

## The Second Eye Movements Verification and Identification Competition

Pawel Kasprowski  
Silesian University of Technology  
Akademicka 16, Gliwice, Poland  
pawel.kasprowski@polsl.pl

Katarzyna Hareźlak  
Silesian University of Technology  
Akademicka 16, Gliwice, Poland  
katarzyna.harezlak@polsl.pl

### Abstract

*The idea concerning usage of the eye movement for human identification has been known for 10 years. However, there is still lack of commonly accepted methods how to perform such identification. This paper describes the second edition of Eye Movement Verification and Identification Competition (EMVIC), which may be regarded as an attempt to provide some common basis for eye movement biometrics (EMB). The paper presents some details describing the organization of the competition, its results and formulates some conclusions for further development of EMB.*

### 1. Introduction

Eyes are one of the most complicated human organs and the analyses of eye movements may reveal a lot of information about a human being. There are a lot of studies that analyze eye movements in order to diagnose specific diseases or to recognize the state of mind [1]. However, surprisingly, there is only little research trying to differentiate people on the basis of their eye movements characteristics.

The paper presents results of The Second Eye Movement Verification and Identification Competition (EMVIC2014), which was organized in conjunction with IJCB 2014 and was one of the official conference's competitions.

The paper is organized as follows: Section 2 briefly describes the state of the art in the eye movement biometrics (EMB) including information about the first EMVIC challenge. Section 3 presents the competition, focusing on description of the dataset made accessible to the participants and the competition rules. Section 4 presents the results of the competition. There is a detailed analysis of competition's results presented in section 5 and, finally, section 6 contains a summary and further plans.

### 2. The state of the art

The idea concerning usage of the eye movements for biometric authentication was presented for the first time in [2]. First experiments, aiming at developing this idea, used

a jumping point stimulus, during which the user was instructed to follow with eyes the point appearing on the screen [3][4]. The subsequent research involved other stimuli in form of static point [5] and free image observations [6]. The first results of such authentication were quite promising, although error rates were too high to reliably use an eye movements signal in practical applications.

In 2006, Silver et al [7] proposed the first known combination of eye movements biometrics and keystroke dynamics. The data from both modalities was recorded during one experiment. However, the results from keyboard dynamics were reported to give much lower error rates and only some limited properties of eye movements (like number of fixations and average fixation length) were used for the purpose of the authentication.

In [8] a 'task-independent' authentication was proposed for the first time. Such authentication didn't depend on stimulus presented, yet used an eye movement signal recorded while a subject was watching a movie.

The first attempt to model eye movements for authentication was done in [9] where so-called Oculomotor Plant Mathematical Model (OPMM), developed by the authors, was used during jumping point based authentication.

In 2011 Deravi et al [10] published the paper, in which they checked possibility of identifying people based on the way they look at static images. However, there were only three subjects in their experiment so the results cannot be considered as reliable.

As eye movement during reading is one of the most investigated subjects in cognitive sciences, there was also an attempt to authenticate people based on their reading patterns [11]. However, the setup of such experiment is difficult due to so called memory effect – when people already know the text they only skim it instead of reading.

In [12] Biedert et al analyzed eye movements of subjects during their normal activity (opening mails, reading documents). They tried to prove that it is possible to estimate whether a subject is familiar with a computer desktop and in that way identify an intruder (who is supposedly not familiar with it).

## 2.1. EMVIC 2012

EMVIC 2012 was the First Eye Movement Verification and Identification Competition, organized by Kasprowski and Komogortsev in 2012 [13]. The main purpose of the competition was to popularize eye movement based biometrics and to provide the single reference point for further research. There were four different datasets published and competition was opened in two different places – well known Kaggle page ([www.kaggle.com](http://www.kaggle.com)) and native competition page ([www.emvic.org](http://www.emvic.org)).

The participant's task for each dataset was to find proper identification of subjects for unlabeled samples based on some available labeled (training) samples, recorded for the same subjects. The results for all participants were compared according to their accuracy (ratio of correctly classified samples). The competition description and the obtained results are provided in several publications [13][14][15].

