BioEye 2015: Competition on Biometrics via Eye Movements

Oleg V. Komogortsev and Ioannis Rigas
Department of Computer Science
Texas State University, San Marcos, USA
ok11@txstate.edu, rigas@txstate.edu

Abstract

Biometric recognition via eye movement-driven features is an emerging field of research. Eye movement cues are characterized by their non-static nature, the encapsulation of physical and behavioral traits, and the possibility to be recorded in tandem with other modalities, e.g. the iris. The BioEye 2015 competition was organized with the aim to boost the evolution of the eye movement biometrics field. The competition was implemented with a particular focus on the issues facing the researchers in the domain of the eye movement recognition, e.g. quality of the eye movement recordings, different visual stimulus types, and the effect of template aging on the resulting recognition accuracy. This paper describes the details and the results of the BioEye 2015 competition, which provided the largest to date biometric database containing records from 306 subjects, stimulus of two types, and recordings separated by short-time and long-time intervals.

1. Introduction

Nature endows the human beings with an admirable diversity of characteristics suitable to be used for the task of biometric recognition. Some of these features, e.g. the fingerprints, are quite distinctive and have drawn the scientific interest for over a century [1]. The advent of the computer technology during the second half of the last century facilitated a more systematic research on various other traits such as the face, the iris, and many others [2].

The application of biometric recognition was initially restricted in the fields of forensic analysis and law enforcement. Nowadays, biometric recognition is gradually becoming part of our everyday life. The wide adoption of biometrics is strongly supported by the changes related to the transference and storage of the private data, the automation of the commercial transactions, and the inherent difficulties faced by the traditional security methods such as passwords. Although the fingerprints, face, and iris appear to be the most frequently used modalities, their standalone utilization is subject to various drawbacks, e.g. bad image quality issues, and vulnerability to spoofing attacks. Thus, their joint use along with other modalities can produce more robust systems, and lead to a wider adoption of the biometric applications.

Although the concept of using the eye movements for biometric recognition is very recent [3], the research of the various properties of the eye movements is more than a century old. During the early studies of Javal [4], it was observed that during a task of reading the eyes make quick movements (called saccades) from one point of focus to another (called fixations). Many studies explored the interconnections of the eye movements with the cognitive activity and the guidance of visual attention, e.g. [5, 6]. Also, the studies of Noton and Stark [7, 8] led to the formation of the so-called Scanpath Theory, which supported both the context-based and the idiosyncratic character of the viewing behaviors.

Eye movements provide the opportunity to decode the information about brain zones that participate in the visual processing and generation of the neural commands for the oculomotor plant, represented by the extraocular muscles, the eye globe, and surrounding tissues. Both the brain and the oculomotor components are well hidden from the external exposure, thus making it almost impossible to conduct replication for a successful spoofing attack [9]. Also, eye movements can be recorded remotely without the requirement of touching the device or extracting any substance from the human body. Although the behavioral nature of the eye movements can limit the permanence of the extracted features, it can provide advantages in terms of the anti-spoofing robustness. It is also important to note that eye movement features can be employed for the detection of health states, e.g. concussions [10], thus forming a use case where a biometric framework is employed not only for the identity recognition but also for monitoring the health condition of the human being. Last, the fact that eye movements are recorded from the face region makes possible their incorporation in multi-modal systems along with iris, face, and peri-ocular features.

The BioEye 2015 (www.bioeye.info) competition was organized to advance the research domain of the eye movement biometrics. The main contributions of the BioEye 2015 competition can be summarized as follows:

1) The competition database is the largest publically available dataset with records from 306 subjects.

2) The eye movement recordings were conducted to achieve the highest possible quality of the recorded data.

3) The competition allowed the participants to validate their algorithms for eye movement recordings captured in response to different visual stimuli.

4) The competition allowed the participants to check the stability of their algorithms in the case of template aging.

