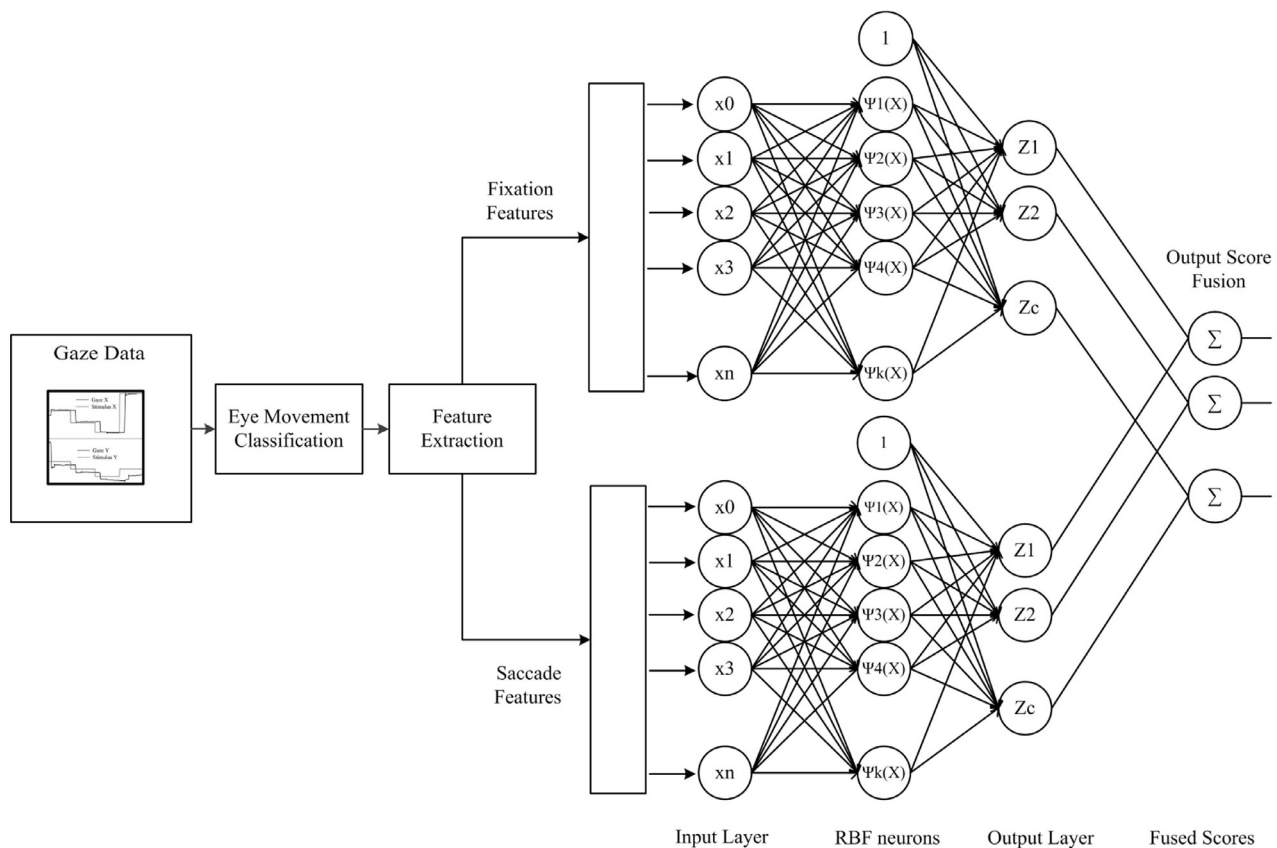


Fig. 2. Schematic of the proposed framework.

Table 3
Details about the database.

Dataset name	RAN_30min	RAN_1year	TEX_30min	TEX_1year
Number of subjects	153	37	153	37
Stimulus	White dot moving in a dark background	White dot moving in a dark background	Text excerpt	Text excerpt
Duration of experiment	100 s	100 s	60 s	60 s
Interval between training and testing data	30 min	1 year	30 min	1 year

4.1.2. Text stimulus (TEX_30min & TEX_1year)

The task, in this case, was reading text excerpts from the poem of Lewis Carroll “The Hunting of the Snark”. The duration of this experiment was 60 s.

