Towards Task-Independent Person Authentication Using Eye Movement Signals

Tomi Kinnunen and Filip Sedlak and Roman Bednarik*
School of Computing
University of Eastern Finland, Joensuu, Finland

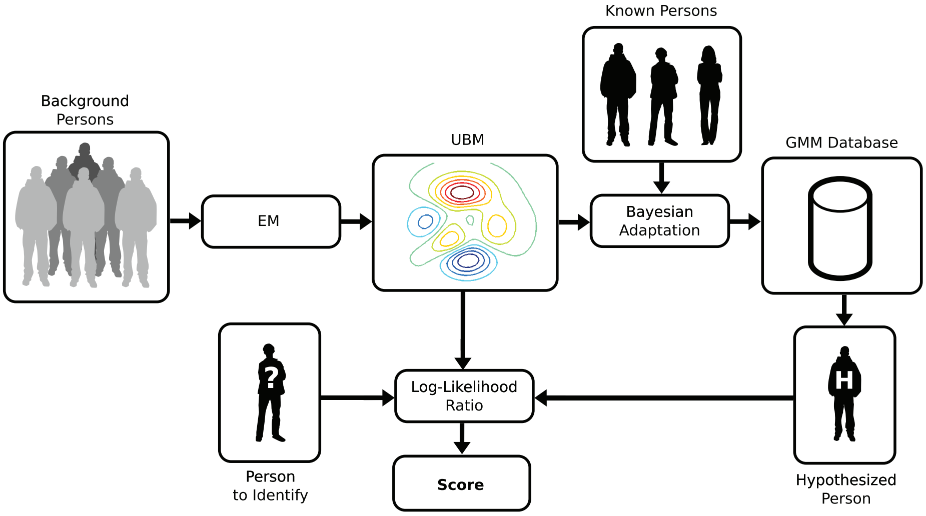


Figure 1: *A Gaussian mixture model (GMM) with universal backround model (UBM). User-dependent models are adapted from the UBM and the recognition score is normalized using the UBM score.*

Abstract

We propose a person authentication system using eye movement signals. In security scenarios, eye-tracking has earlier been used for gaze-based password entry. A few authors have also used physical features of eye movement signals for authentication in a *task-dependent scenario* with matched training and test samples. We propose and implement a *task-independent* scenario whereby the training and test samples can be arbitrary. We use short-term eye gaze direction to construct feature vectors which are modeled using Gaussian mixtures. The results suggest that there are person-specific features in the eye movements that can be modeled in a task-independent manner. The range of possible applications extends beyond the security-type of authentication to proactive and user-convenience systems.

CR Categories: K.6.5 [Management of Computing and Information Systems]: Security and Protection—Authentication; I.5.2 [Pattern Recognition]: Design Methodology—Feature evaluation and selection;

Keywords: Biometrics, eye tracking, task independence

1 Introduction

The goal of *biometric person authentication* is to recognize persons based on their physical, behavioral and/or learned traits. *Physical* traits, such as fingerprints, are directly measured from one's body with the aid of a biometric sensor. *Behavioral* traits, such as handwriting and speech, on the other hand, involve physical action performed by the user, and hence involve a time-varying component which can be controlled by conscious action. The behavioral signal captured by the sensor hence includes a (possibly very complex) mixture of the cognitive (behavioral) component and the physical component corresponding to the individual physiology.

Human eyes provide a rich source of information about the identity of a person. The biometric systems utilizing commonly known physical properties of eyes - iris and retinal patterns - achieve high recognition accuracy. The eyes, however, have a strong behavioral component as well, the eye movements. The eye, the *oculomotor plant* and the human visual system develop individually for each person and thus, it is reasonable to hypothesize that some part of the resulting eye-movements is individual too. In this paper we study eye movement features to recognize persons, independently on the task. Thus, eye movements could provide complementary information for iris, retina, or face recognition. In addition, an eye tracker enables also *liveness detection*, that is, validating whether the biometric template originates from a real, living human being.