2. Research on Eye Movement Biometrics

2.1. Prior Art

The idea for using the eye movements for biometric purposes was initially suggested by Kasprowski and Ober [3] who employed the Cepstrum transform for signal analysis. The employed visual stimulus was a 'jumping' point of light. The verification performance for a group of nine subjects revealed a 1.36% False Acceptance Rate (FAR), and a 12.59% False Rejection Rate (FRR). In the study of Bednarik et al. [11], eye movement features were extracted using the Fast Fourier Transform (FFT) and Principal Component Analysis (PCA). They used a variety of stimuli, e.g. static images, text, and a moving cross, and they tested the identification scenario reporting a Rank-1 Identification Rate (IR) of 56% for a database of 12 subjects. The work of Kinnunen et al. [12] was inspired by the similarities between eye movement and speech signals. They applied voice recognition techniques for the analysis of eye movements in a task-independent scenario involving various stimuli. The achieved Equal Error Rate (EER) for 17 subjects was about 30%. In the work of Komogortsev et al. [13], a mathematical model of the eye—Oculomotor Plant Mathematical Model (OPMM)—was created and tested in a task of identity recognition. The model evolved to a method for the recognition via Oculomotor Plant Characteristics (OPC) [14]. The experiments for the OPC biometrics were performed with a pool of 59 subjects and by using a 'jumping' point of light as the stimulus. The achieved Half Total Error Rate (HTER) was 19%. Liang et al. [15] employed the mutual information between features extracted from the eye tracking data, and tested their approach for a stimulus consisting of video clips. The identification accuracy for a small group of 5 subjects was 82%. The method presented by Rigas et al. [16] employed graph matching techniques for analyzing the spatial characteristics of fixations while observing images of human faces. The experiments were performed with 15 subjects, and yielded minimal EER of 30%. Utilization of graph representation and face images was adopted also in the recent work of Cantoni et al. [17]. The developed gaze analysis technique (GANT) was tested on a large set of 112 subjects, and the achieved EER was 25%. In the work of Holland and Komogortsev [18], a variety of features extracted from the eye movements were used to form the Complex Eye Movement Pattern (CEM-P) biometrics. For a pool of 32 subjects, the minimal EER was 28% and the best Rank-1 IR was 53%. The CEM-P features served as the basis for the Complex Eye Movement Behavior (CEM-B) biometrics [19], where eye movement features were represented as statistical distributions. The best results for a database of 32 subjects were a minimal EER of 16.5% and a maximum Rank-1 IR of 83.7%. In the study of Yoon et al. [20] a scheme based on Hidden Markov Models (HMM) was used to analyze gaze for cognitive stimulus consisted of dot-patterns. The best achieved Rank-1 IR was 76% for a set of 12 subjects. Rigas and Komogortsev [21] presented a method based on Fixation Density Maps (FDMs) for representing visual attention as a biometric template. They used a large database of 200 subjects with the resulting best EER of 12.1% and Rank-1 IR of 39.4%.

2.2. Previous Competitions

Two competitions for eye movement biometrics were organized previously, in an effort to introduce eye movement biometrics to a broader audience. The first Eye Movement Verification and Identification Competition (EMVIC 2012) [22] was organized in 2012 as a part of IEEE BTAS 2012. In all datasets, the used visual stimulus was a 'jumping' point of light. The four provided datasets were recorded from 37, 75, 29, and 27 subjects respectively. Two of the participants described their works for the competition datasets in scientific papers, Cuong et al. [23], and Rigas et al. [24] (second and third place in the competition). The second Eye Movement Verification and Identification Competition (EMVIC 2014) [25] was organized in 2014 as part of the IEEE IJCB 2014. In this case the visual stimulus consisted of face images, and the total number of recorded subjects was 34. The method that won the first place is described in the work of Monaco [26].

3. Competition Impact

Data collected in response to the BioEye 2015 website revealed an increased interest for the field of eye movement biometrics spanned in different regions of the world. The website had visitors from 54 different countries. There was a total of 64 registrations in the competition. Most of the registered users were members of academic institutions, whereas a smaller number was from research labs/centers, and industry/private companies. Seven from the registered users followed the final submission process and participated in the evaluation procedure. Collectively for all datasets, a total of 200 submissions were uploaded during the 26 days of the evaluation phase of the competition.

4. Competition Database

4.1. Recording Procedure

4.1.1   Recording Sessions
In order to facilitate the investigation of the template aging effects, we captured eye movement data in three sessions.

Each session consisted of multiple stimuli. Among recorded stimuli we selected two, described bellow, as the most suitable as a starting point for the participants of the competition. The 'same-stimulus' recordings for the first and the second session were separated by a short-time interval ranging from 13 to 42 minutes (M = 19.5, STD = 4.2). The 'same-stimulus' recordings for the third session were separated from the other two sessions by a long-time ranging from 8 to 13 months (M = 10.5, STD = 1.5).

4.1.2 Participants

The data for the first two recording sessions were captured from 306 subjects (165 male/141 female), ages 18-46 (M = 22, STD = 4.3). The third session was performed for a subset of 74 participants from the original subject pool of 306 individuals. During each session, every participant performed one recording for each of the two types of the used visual stimuli. As a result, the total number of the unique recordings of the BioEye 2015 datasets was 1372. All recordings were performed in the Human Computer Interaction Laboratory of Texas State University. Texas State University's institutional review board approved the study, and the participants provided informed consent.