A comprehensive list of the datasets and parameters are shown in Table 3.

4.2. Evaluation metrics

The proposed algorithm has been evaluated in the labeled development set. Rank-1 accuracy and EER are used for evaluating the algorithm. Rank-1 (R1) accuracy is defined as the ratio of the total number of correct detections to the number of samples used. EER is the percentage at which False Acceptance Rate (FAR) and False Rejection Rate (FRR) are equal. Detection Error Trade-off (DET) curves are shown for all the datasets. Rank(n) accuracy is the number of correct detections in the top n candidates. Cumulative match characteristics (CMC) is the cumulative plot of rank(n) accuracy. CMC curves are also plotted for all the four datasets. The evaluation set in the BioEye2015 dataset is unlabeled. However, we report the R1 accuracy as obtained from the public results [22] of the competition.

4.3. Results

4.3.1. Performance in the development datasets

The model was trained using 50% of data in the development datasets. We have trained and tested the algorithm on completely

disjoint sessions to test its generalization ability. For example, in RAN_30min sequence there are 153 samples available for two different sessions. We have trained the Algorithm only on the first session (using a random 50% subset of the data). The evaluation was carried out on the session 2 data. We have not used the data from the same session for training and testing since it won't account for intersession variability.

The average R1 accuracy and EER were calculated from random 50% subsets of development datasets. This procedure was repeated 100 times and the average R1 accuracy and EER were obtained. The results obtained along with the standard deviations are given in Table 4.

The R1 accuracy in RAN_30min and TEX_30min databases are above 90% indicating the robustness of the proposed framework. The EER on RAN_30min database is found out to be 2.59%, comparable to the accuracy levels of fingerprint (2.07% EER) [34], voice recognition systems, and facial geometry (15% EER) [35] biometrics.

Table 4
Results in the development datasets.

	RAN_30	RAN_1yr	TEX_30	TEX_1yr
R1	90.10 ± 2.76	79.31 ± 6.86	92.38 ± 2.56	83.41 ± 6.98
EER	2.59 ± 0.71	10.96 ± 4.59	3.78 ± 0.77	9.36 ± 3.49

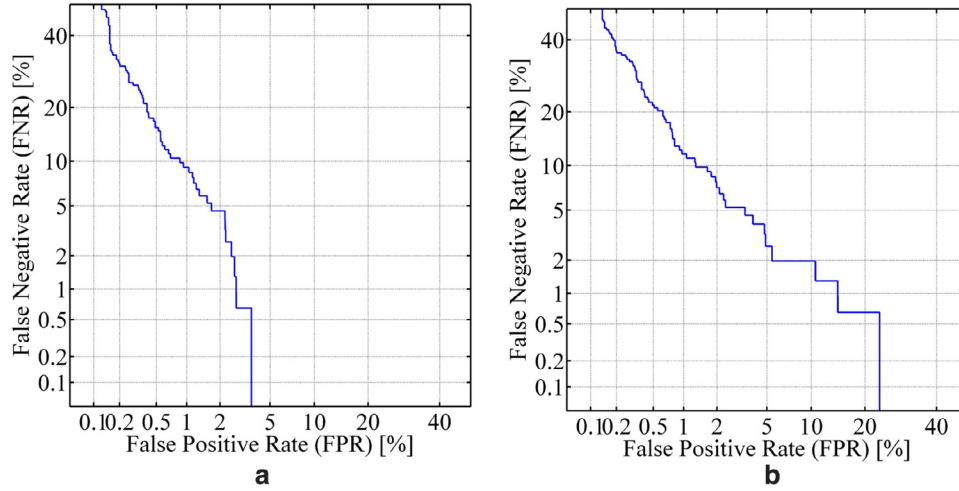


Fig. 3. DET curve for (a) RAN_30min and (b) TEX_30min.