Another advantage of biometric eye movement authentication would be that it enables unnoticeable *continuous authentication* of a person; the identity of the user can be re-authenticated continuously without any interaction by the user. From the technological viewpoint and accessibility, the accuracy of eye trackers is continuously improving; at the same time, eye trackers have already been integrated with some low-cost webcams, see e.g. [Cogain]. It can be hypothesized that future notebooks and mobile devices will have integrated eye trackers as standard equipment.

*e-mail: {tkinnu,fsedlak,bednarik}@cs.joensuu.fi

ETRA 2010, Austin, TX, March 22 – 24, 2010.
© 2010 ACM 978-1-60558-994-7/10/0003 $10.00

1.1 Related Work

The idea of using eye-movements for person authentication is not new. The currently adopted approaches consider eye gaze as an alternative way to input user-specific PIN numbers or passwords; see [Kumar et al. 2007] for a recent review of such methods. In these systems, the user may input the pass-phrase by, for instance, gaze-typing the password [Kumar et al. 2007], using *graphical passwords* such as human faces [Passfaces], or using *gaze gestures* similar to mouse gestures in web browsers [Luca et al. 2007]. In such systems, it is important to design the user interface (presenting the keypad or visual tokens; providing user feedback) to find acceptable trade-off between security and usability - too complicated cognitive task becomes easily distracting. The eye gaze input is useful in security application where *shoulder surfing* is expected (e.g. watching over one's shoulder when he/she is typing his/her PIN code in an ATM device). However, it is still based on the "what-the-user-knows" type of authentication; it might also be called *cognometric* authentication [Passfaces].

In this paper, we focus on biometric authentication, that is, who the person *is* rather than what he/she knows. Biometric authentication could be combined with the cognometric techniques, or used as the only authentication method in low-security scenarios (e.g. customized user profiles in computers, adaptive user interfaces). We are aware of two prior studies that have utilized physical features of the eye movements to recognize persons [Kasprowski 2004; Bednarik et al. 2005]. In [Kasprowski 2004], the author used a custom-made head-mounted infrared oculography eye tracker ("OBER 2") with sampling rate of 250 Hz to collect data from $N = 47$ subjects. The subjects followed a jumping stimulation point presented on a normal computer screen at twelve successive point positions. Each subject had multiple recording sessions and one task consisted of less than 10 seconds of data and lead to fixed-dimensional (2048 data points) pattern. The stimulation reference signal, together with robust fixation detection, was used for normalizing and aligning the gaze coordinate data. From the normalized signals, several features were derived: average velocity direction, distance to stimulation, eye difference, discrete Fourier transform (DFT) and discrete wavelet transform (DWT). Five well-known pattern matching techniques, as well as their ensemble classifier, were then used for verifying the identity of a person. Average recognition error of the classifier ensemble around 7 % to 8 % was reported.

In an independent study [Bednarik et al. 2005], the authors used a commercially available eye tracker (Tobii ET-1750), built in a TFT panel, to collect data from $N = 12$ subjects at sampling rate of 50 Hz. The stimulus consisted of a single cross shown in the middle of the screen and displayed for 1 second. The eye-movement features consisted of pupil diameter and its time derivative, gaze velocity, and time-varying distance between the eyes. Discrete Fourier transform (DFT), principal component analysis (PCA) and their combination were then applied to derive low-dimensional features, followed by k-nearest neighbor template matching. The time derivative of the pupil size was found to be the best *dynamic* feature, yielding identification rate of 50 % to 60 % (the best feature, distance between the eyes, was not considered interesting since it can be obtained without an eye tracker).

1.2 What is New in This Study: Task Independence

Both of the studies [Kasprowski 2004; Bednarik et al. 2005] are inherently *task-dependent*: they assume that the same stimulus appears in training and testing. This approach has the advantage that the reference template (training sample) and test sample can be accurately aligned. However, the accuracy comes with a price paid on convenience: in security scenario, the user is forced to perform a certain task which becomes both learned and easily distracting. Design of the authentication stimulus should be done with great care to avoid learning effect of the user; if the user learns the task, such as the order of the moving object, his behavior changes which may lead to increased recognition errors. From the security point of view, the repetitious task can easily be copied and intrusion system built to imitate the expected input.