The most interesting finding of the competition regarded differences in the accuracy of the results obtained for various datasets. Although all four datasets contained eye movement registered for the same type of stimulus (jumping point), the results differed significantly among datasets. It showed that the quality of data, data collecting scenarios and device used may significantly influence results, which was analyzed in subsequent publications [16][17].

## 2.2. Summary

Possibly the contest announced as part of one of the most important IEEE biometric conferences influenced the popularity of eye movements biometric, because there have been several new papers published since 2012 [18][19][20][21].

In all the aforementioned publications regarding eye movements the authors retrieved different features of an eye movement signal. In most cases the signal was preprocessed to divide it into fixations (moments when the eye is looking at one place) and saccades (rapid movements from one fixation point to another). Some authors focused on fixation information, retrieving their sequences [22] or on identifying eye micro-movements during fixations [5]. Other authors focused on saccades calculating their velocities and accelerations [18][19]. There were also approaches using raw signals and their different transformations [15][4].

## 3. Competition

This section presents the competition setup – dataset used, competition procedure and submission opportunities.

### 3.1. Dataset

A head mounted Jazz-Novo eye tracker that records eye

positions with 1000Hz frequency was used to collect eye movement data. 34 subjects took part in the experiment for whom overall 56 sessions were recorded. Every session consisted of an initial calibration and subsequent presentations of images. To obtain comparable results there were face images, photographs of different people, used as stimuli. Every face appearing on the screen was cropped in the way ensuring the same location of eyes for every picture. No further processing was applied.

Similar stimuli have already been used in the eye movement based biometric identification [22][20]. In [22] there were 10 faces showed to participants during eight sessions. Every participant taking part in each session looked at the same 10 faces but presented in a different order. The subject's task was to look freely at each face for 4 seconds. The participants were identified using the information about all ten faces observations (the whole session). In [20] there were 16 face images used in each session and presented to participants for 10 seconds. Similarly to [22] so called fixation models were built separately for every session (16 observations) and used for identification of other sessions. In both cases the same set of face images was used for both model creation (learning) and evaluation (testing) stages.

Contrary to free observations employed in [22] and [20] in the experiment described here the participants' task was to look at a face on the screen and assess, by pressing one of the two possible buttons, if they recognized the face or not. When the participant pressed the button, the face disappeared. Every such task was recorded as a separate sample. Similarly, contrary to experiments conducted in [22] and [20], in the presented approach a subject identification was based on one sample (one photo and related button click), not on all images presented during the session.

The length of an observation of each face differed because participants could finish the observation freely when they made a decision regarding an observed face familiarity. The average sample length was 2429 msec. but lengths ranged from 891 msec. to 22012 msec.

Every person took part in at least one 'session' – a sequence of face observations. As it was mentioned earlier there were overall 56 sessions provided for 34 participants of whom 22 took part in two sessions with at least one week interval between them. The presented experiment consisted of 24 observations (24 different faces) in the first its session and of 27 observations (27 faces different from those used in the first session) in the second one. The total number of separate observations was 1430.

The samples recorded for each user during the first session were used as the training set (most of 24 observations). Data recorded during the second session was used as testing samples (27 samples for 21 users and 26 samples for one user). Hence, the training set consisted of 837 samples of 34 subjects and the testing set included 593

samples of 22 subjects.

The properties of the dataset described above (especially various size of samples and different stimuli for every sample recorded) made it, in the authors opinion, one of the most challenging datasets used for eye movement biometrics so far.

### 3.2. The competition procedure

Prospective competitors could register on the web page ([www.emvic.org](http://www.emvic.org)) and download the dataset. The dataset consisted of two parts: training data - set of known (labeled) samples, and testing data - set of samples with unknown classification (unlabeled). The competitors task was to analyze the labeled samples, develop classifiers and use them to classify the unlabeled samples from the test data file. The results of classification were expected to be delivered in a text file in a format specified on the web page.