4.1.3 Apparatus

The recording device used for capturing the eye movements was an EyeLink 1000 eye tracker [27], operating with a sampling frequency of 1000 Hz. During the experiments, the device operated in a monocular mode and captured the movements from the left eye only. The recording setup included a head stabilizer with a chinrest for limiting head movements. The visual stimuli were presented on a flat screen monitor with dimensions 474 x 297 mm, and resolution of 1680 x 1050 pixels. The distance between the stimulus screen and the subject's head was 550 mm. The primary eye position corresponded to the center of the screen with a vertical offset of 35 mm above the eye level.

4.1.4 Visual Stimuli

Random Dot Stimulus (RAN): the RAN stimulus consisted of a white circular point with its center marked with a black dot, presented in a black background on a computer screen. The point made 100 'jumps' from one position to another in a uniformly distributed random pattern. Subjects were instructed to follow the white dot with their eyes. The time interval between two consecutive jumps was 1 second, and the total duration of this stimulus was 1 minute and 40 seconds. Figure 1 shows an example of the RAN stimulus and the respective captured eye movement signal.

Text Reading Stimulus (TEX): the TEX stimulus consisted of text excerpts from the poem of Lewis Carol 'The Hunt for the Snark'. The particular content of the used excerpts provokes an active processing of the text, thus allowing for an inspection of the cognitive aspects of the eye movements. The total time given to each participant to freely read the text excerpts was 1 minute. Figure 2 shows an example of the TEX stimulus and the respective captured eye movement signal.
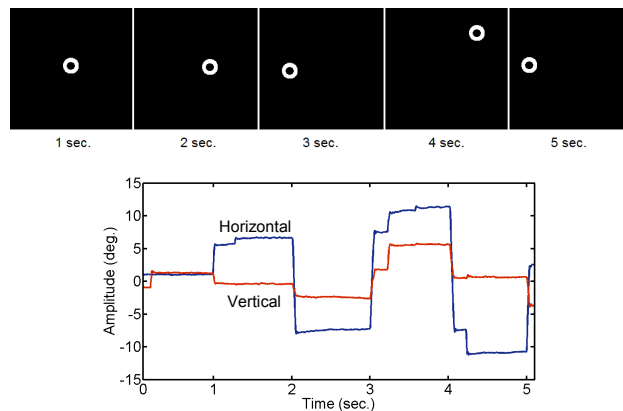

Figure 1. Example of the used RAN stimulus (five positions are shown) and the respective eye movement signal.
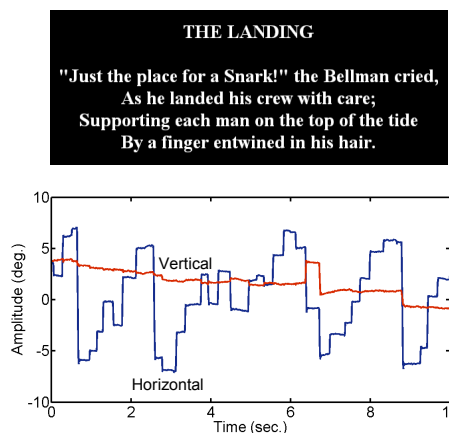

Figure 2. Example of the used TEX stimulus and the respective eye movement signal.

4.1.5 Recordings Quality

During the experiments we followed a strict recording protocol in order to ensure the creation of a very high quality database of eye movement recordings. A calibration process was performed before every new recording, and the quality of each calibration was quantified using accuracy measures. The measured average calibration accuracy across the three sessions for the RAN stimulus was 0.49° (STD = 0.16°), and for the TEX stimulus was 0.48° (0.16°). After the completion of the recordings we calculated the average recording validity across the three sessions. Recording validity is defined as the percent of samples that were successfully captured by the eye tracker during the whole duration of each recording. Usually, the sources of invalidity are related to blinking (most common), out of range movements, squinting, eye moisture, excessive mascara, etc. The average recording validity across the three sessions for the RAN stimulus was measured to 96.7% (STD = 4.7%), and for the TEX stimulus the validity was 94.3% (STD = 5.7%).