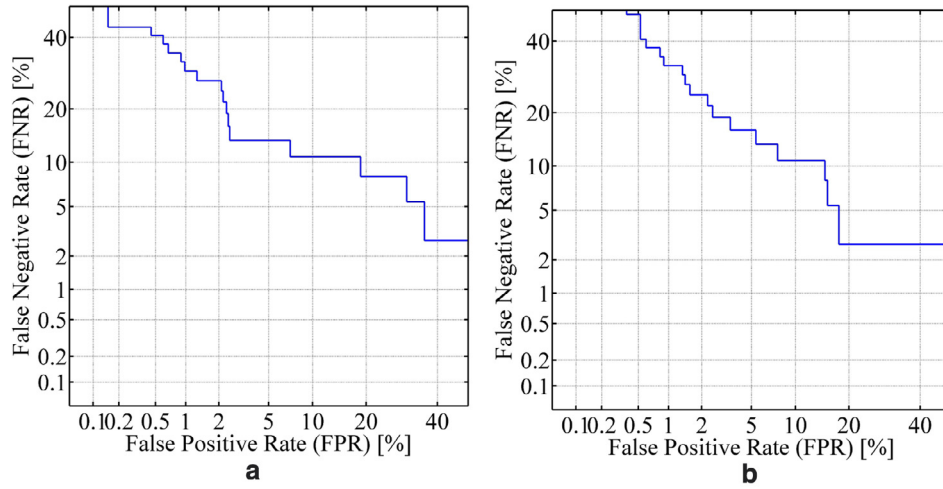


Fig. 4. DET curve for (a) RAN_1year and (b) TEX_1year.

Table 5

Comparison of R1 accuracy in the entire development dataset.

	RAN_30	RAN_1yr	TEX_30	TEX_1yr
Our method (%)	89.54	81.08	85.62	78.38
Baseline (%)	40.52	16.22	52.94	40.54

R1 accuracy (Table 5) of the proposed algorithm obtained from the development set was compared with the baseline algorithm (CEM-B) [21]. The average cumulative matching characteristics curves for the four datasets are shown in Figs. 5 and 6.

The Detection Error Trade-off (DET) curves for the development datasets are shown in Figs. 3 and 4. In Fig. 3(a) and (b), FNR becomes very small as FPR increases indicating a good separation from impostors. The reduction in FNR may be because of the addition of scores of all the fixations and saccades in the score fusion stage. Impostor scores are considerably smaller than genuine scores in the proposed approach. The performance in 1-year sessions is poor compared to 30-min sessions indicating template aging effects.

4.3.2. Performance in the evaluation sets

The evaluation part of the database is unlabeled. However, the results of the competition are available on the website [22]. The evaluation set of the dataset had only one unlabeled data for every labeled

sample. We have used this one to one correspondence assumption in the final stage of the algorithm.

Let there be n labeled and n unlabeled recordings. The task is to assign each unlabeled file to a labeled file. The scores obtained from RBF output stage were stored in a matrix D (with dimension $n \times n$). $D(i, j)$ denotes the normalized similarity score between i th labeled and j th unlabeled samples. We have selected the best match for each unlabeled recording using Algorithm 3. The use of this one to one assumption improved the results. However, this assumption may not be suitable for practical biometric identification/verification scenarios.

Algorithm 3: One to one matching.

Data: D (Score matrix)

Result: Matches

$[n, n] = \text{size}(D)$;

for $i \leftarrow 1$ **to** n **do**

$[\text{row}, \text{col}] = \text{find}(D == \max(D(:)))$;

$D(\text{row}, :) = -\infty$;

$D(:, \text{col}) = -\infty$;

$\text{pair} = [\text{row}, \text{col}]$;

$\text{Matches.append}(\text{pair})$;

end

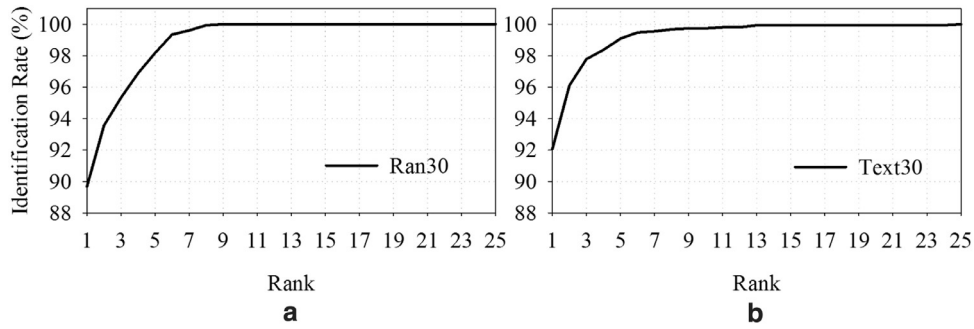


Fig. 5. CMC curve for (a) RAN_30min and (b) TEX_30min.