Here we take a step towards eye movement biometrics where we have minimal or absolutely no prior knowledge of the underlying task. We call such problem as *task-independent* eye-movement biometrics. We stem this terminology from other behavioral biometric tasks, such as *text-independent* speaker [Kinnunen and Li 2010] and writer [Bulacu and Schomaker 2007] identification.

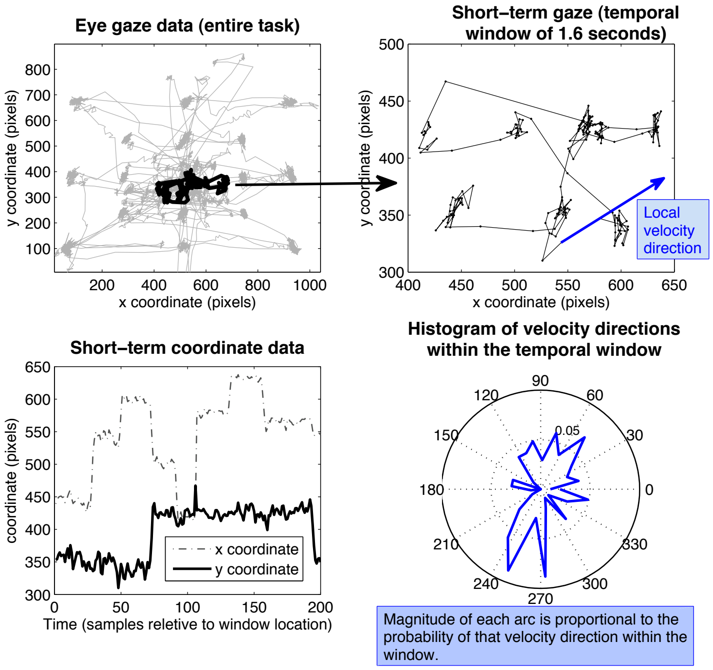
2 Data Preprocessing and Feature Extraction

The eye-coordinates are denoted here by $x_{\text{left}}(t), y_{\text{left}}(t), x_{\text{right}}(t), y_{\text{right}}(t)$ for each timestamp t. Some trackers, including the one used in this study, provide information about pupil diameter of both eyes as well; we focus on the gaze coordinate data only. The eye movement signal is inherently noisy and includes missing data due to blinks, involuntary head movements, or corneal moisture tear film and irregularities, for instance. Typically, the eye-tracking device gives information about the missing data in a validity variable. We assume continuity of the data within the missing parts and linearly interpolate the data in-between the valid segments. Interpolation is applied independently for all four coordinate time series.

It is a known fact that the eye is never still and constantly moves around the fixation center. While most eye-trackers cannot measure the microscopic movements of the eye – the tremor, drift, and microsaccades – due to accuracy and temporal resolution, we concentrate on the movements on a coarser level of detail that can be effectively measured. Even during a single fixation, the eye constantly samples the region of interest. We hypothesize that the way the eye moves during the fixation and smooth pursuit have bearing on some individual property of the oculomotor plant.

We describe the movement as a histogram of all angles the eye travel during a certain period. Similar technique was used for the task-dependent recognition in [Kasprowski 2004]. Figure 2 summarizes the idea. We consider short-term data window (L samples) that expands over a temporal span of few seconds. The local velocity direction of the gaze (from time step t to time step $t + 1$) is computed using trigonometric identities and transformed into a normalized histogram, or discrete probability mass function $\boldsymbol{z} = (p(\theta_1), p(\theta_2), \ldots, p(\theta_K))$, where $p(\theta_k) > 0$ for all k and $\sum_k p(\theta_k) = 1$. The K histogram bin midpoints are pre-located at angles $\theta_k = k(2\pi/K)$ where $k = 0, 1, \ldots, K - 1$. The data window is then shifted forward by S samples (here we choose $S = L/2$). The method produces a time sequence $\{\boldsymbol{z}_t\}, t = 1, 2, \ldots, T$ of K-dimensional feature vectors which are considered independent from each other and used in statistical modeling as explained in the next Section.