4.2. Datasets Formation

4.2.1   Original Datasets

The eye movement recordings were organized in four datasets to allow the participants to test their algorithms for different parameters, i.e. different visual stimuli, and different time intervals between the recordings. The datasets 'RAN_Short' and 'TEX_Short' were assembled by the recordings captured within the short-time interval described in the Section 4.1.1 using 306 subjects. The datasets 'RAN_Long' and 'TEX_Long' were assembled by the recordings captured within the long-time interval from the subset of 74 subjects.

The datasets were split in two halves. Recordings from one half of the subjects created the development datasets, and the recordings from the other half created the evaluation datasets. This split of the datasets was performed to ensure that recordings from the same subject would not appear in both datasets, thus reducing any possible overfitting effects while training the algorithms.

Although the raw recordings were recorded with 1000 Hz sampling rate, the signals were downsampled to 250 Hz. This pre-processing was done in order to: 1) mitigate high frequency noise, 2) reduce the number of samples for efficient processing, and 3) allow for a reasonable storage and transmission size for the datasets. Previous research [18] indicates that the reduction of sampling rate to 250 Hz does not substantially affect recognition accuracy.

4.2.2   Additional Datasets

Two additional datasets were given to the participants after the end of the original submission period. These datasets were employed to provide further insights into the robustness of the algorithms. The datasets were assembled by using multiple recordings from a subset of 100 subjects for the RAN stimulus, selected as the most challenging stimulus. The recordings were conducted in time intervals from 2 weeks to 10 months. The number of unlabeled recordings for each subject randomly varied between 1 and 6, in contrast to the original datasets where each subject had exactly one labeled and one unlabeled sample. For the dataset 'RAN_MultiRec_250Hz' we employed the same downsampling process used for the original datasets. For the dataset 'RAN_MultiRec_1000Hz' we used the raw data recorded at 1000 Hz. This design allowed to infer: 1) stability of an algorithm for an unknown amount of unlabeled samples recorded with various time intervals, 2) stability of an algorithm to a change in the sampling rate.

5. Competition Procedure

5.1. Acquisition of the Results

Participants had a month to work with the development datasets before evaluation datasets became available. Participants had 26 days to submit the results for the evaluation datasets and were allowed one submission per day. After each submission, participants were able to view their own results and the best results from other competitors in an anonymized format. After the end of the original submission period, participants were requested to also label the additional datasets described above. The participants where asked to use their best algorithm without modifying the parameters/thresholds of the algorithm. Only one submission for the additional datasets was allowed.

5.2. Evaluation Metrics

The evaluation metric for the competition was the Rank-1 Identification Rate (IR), defined as the ratio of the total number of correctly identified unlabeled recordings to the total number of the unlabeled recordings. Since the competition database consisted of four separate datasets, the Rank-1 IR performances were evaluated for each dataset separately. Final result for the determination of the winner was calculated by the following formula:

$$IR_f = w_{D1} \cdot IR_{D1} + w_{D2} \cdot IR_{D2} + w_{D3} \cdot IR_{D3} + w_{D4} \cdot IR_{D4}$$

where $D1$ = 'RAN_Short', $D2$ = 'RAN_Long', $D3$ = 'TEX_Short', $D4$ = 'TEX_Long', and $w_{D1} = 0.3$, $w_{D2} = 0.2$, $w_{D3} = 0.3$, $w_{D4} = 0.2$. These weights were selected to accommodate for the smaller number of subjects in the long-time interval datasets.

6. Description of Participating Methods

This section presents the descriptions of the developed methods, as provided by the participants of the competition.

Method 1(ranked first)
Anjith George and Prof. Aurobinda Routray, Indian Inst. of Tech. Kharagpur
In this method, a large number of statistical features were extracted from position, velocity, and acceleration profiles of fixations and saccades. The redundant and correlated features extracted in this stage were removed via a feature selection process. Several algorithms were tested, and the best IR was achieved by a Neural Network based framework. The developed approach did not use the provided validity and the stimulus position information. However, it used the one-to-one correspondence between the unlabeled and the labeled recordings during the classification.

Method 2 (ranked second)
Narishige Abe, Fujitsu Laboratories Ltd.
In this method, the used features included fixation/saccade duration, acceleration, and frequency. Recordings were divided into local blocks and the features related to the fixation/saccade information or frequency (Mel-Frequency Cepstral Coefficients) were extracted from the blocks, and converted to vectorized data. The extracted features were compared using a simple vector comparison method. The

developed approach did not use the stimulus position information, but it used the validity information and the one-to-one correspondence between recordings during the classification.