The main metric used for the results evaluation was accuracy, defined as the number of test samples classified correctly to the number of all test samples. It was possible to send more than one submission but the number of submissions was limited to one per day.

The gathered results were systematically published on the competition web site. To avoid over-fitting resulting from intensive usage of the feedback to improve classification – only the results calculated for 80% of samples were made accessible. It changed during the last three days before deadline.

## 4. The results

To download the dataset it was necessary to create an account on [www.emvic.org](http://www.emvic.org) web site. There were 82 participants that registered and downloaded the dataset, which shows the growing interest in eye movement biometrics. However, only 19 users uploaded their results to be evaluated. Nevertheless, because users could submit more than one result there were overall 176 submissions sent.

The task for the participants was to guess a correct identity of every sample, which was not trivial undertaking. Firstly, there were samples with duration less than one second. Secondly, every sample was a recording of an observation of a different image. Because there were 34 subjects (i.e. 34 possible results), the expected random guess result was 2.9%.

The best result achieved by the winner of the competition was about 40% (see Table 1). Several participants reported very good results while doing cross validation in the training set and then they were surprised that their results for the testing set were much worse. It was especially noticeable in the case of the participants that achieved good results in EMVIC2012. The reasons for such phenomenon were analyzed in section 5.2.

Table 1. Results of the EMVIC challenge.

Rank	Participant	Result
1	Vinnie Monaco, Pace University	39.63%
2	Narishige Abe, Stanford University	35.24%
3	Subhadeep Mukhopadhyay, Temple University	26.48%
4	Dragan Gamberger, Rudjer Boskovic Institute, Zagreb Croatia	25.97%
5	Vitor Yano, University of Campinas	21.08%

All results are available on [www.emvic.org](http://www.emvic.org)

Contrary to the previous EMVIC [13] the three best submissions used methods that may be generally described as time series analyses. Submissions that used eye movements related features such as fixations, velocities or eye spatial positions, have got, in this year competition, lower scores – possibly because there was less data available for both training and testing sets.

## 5. Analysis of the participants' results

There were 176 submissions, so every sample could have been classified correctly from 0 to 176 times. Figure 1 shows the distribution of number of classifications done correctly. It shows that most samples had 10-20 correct classifications, which corresponds to the range of 5 to 11% of possible assignments to obtain.

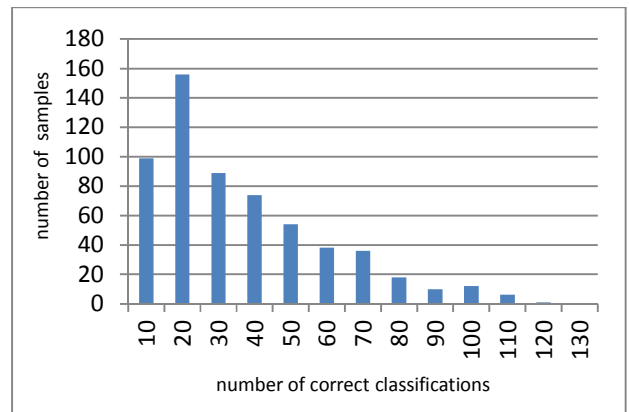


Figure 1: Histogram of number of samples to number of correct classifications.

Having such results the obvious question to ask was what properties of a sample influence the recognition rate, i.e. whether it is possible to predict a recognition rate of the sample.

At first it was analyzed if there is any correlation between a recognition rate and sample's length. It could be

supposed that results should be better if more data is available (i.e. a sample is longer). However, it occurred that there is no correlation between these two values (with Pearson correlation coefficient equal to  $-0.00015$ ). It means that shorter samples were comparatively difficult to classify with longer samples (see Figure 2).