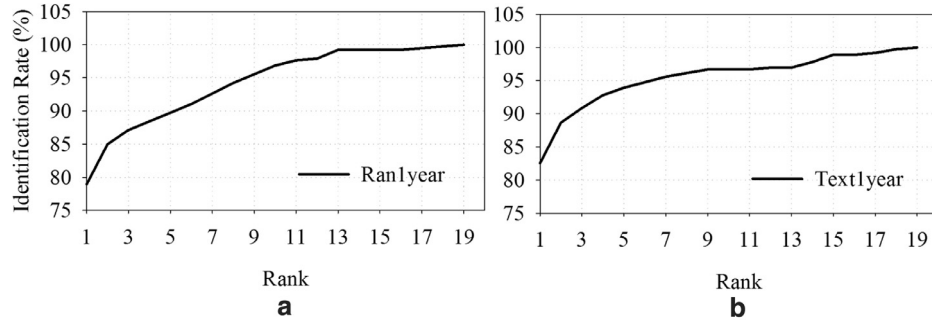


Fig. 6. CMC curve for (a) RAN_1year and (b) TEX_1year.

Table 6
Comparison of R1 accuracy with baseline method in evaluation dataset.

	RAN_30	RAN_1yr	TEX_30	TEX_1yr
Our method (%)	93.46	83.78	89.54	83.78
Our method* (%)	98.69	89.19	98.04	94.59
Baseline (%)	33.99	40.54	58.17	48.65

* With one to one assumption.

The proposed method has been found to outperform all the other methods even without the one to one assumption indicating the robustness for biometric applications. The results with and without this assumption are shown in Table 6.

4.4. Computational complexity

The algorithm has been implemented in an Intel Core i5 CPU, 3.33 GHz PC with 4 GB RAM. The average training time for the network without code optimization (single-threaded) in MATLAB is about 400 s (with 153 samples). In the testing phase, for predicting one unlabeled recording, it takes on an average 0.21 s (in TEX_30min). The time taken for training and testing phase can be improved considerably by implementation in C, C++, using parallel processing platforms like Graphical Processing Units (GPU).

4.5. Discussions

4.5.1. Performance of the algorithm

The R1 accuracy of the proposed method is high in both TEX and RAN datasets, which indicates the possibility of developing a task-independent biometric system. The EER and R1 accuracy achieved show the robustness of the proposed score fusion approach. The selected features show good discrimination ability in both stimuli. The accuracy with 1-year datasets is comparatively lesser than that with the 30-min datasets. This lower accuracy may be attributed to template aging effects. Some of the selected features may show variability over time [36,37].

The feature selection was carried out in 30-min datasets due to the availability of a large number of subjects. Feature selection with 1-year datasets may lead to overfitting because of fewer subjects. This issue can be solved by using the feature selection in 1-year datasets with a larger number of subjects, which may identify features that are robust against template aging. However, the results show significant improvement compared to the state of the art methods. The proposed algorithm was ranked first in the BioEye 2015 [22] competition.

4.5.2. Limitations

Controlled experimental setup was used to collect the data used in this work. The sampling rate and quality of data used in the present work were very high since it was collected in lab conditions using chinrest. Accurate estimation of the features in noisy, low sampling rates is necessary for the use in a practical biometric scenario. The nature of eye movements may be affected by the level of alertness, fatigue, emotions, cognitive loading, etc. Consumption of caffeine and alcohol by the subjects may affect the performance of the proposed algorithm. The features selected for biometrics should be invariant to such variations. Only two sessions of data were available for each subject. Intersession variability and template aging effects need to be studied further. Lack of publicly available databases containing a large number of samples (to account for template aging, uncontrolled environment, affective states, intersession variability) is another problem. Creation of a large database with such variability could provide more robust solutions.