Method 3 (ranked third)
Pawel Kasprowski, Silesian University of Technology
In this method, the used features were position, velocity, acceleration, and jerk. For the RAN datasets, the data were divided to chunks according to the stimulus position. The chunks were organized into four subsets (NE, NW, SE, SW) and Dynamic Time Warping distances were calculated between all chunks. The distances were used to build a classifier based on either the support vector machines (SVM) or the random forests (RF). For the TEX datasets, histograms with different bins were used instead of the distances. Combination of the results from different classifiers was done using a simple voting scheme. The developed approach used the validity information to remove invalid data, and the one-to-one correspondence between recordings during the classification.

Method 4 (ranked forth)
Thomas Kübler, Un. Tuebingen
In this method, the used features were saccade direction (binned degrees as well as horizontal/vertical ratio), amplitude, velocity and acceleration profiles, fixation duration profiles, Gaussian Mixture model parameters of fixation and saccade mean velocity and velocity spread, blink rate and duration, and stimulus hit accuracy. Based on the features a pairwise similarity score was calculated (using chi-square distance between histograms for the velocity, acceleration and fixation duration profiles). Linear weights for the individual scores were calculated based on the training dataset. The developed approach used the validity and the stimulus information, but did not use the one-to-one correspondence between recordings during the classification.

Method 5 (ranked fifth)
Bruno Galmar, Independent Researcher, Taiwan
In this method, the first stage involved the classification of the fixations versus saccades and the measurement of some characteristics/features of both fixations and saccades (position, velocity, and acceleration). In the next stage feature selection was performed via an empirical approach based on the maximization of a specific criterion, i.e. the number of unique assigned labels. The selected features were used for the KNN classification with the Mahalanobis distance. The developed approach used the stimulus position and validity information, as well as the one-to-one correspondence between recordings during classification.

Method 6 (ranked sixth)
Robert Bixler, University of Notre Dame
In this method, different eye movement classification techniques were explored and the dispersion-based I-DT

method was selected as the optimum, based on the assessment using qualitative and quantitative eye movement classification scores. Then, different eye movement features were calculated from the fixation and saccadic characteristics (duration, amplitude, velocity, scanpath). Different classifiers were trained with the labeled data, and during the evaluation with the testing set the best performing classifier was found to be the support vector machines (SVM).

Method 7 (ranked seventh)
Andrey Kuehlkamp, University of Notre Dame
In this method, the signal was processed using an implementation of the I-DT (dispersion threshold) method to identify fixations and saccades. Then, several features were extracted related to the fixation start time, duration, position, and saccade start time, duration, amplitude, and mean velocity. Each feature was treated as a histogram, and the similarity between histograms of different subjects was compared using Dynamic Time Warping. The developed approach did not use any of the provided information regarding the stimulus position, the validity, and the one-to-one correspondence between recordings.

7. Competition Results

7.1. Example Baselines

In order to provide the participants with some baseline results, we calculated results on the competition datasets by using a method already published in bibliography. This method was the Complex Eye Movement Behavior (CEM-B) biometrics [19]. Since the algorithm was used to provide baseline results, its most simple instantiation was employed, i.e. feature extraction, comparison, and combination using simple sum. Direct classification was used without employing the one-to-one correspondence of unlabeled and labeled recordings. In Table 1, we show the baseline results achieved by the CEM-B algorithm for the development and the evaluation datasets accordingly. The value of the field 'FINAL' is calculated using the combination formula described in Section 5.2.

Table 1. Example baseline results by using the CEM-B algorithm for the development and evaluation datasets.

Development Datasets (half gallery - half probe)				
Rank-1 Identification Rate %				
RAN_Short	**RAN_Long**	**TEX_Short**	**TEX_Long**	**FINAL**
40.5	16.2	52.9	40.5	39.4
Evaluation Datasets (labeled gallery - unlabeled probe)				
Rank-1 Identification Rate %				
RAN_Short	**RAN_Long**	**TEX_Short**	**TEX_Long**	**FINAL**
34.0	40.5	58.2	48.6	45.5

Table 2. Final results for the original evaluation datasets of BioEye 2015 competition.