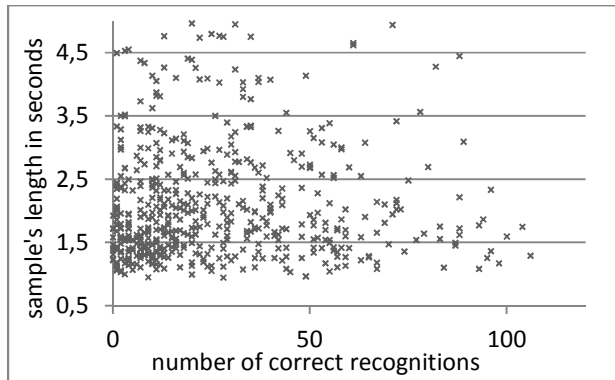


Figure 2: Pictorial representation of the correlation between number of correct recognitions (horizontal) and a sample's length.

Additionally, the ANOVA test with the number of correct results as an independent value was performed. There were several different sample's properties used as dependent values, including: subject id, an image related to recorded sample, and familiarity of an image.

As the values of the results were skewed with long right tail, the numbers were square root transformed to get the normal distribution. Table 2 presents the significance of each dependent property calculated with ANOVA test. There was also information about the number of classes (distinct values) for every property included in the second column.

Table 2. Dependency of recognition rate on some nominal samples' properties.

Property	Distinct values	Significance (p-value)
Subject id	22	$<0.0001$
Image observed	24	0.981
Familiarity	2	0.917

The analyses of the outcome showed that recognition rate did not depend on the image being observed. Hence, it was impossible to provide guidelines on what kind of images (faces in this example) were better for the use in identification tasks.

Likewise, the familiarity of the face did not influence recognition rates. It means that it was comparatively difficult to recognize a person observing a familiar face and an unfamiliar one.

The only significant dependency found was relation of the rate and the subject id. There were subjects that were much easier to identify than others. The detailed description of this finding is presented in the next section.

### 5.1. Differences in recognition rates among subjects

To obtain reliable statistics only the best submissions of five best participants were taken into account in this section. Because the previous section finished with the conclusion that the recognition results were highly correlated with the subject id, the difference in accuracy among subjects had to be analyzed.

Figure 3 presents the identification accuracy aggregated for every subject. As it can be observed there were indeed considerable differences in the results with average accuracy equal to 0.31 and deviation 0.23.

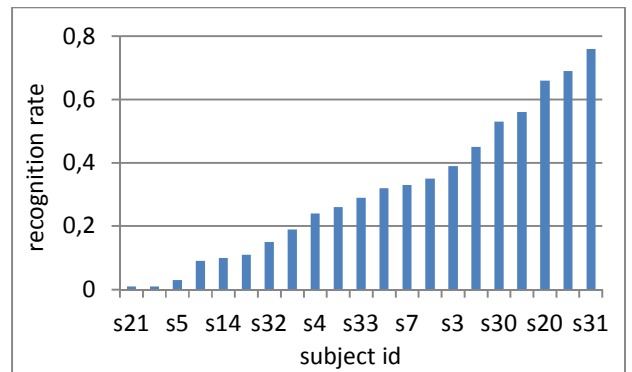


Figure 3: Recognition rates for subjects.

Moreover, it occurred that the strong correlations in recognition rates per subject among the 5 best submissions existed. It means that e.g. for all submissions subject s32 was recognized very well while subject s21 was not recognized by them almost at all.

It must be remembered that each submission taken into account here was made by a different participant and was built using a different algorithm, so such the strong correlation shows that there were subjects difficult to identify regardless of the method used.

Values of Pearson correlation coefficient are presented in table 3. Headers of rows and columns include identifiers of the 5 considered submissions.



Table 3. Correlation between recognition rates calculated for best submissions of five best subjects.

	p1	p2	p3	p4	p5
p1	1,00	0,77	0,67	0,71	0,73
p2	0,77	1,00	0,69	0,67	0,58
p3	0,67	0,69	1,00	0,67	0,76
p4	0,71	0,67	0,67	1,00	0,77
p5	0,73	0,58	0,76	0,77	1,00

Some effort has been made to find, which factors related to a subject influence recognition rate. There were different factors analyzed, like an average observation duration, gender, number of fixations, yet with no significant dependencies found. The only correlation that has been found regarded the recognition rate and the quality of the calibration, in the respect to the vertical axis (value obtained from a calibration procedure performed prior to the experiment). The correlation was 0.53, which confirmed the earlier expectation that it is easier to identify people when data with better quality is available.