5. Conclusions

This work proposes a novel framework for biometric identification based on dynamic characteristics of eye movements. The raw eye movement data is classified into a sequence of fixations and saccades. We extract a large set of features from fixations and saccades to characterize each individual. The important features extracted from fixations and saccades are identified based on a backward selection framework. Two different Gaussian RBF networks are trained using features from fixations and saccades separately. In the detection

phase, scores obtained from both RBF networks are used to get the subject's identity. The high accuracy obtained shows the robustness of the proposed algorithm. The proposed framework can be easily integrated into the existing iris recognition systems. A combination of the proposed approach with conventional iris recognition systems may give rise to a new counterfeit-resistant biometric system. The comparable accuracy in distinct types of stimuli indicates the possibility of developing a task-independent system for eye movement biometrics. The proposed method can also be used for continuous authentication in desktop environments. Robustness of the algorithm against lower sampling rates, calibration error and noise can be explored in future. The effect of duration of data on the level of accuracy is another topic to be investigated.

Acknowledgments

The authors would like to thank the organizers of BioEye 2015 competition for providing the data.

References

- [1] A.K. Jain, P. Flynn, A.A. Ross, *Handbook of Biometrics*, Springer Science & Business Media, 2007.
- [2] A.K. Jain, A. Ross, S. Prabhakar, An introduction to biometric recognition, *IEEE Trans. Circuits Syst. Video Technol.* 14 (1) (2004) 4–20.
- [3] L. Wang, X. Geng, L. Wang, X. Geng, *Behavioral Biometrics For Human Identification: Intelligent Applications*, IGI Global, 2009.
- [4] S. Marcel, J.D.R. Millán, Person authentication using brainwaves (EEG) and maximum a posteriori model adaptation, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (4) (2007) 743–752.
- [5] K.N. Plataniotis, D. Hatzinakos, J.K. Lee, ECG biometric recognition without fiducial detection, in: *Proceedings of the Special Session on Research at the Biometric Consortium Conference, 2006 Biometrics Symposium*, IEEE, 2006, pp. 1–6.
- [6] I.B. Group, *Independent Testing of Iris Recognition Technology*, Final Report, NBCHC030114/0002, 2005.
- [7] C. Roberts, Biometric attack vectors and defences, *Comput. Secur.* 26 (1) (2007) 14–25.
- [8] S. Schuckers, L. Hornak, T. Norman, R. Derakhshani, S. Parthasaradhi, Issues for liveness detection in biometrics, in: *Proceedings of the IEEE Biometric Consortium Conference*, New York, 2002.
- [9] R.J. Leigh, D.S. Zee, *The Neurology of Eye Movements*, vol. 90, Oxford University Press, New York, 1999.
- [10] T. Kinnunen, F. Sedlak, R. Bednarik, Towards task-independent person authentication using eye movement signals, in: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, 2010, pp. 187–190.
- [11] P. Kasprowski, J. Ober, Eye movements in biometrics, in: *Biometric Authentication*, Springer, 2004, pp. 248–258.
- [12] R. Bednarik, T. Kinnunen, A. Mihaila, P. Fränti, Eye-movements as a biometric, in: *Image Analysis*, Springer, 2005, pp. 780–789.
- [13] O.V. Komogortsev, S. Jayarathna, C.R. Aragon, M. Mahmoud, Biometric identification via an oculomotor plant mathematical model, in: *Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications*, ACM, 2010, pp. 57–60.
- [14] O.V. Komogortsev, A. Karpov, L.R. Price, C. Aragon, Biometric authentication via oculomotor plant characteristics, in: *Proceedings of the 5th IAPR International Conference on Biometrics (ICB)*, IEEE, 2012, pp. 413–420.
- [15] C.D. Holland, O.V. Komogortsev, Complex eye movement pattern biometrics: the effects of environment and stimulus, *IEEE Trans. Inf. Forensics Secur.* 8 (12) (2013) 2115–2126.