Participant	Rank-1 Identification Rate %				
	RAN_Short	RAN_Long	TEX_Short	TEX_Long	FINAL
A. George and Prof. A. Routray *Indian Inst. of Tech. Kharagpur*	98.7	89.2	98.0	94.6	**95.8**
N. Abe *Fujitsu Laboratories Ltd.*	88.9	59.5	91.5	67.6	**79.5**
P. Kasprowski *Silesian Un. of Technology*	75.8	54.1	73.2	45.9	**64.7**
Thomas Kübler *Un. Tuebingen*	51.6	40.5	57.5	40.5	**49.0**
B. Galmar *Independ. Research., Taiwan*	30.1	24.3	49.0	27.0	**34.0**
R. Bixler *Un. of Notre Dame*	20.3	5.4	13.7	10.8	**13.4**
A. Kuehlkamp *Un. of Notre Dame*	10.5	10.8	11.8	10.8	**11.0**

7.2. Results on the Original Datasets

This section presents the final results of the participants for the original datasets of the BioEye 2015 competition. As already explained, the participants were allowed to make more than one submission. Table 2 presents the top achieved results from all their submissions, by taking into consideration the submissions for every dataset separately.

The top achieved overall Rank-1 IR was 95.8%. However, in some specific datasets the performance reached even higher levels. It is worth noting that even the participants with the lowest accuracy were capable of achieving rates that are considerably higher than the level of chance, which is 0.7% for the 'Short Term' datasets, and 2.8% for the 'Long Term' datasets.

With a close inspection of the table we can observe a clear trend: irrespectively of the stimulus, the rates for the 'Short Term' datasets are relatively higher than the rates for the 'Long Term' datasets. Overall, the performance loss ranges from 0.4% to 29.4% (M = 13.6%, STD = 10.1%). For the top-3 methods this loss ranges from 9.5% to 29.4% (M = 20.2%, STD = 10.0%). These results seem to confirm the observations from the previous studies [25], [28], investigating the impact of the temporal proximity of recordings on the eye movement biometric performance.

A quick inspection of the performances for different types of visual stimulus does not reveal an analogous strong trend. However, by averaging for each participant the performances for different stimuli—irrespective of the time interval—the impact of the stimulus on the performance of each algorithm is noticeable. In the vast majority of algorithms (5 out of 7) the preferred stimulus was the TEX. Also, the calculated absolute difference in the average performances of algorithms between the two types of stimulus was found to range from 0% to 18.9% (M = 4.9%, STD = 4.9%).

7.3. Results on the Additional Datasets

In this section, we present the results for the additional datasets used to provide further insight into the robustness of the developed algorithms. As we can see in Table 3, the employment of more challenging datasets leads to the degradation in performance. For the case of the 250 Hz dataset, the absolute loss ranges from 6.4% to 38.6% (M = 23.8%, STD = 11.8%). For the case of the 1000 Hz dataset, containing the raw recordings, there is a further drop in performance with respect to the 250 Hz dataset, ranging from 0.28% to 14.1% (M = 5.4%, STD = 6.0%). In the discussion part, we provide further analysis regarding the possible reasons for the observed deterioration in performance.

Table 3. Results for the additional datasets.

Participant	Rank-1 Identification Rate %		
	FINAL	MultiRec (250 Hz)	MultiRec (1000Hz)
CEM-B (baseline)	45.5	23.4	27.4
A. George/Prof. A. Routray *IITech. Kharagpur*	95.8	70.1	56.0
N. Abe *Fujitsu Laboratories Ltd.*	79.5	44.3	30.4
P. Kasprowski *Silesian Un. of Technology*	64.7	31.5	N/A
T. Kübler *Un. Tuebingen*	49.0	10.3	14.4
B. Galmar *Independ. Research., Taiwan*	34.0	13.3	12.2
R. Bixler *Un. of Notre Dame*	13.4	7.1	6.8
A. Kuehlkamp *Un. of Notre Dame*	11.0	2.2	2.7

8. Discussion

The competition procedure and the obtained results allow reaching useful conclusions. The results obtained for the original datasets by the best performing algorithms reach relatively high levels. This fact is particularly encouraging for the future of the eye movement-driven biometrics. Also, it highlights the importance of the adopted procedures for ensuring recordings of a high quality, which in turn can be beneficial during the feature extraction process. However, it should be taken into consideration that the specific results should be interpreted taking into account the structure and the limitations of the employed datasets. First, the original competition datasets contained exactly one unlabeled recording for every labeled recording. This information can be utilized during the classification process and can lead to more optimistic identification rates as indeed it occurred per feedback from the participants. However, in real world identification the conditions are substantially more challenging than the environment employed for the recordings of the original datasets. Second, a unique performance metric (Rank-1 IR) was utilized in order to facilitate comparison of the methods. This can lead to approaches that are optimized to tackle this specific metric. Finally, the participants were allowed to make more than one submission for the competition (limited, though, to one per day).