## 5.2. Memory effect and data dependency

As it was mentioned in Section 4 several of the challenge participants reported very good results obtained using cross validation for the training set, while their results for the testing set were much worse. It regarded the original dataset, in which (1) data recorded for each user during the first session was used as training samples, (2) data recorded during the second session was used as testing samples, (3) the interval between two sessions was at least one week.

To evaluate to what extent such division of samples influenced the performance (the final results), the organizers decided to publish, just after reaching the challenge deadline, a new dataset, in which samples from both sessions were randomly spread into training and testing sets constituting a new dataset B. The participants were asked to test their best submissions using these newly prepared sets.

During some previous research [16][17] it was proven that the time interval between samples' recordings highly influences the results. Indeed, it occurred (Table 4) that the classification results were better in the case of each of six participants that submitted the result for the second dataset.

This outcome confirmed that there was some similarity between samples of the same person collected in short intervals that was absent in samples of that person collected after some time.

That is why the classification was easier when samples from the same session of the same subjects were included in both training and testing sets – as it was in the case of dataset B. The differences are especially visible for some participants obtaining low accuracy while using the

original dataset. It seems that their methods provided the higher ability to extract properties of the signal that were dependent on time interval and used them very efficiently. However, without this information the recognition rate was much lower.

Table 4. Comparison between results for original dataset and the dataset B.

Participant	Original result	Result for dataset B	Improvement
1	39.63%	72.38%	183%
2	35.24%	82.33%	234%
3	26.54%	70.84%	267%
4	25.97%	59.17%	228%
5	16.19%	72.55%	448%
6	4.89%	86.44%	1768%

## 6. Summary

Due to the low accessibility to eye trackers, which were very complicated and expensive devices, eye movement biometrics (EMB) was in past considered more as an academic research problem than as a solution applicable in practice. However, recently the situation has changed. It is possible to obtain an eye tracker producing data with sufficient quality for the price less than \$100 (e.g. TheEyeTribe or Tobii EyeX). This fact will probably focus more attention on EMB as it will be possible to perform even for ordinary users. This phenomenon has already been visible during EMVIC 2014 – there were more participants taking part than in the first edition and EMVIC web page was attracting from 100 to even 1000 visitors per day from all over the world.

The results of the competition show that there is still a lot of work to be done to make EMB easy, fast and reliable. The setup proposed for the competition assumed that the user should be identified based on one observation of some image with no assumption regarding this observation's length. Such setup is very convenient for users, however it occurred that it is difficult to be properly used in EMB, as the best result was only 40% of correct classifications.

An eye movement signal consists of several elements. One of them is (1) physiological – it depends on properties of an oculomotor plant (set of muscles and nerves responsible for providing eye movements). This element is obviously repeatable in multiple trials for the same subject. The other two elements are behavioral. These elements may be divided into (2) long term element (dependent on subject's experience) and (3) short term element (dependent on subject's current attitude, tiredness etc.). As it was shown during the competition it may be difficult to distinguish all these three elements from each other. Especially the third one (short term behavior), which is

repeatable within one session and is different for different sessions may influence the results. The further development of EMB field should regard extraction of the mentioned element to obtain eye movement signal that is repeatable for the same subject across multiple sessions. These issue will also be taken into account while preparing the next competition edition.

## Acknowledgements

Organizers of the competition would like to thank our official partner, SensoMotoric Instruments (SMI), who awarded an SMI RED-m eye tracker to the winner of the EMVIC and a 250 Euro travel grant for each of the two runners up.