- [16] I. Rigas, G. Economou, S. Fotopoulos, Biometric identification based on the eye movements and graph matching techniques, *Pattern Recognit. Lett.* 33 (6) (2012a) 786–792.
- [17] I. Rigas, G. Economou, S. Fotopoulos, Human eye movements as a trait for biometric identification, in: *Proceedings of the IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, IEEE, 2012b, pp. 217–222.
- [18] Y. Zhang, M. Juhola, On biometric verification of a user by means of eye movement data mining, in: *Proceedings of the Second International Conference on Advances in Information Mining and Management (IMMM 2012)*, 2012, pp. 85–90.
- [19] V. Cantoni, C. Galdi, M. Nappi, M. Porta, D. Riccio, Gant: gaze analysis technique for human identification, *Pattern Recognit.* 48 (4) (2015) 1027–1038.
- [20] C. Holland, O.V. Komogortsev, Biometric identification via eye movement scan-paths in reading, in: *Proceedings of the International Joint Conference on Biometrics (IJCB)*, IEEE, 2011, pp. 1–8.
- [21] C.D. Holland, O.V. Komogortsev, Complex eye movement pattern biometrics: analyzing fixations and saccades, in: *Proceedings of the International Conference on Biometrics (ICB)*, IEEE, 2013, pp. 1–8.
- [22] BioEye 2015, BTAS, 2015, Competition on biometrics via eye movements, <https://bioeye.cs.txstate.edu/>, (accessed 10.12.15).
- [23] H. Collewijn, C.J. Erkelens, R. Steinman, Binocular co-ordination of human horizontal saccadic eye movements, *J. Physiol.* 404 (1) (1988) 157–182.
- [24] C.M. Harris, I. Abramov, L. Hainl, Instrument considerations in measuring fast eye movements, *Behav. Res. Methods Instrum. Comput.* 16 (4) (1984) 341–350.
- [25] S.R. Krishnan, C.S. Seelamantula, On the selection of optimum Savitzky–Golay filters, *Signal Process.* 61 (2) (2013) 380–391.
- [26] A. Savitzky, M.J. Golay, Smoothing and differentiation of data by simplified least squares procedures, *Anal. Chem.* 36 (8) (1964) 1627–1639.
- [27] C.D. Holland, O.V. Komogortsev, Biometric verification via complex eye movements: the effects of environment and stimulus, in: *Proceedings of the 2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, IEEE, 2012, pp. 39–46.
- [28] D.D. Salvucci, J.H. Goldberg, Identifying fixations and saccades in eye-tracking protocols, in: *Proceedings of the 2000 Symposium on Eye Tracking Research & Applications*, ACM, 2000, pp. 71–78.
- [29] M.R. Harwood, J.P. Herman, Optimally straight and optimally curved saccades, *J. Neurosci.* 28 (30) (2008) 7455–7457.
- [30] R. Kohavi, G.H. John, Wrappers for feature subset selection, *Artif. Intell.* 97 (1) (1997) 273–324.
- [31] H.H. Goossens, A. Van Opstal, Human eye-head coordination in two dimensions under different sensorimotor conditions, *Exp. Brain Res.* 114 (3) (1997) 542–560.
- [32] D.S. Broomhead, D. Lowe, Radial basis functions, multi-variable functional interpolation and adaptive networks, Technical Report, DTIC Document, 1988, (accessed 10.12.15).
- [33] F. Schwenker, H.A. Kestler, G. Palm, Three learning phases for radial-basis-function networks, *Neural Netw.* 14 (4) (2001) 439–458.
- [34] D. Maio, D. Maltoni, R. Cappelli, J.L. Wayman, A.K. Jain, FVC2004: third fingerprint verification competition, in: *Biometric Authentication*, Springer, 2004, pp. 1–7.
- [35] P.J. Phillips, W.T. Scruggs, A.J. O'Toole, P.J. Flynn, K.W. Bowyer, C.L. Schott, M. Sharpe, FRVT 2006 and ICE 2006 large-scale experimental results, *IEEE Trans. Pattern Anal. Mach. Intell.* 32 (5) (2010) 831–846.
- [36] O.V. Komogortsev, C.D. Holland, A. Karpov, Template aging in eye movement-driven biometrics, in: *SPIE, Biometric and Surveillance Technology for Human and Activity Identification XI*, 2014, pp. 90750A–90750A-9.
- [37] P. Kasprowski, The impact of temporal proximity between samples on eye movement biometric identification, in: *Computer Information Systems and Industrial Management*, Springer, 2013, pp. 77–87.