The descriptions of the methods, provided by the participants, allow for an informative overview of the possible eye movement features that can be employed for recognition purposes. Among them, the most informative seem to be the features related to the duration, position, amplitude, velocity, and acceleration of the eye movements. Also, some features extracted from the spectral analysis of the signal can provide some valuable information. It is noteworthy that the best performing method of the competition was based on a pattern recognition/machine learning framework, which argues for further investigation into such techniques to improve overall accuracy of the eye movement biometrics methods.

One of the goals of this competition was to evaluate the impact of the template aging on different algorithms. The observed trend was a decrease in performance for the recordings captured with the longer time intervals. This can be expected for the eye movement biometrics where the features can be frequently modulated by behavioral factors, which are more susceptible to the time-related alterations than the static physical cues. Also, it should be noted that the number of the available recordings for the 'Long' period was smaller than for the 'Short' period recordings. This emphasizes the performance difference even further, and should be taken into consideration during the relative comparison of the presented results.

Regarding the different types of used visual stimuli, the observed differences were less pronounced. Of course, the signals generated for different stimuli can exhibit different overall shapes and specialized characteristics. However, the participants were in general capable of modifying their algorithms so that they can extract the corresponding features for both cases of visual stimuli. Most of the algorithms worked better for the case of the TEX stimulus than the RAN stimulus. This phenomenon can be possibly attributed to a more complex response of the human visual system to a task of reading and the lack of the randomization that was present in the RAN stimulus.

The use of the additional datasets facilitated further analysis of the developed algorithms. The inclusion of multiple unlabeled recordings per subject leads to a reduction in accuracy for all algorithms. One of the reasons for that is the added complexity of this specific identification scenario. The one-to-one correspondence assumption becomes invalid, and also, the inclusion of more unlabeled recordings per subject can deteriorate the discriminability of the features. Also, the recordings were captured in time intervals that were different from the original datasets. In our opinion the specific rates achieved for this more complex scenario are expected to be more representative of the biometric potential of the extracted eye movement features, and the resulting performance is, possibly, closer to what can be achieved in a more realistic use of an eye movement-driven biometrics system.

In the case of the 1000 Hz raw recordings dataset, the participants were asked to apply the exact same code version that they used for the original datasets, so that the developed code robustness can be evaluated for the signals of different characteristics. In general, there was an observed further drop in performance. This can be attributed to the differentiating characteristics of the signal (noise, invalidity) for which the already developed codes could not accommodate automatically. Given the requirement of using the exact same code, which limited the possibility of making any improvements, it is possible to hypothesize that underlying algorithms can be fine-tuned to provide higher performance than what is reported here.

9. Conclusion

The goal of the BioEye 2015 competition was to provide the opportunity to the researchers to analyze a large high-quality database of eye movement recordings and develop their methods in order to tackle the task of biometric identification. The results showed that there is a wide gamut of features and signal processing approaches that can be employed for the analysis of the eye movements and the extraction of the distinctive features that allow establishing the identity of a person. The organizing committee hopes that the ideas generated during the competition and from the presentation of the results will contribute to the understanding of the challenges of the eye movement-driven biometrics, and further advance the research in this field.

10. Acknowledgements

References

[1] P.-J. Coulier, "Les vapeurs d'iode employees comme moyen de reconnaitre l'alteration des ecritures.," *In L'Annee scientifique et industrielle, Figuier, L. Ed., Hachette,* vol. 8, pp. 157-160, 1863.

[2] A. K. Jain, P. J. Flynn, and A. Ross, *Handbook of Biometrics*: Springer-Verlag New York, Inc., 2007.

[3] P. Kasprowski and J. Ober, "Eye Movements in Biometrics," in *Biometric Authentication*. vol. 3087, D. Maltoni and A. K. Jain, Eds., ed: Springer Berlin Heidelberg, 2004, pp. 248-258.

[4] E. Javal, "Essai sur la physiologie de la lecture.," *Annales d'Oculistique,* vol. (79), 97-117, (80), 135-147, 240-274, 1878.

[5] A. L. Yarbus, *Eye Movements and Vision*: New York: Plenum Press, 1967.

[6] M. A. Just and P. A. Carpenter, "A theory of reading: from eye fixations to comprehension," *Psychol Rev.,* vol. 87, pp. 329-354, 1980.

[7] D. Noton and L. Stark, "Scanpaths in eye movements during pattern perception," *Science (New York, N.Y.),* vol. 171, pp. 308-311, 1971.

[8] D. Noton and L. Stark, "Scanpaths in saccadic eye movements while viewing and recognizing patterns," *Vision Research,* vol. 11, pp. 929-932, 1971.