## References

- [1] Duchowski A.: A Breadth-First Survey of Eye Tracking Applications. Behavior Research Methods, Instruments & Computers (BRMIC), 34(4). 2002
- [2] Kasprowski, P., Ober, J. Eye movement tracking for human identification, 6th World Conference BIOMETRICS'2003, London, 2003.
- [3] Kasprowski, P.: Human identification using eye movements. Doctoral thesis. Silesian University of Technology, Poland, 2004.
- [4] Kasprowski, P., Ober, J.: Enhancing eye-movement-based biometric identification method by using voting classifiers. In Defense and Security (pp. 314-323). SPIE, International Society for Optics and Photonics. 2005.
- [5] Bednarik, R.; Kinnunen, T.; Mihaila, A.; Fränti, P.: Eye-movements as a biometric, In 14 Scandinavian Conference on Image Analysis, Lecture Notes in Computer Science, Springer-Verlag, vol. 3540, pp. 780-789, 2005.
- [6] Maeder, A. J., Fookes C.: A visual attention approach to personal identification. Eighth Australian and New Zealand Intelligent Information Systems Conference, 2003.
- [7] Silver, D.L., Biggs, A: Keystroke and Eye-Tracking Biometrics for User Identification. IC-AI. 2006.
- [8] Kinnunen T., Sedlak F., Bednarik R.: Towards task-independent person authentication using eye movement signals. Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ACM New York, NY, USA 2010.
- [9] Komogortsev O., Jayarathna S., Aragon C.R., Mahmoud M.: Biometric Identification via an Oculomotor Plant Mathematical Model. Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications, ACM New York, NY, USA 2010.
- [10] Deravi, F., Shivanand P. G.: Gaze Trajectory as a Biometric Modality. BIOSIGNALS. 2011.
- [11] Holland, C., Komogortsev, O. V.: Biometric identification via eye movement scanpaths in reading. In Biometrics (IJCB), 2011 International Joint Conference on (pp. 1-8). IEEE. 2011.
- [12] Biedert, R., Frank, M., Martinovic, I., Song, D. Stimuli for Gaze Based Intrusion Detection. In Future Information Technology, Application, and Service (pp. 757-763). Springer Netherlands 2012.
- [13] Kasprowski, P.; Komogortsev, O. V.; Karpov, A.: First Eye Movement Verification and Identification Competition at BTAS 2012. IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), 2012.
- [14] Rigas, I.; Economou, G.; Fotopoulos, S.: Human eye movements as a trait for biometrical identification. IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS), pp. 217-222, 2012.
- [15] Cuong, N. V.; Dinh, V.; Ho, L. S. T.: Mel-frequency Cepstral Coefficients for Eye Movement Identification. IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2012).
- [16] Kasprowski, P., Rigas I.: The influence of dataset quality on the results of behavioral biometric experiments. IEEE BIOSIG 2013 Conference, Lecture Notes in Informatics (LNI), Vol P-212, Bonn 2013.
- [17] Kasprowski, P.: The Impact of Temporal Proximity between Samples on Eye Movement Biometric Identification. Computer Information Systems and Industrial Management. LNCS 8104, Springer-Verlag, 2013.
- [18] Zhang, Y., Juhola M.: On Biometric Verification of a User by Means of Eye Movement Data Mining. IMMM 2012, The Second International Conference on Advances in Information Mining and Management. 2012.
- [19] Zhen, L., Tan F., Chi Z.: Video-based biometric identification using eye tracking technique. Signal Processing, Communication and Computing (ICSPCC), 2012 IEEE International Conference on. IEEE, 2012.
- [20] Cantoni, V., et al.: GANT: Gaze analysis technique for human identification. Pattern Recognition, 2014.
- [21] Darwish, A., Pasquier M.: Biometric identification using the dynamic features of the eyes. Biometrics: Theory, Applications and Systems (BTAS), 2013 IEEE Sixth International Conference on. IEEE, 2013.
- [22] Rigas, I.; Economou, G.; Fotopoulos, S.: Biometric identification based on the eye movements and graph matching techniques, Pattern Recognition Letters, Volume 33, Issue 6, 15 April 2012, Pages 786-792, ISSN 0167-8655, 2012.