[9] O. V. Komogortsev, A. Karpov, and C. D. Holland, "Attack of Mechanical Replicas: Liveness Detection With Eye Movements," *IEEE Transactions on Information Forensics and Security,* vol. 10, pp. 716-725, 2015.

[10] O. V. Komogortsev and C. D. Holland, "The application of eye movement biometrics in the automated detection of mild traumatic brain injury," presented at the Ext. Abstr. of 32nd ACM Conference on Human Factors in Computing Systems (CHI), 2014.

[11] R. Bednarik, T. Kinnunen, A. Mihaila, and P. Fränti, "Eye-Movements as a Biometric," in *Image Analysis*. vol. 3540, H. Kalviainen, J. Parkkinen, and A. Kaarna, Eds., ed: Springer Berlin Heidelberg, 2005, pp. 780-789.

[12] T. Kinnunen, F. Sedlak, and R. Bednarik, "Towards task-independent person authentication using eye movement signals," presented at the 2010 Symposium on Eye-Tracking Research & Applications (ETRA), 2010.

[13] O. V. Komogortsev, S. Jayarathna, C. R. Aragon, and M. Mahmoud, "Biometric identification via an oculomotor plant mathematical model," presented at the 2010 Symposium on Eye-Tracking Research & Applications (ETRA), 2010.

[14] O. V. Komogortsev, A. Karpov, L. R. Price, and C. Aragon, "Biometric authentication via oculomotor plant characteristics," in *2012 IAPR International Conference on Biometrics (ICB)*, 2012, pp. 413-420.

[15] Z. Liang, F. Tan, and Z. Chi, "Video-based biometric identification using eye tracking technique," in *2012 IEEE International Conference on Signal Processing, Communication and Computing (ICSPCC)*, 2012, pp. 728-733.

[16] I. Rigas, G. Economou, and S. Fotopoulos, "Biometric identification based on the eye movements and graph matching techniques," *Pattern Recognition Letters,* vol. 33, pp. 786-792, 2012.

[17] V. Cantoni, C. Galdi, M. Nappi, M. Porta, and D. Riccio, "GANT: Gaze analysis technique for human identification," *Pattern Recognition,* vol. 48, pp. 1023-1034, 2015.

[18] C. D. Holland and O. V. Komogortsev, "Complex Eye Movement Pattern Biometrics: The Effects of Environment and Stimulus," *IEEE Transactions on Information Forensics and Security,* vol. 8, pp. 2115-2126, 2013.

[19] C. D. Holland and O. V. Komogortsev, "Complex eye movement pattern biometrics: Analyzing fixations and saccades," in *2013 International Conference on Biometrics (ICB)*, 2013, pp. 1-8.

[20] H.-J. Yoon, T. R. Carmichael, and G. Tourassi, "Gaze as a biometric," in *SPIE*, 2014, pp. 903707-903707-7.

[21] I. Rigas and O. V. Komogortsev, "Biometric Recognition via Probabilistic Spatial Projection of Eye Movement Trajectories in Dynamic Visual Environments," *IEEE Transactions on Information Forensics and Security,* vol. 9, pp. 1743-1754, 2014.

[22] P. Kasprowski, O. V. Komogortsev, and A. Karpov, "First eye movement verification and identification competition at BTAS 2012," in *2012 IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2012, pp. 195-202.

[23] N. V. Cuong, V. Dinh, and L. S. T. Ho, "Mel-frequency Cepstral Coefficients for Eye Movement Identification," in *IEEE 24th International Conference on Tools with Artificial Intelligence (ICTAI)*, 2012, pp. 253-260.

[24] I. Rigas, G. Economou, and S. Fotopoulos, "Human eye movements as a trait for biometrical identification," in *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2012, pp. 217-222.

[25] P. Kasprowski and K. Harezlak, "The Second Eye Movements Verification and Identification Competition," in *2014 IEEE International Joint Conference on Biometrics (IJCB)*, 2014, pp. 1-6.

[26] J. V. Monaco, "Classification and authentication of one-dimensional behavioral biometrics," in *2014 IEEE International Joint Conference on Biometrics (IJCB)*, 2014, pp. 1-8.

[27] *EyeLink. EyeLink 1000 Eye Tracker*. Available: http://www.sr-research.com/

[28] O. V. Komogortsev, C. D. Holland, and A. Karpov, "Template aging in eye movement-driven biometrics," in *SPIE*, 2014, pp. 90750A-90750A-9.