Note that in the example of Fig. 2 most of the fixations have a "north-east" or "south-west" tendency and this shows up in the histogram as an "eigendirection" of the velocity. It is worth emphasizing that we deliberately avoid the use of any fixation detection – which is error prone – but instead use all the data within a given window. This data is likely to include several fixations and saccades. Since the amount of time spent on fixating is typically more than 90 %, the relative contribution of the saccades in the histogram is smaller.



Local velocity direction

Magnitude of each arc is proportional to the probability of that velocity direction within the window.

Figure 2: *Illustration of velocity direction features.*

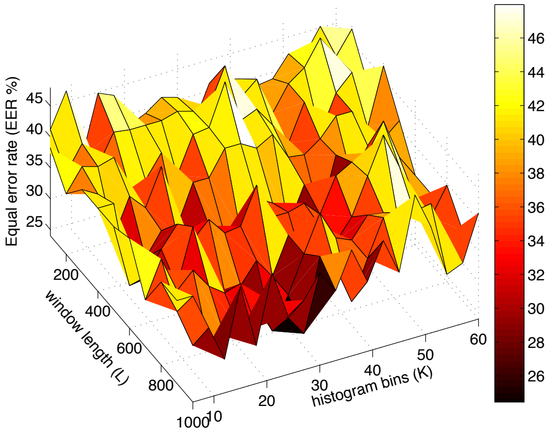


Figure 3: *Effect of feature parameters to accuracy.*

3 Task-Independent Modeling of Features

We adopt some machinery from modern text-independent speaker recognition [Kinnunen and Li 2010]. In speaker recognition, the speech signal is also transformed into a sequence of short-term feature vectors that are considered independent. The matching problem becomes to quantifying the similarity of the given “feature vector clouds”. To this end, each person is modeled using a *Gaussian mixture model* (GMM) [Bishop 2006]. An important advance in speaker recognition was normalizing the speaker models with respect to a so-called *universal background model* (UBM) [Reynolds et al. 2000]. This method (Fig. 1) is now well-established baseline method in that field. The UBM, which is just another Gaussian mixture model, is first trained from a large set of data. The user-dependent GMMs are then derived by adapting the parameters of the UBM to that speaker's training data. The adapted is an interpolated model between the UBM (prior model) and the observed training data. This *maximum a posteriori* (MAP) training of the models gives better accuracy than maximum likelihood (ML) trained models for limited amount of data. Normalization by the UBM likelihood in the test phase also emphasizes the *difference* of the given person from the general population of persons.

The GMM-UBM method [Reynolds et al. 2000] can be shortly summarized as follows. Both the UBM and the user-dependent models are described by a mixture of multivariate Gaussians with some mean vectors $\boldsymbol{\mu}_m$, diagonal covariance matrices $(\boldsymbol{\Sigma}_m)$ and mixture weights (w_m). The density function of GMM is then $p(\boldsymbol{z}|\boldsymbol{\Theta}) = \sum_{m=1}^{M} w_m \mathcal{N}(\boldsymbol{z}|\boldsymbol{\mu}_m, \boldsymbol{\Sigma}_m)$, where $\mathcal{N}(\cdot|\cdot)$ denotes multivariate Gaussian and $\boldsymbol{\Theta}$ denotes collectively all the parameters. The mixture weights satisfy $w_m \geq 0, \sum_m w_m = 1$. The UBM parameters are found via the *expectation-maximization* (EM) algorithm, and the user-dependent mean vectors are derived as,

$$\boldsymbol{\mu}'_m = \alpha_m \tilde{\boldsymbol{z}}_m + (1 - \alpha_m)\boldsymbol{\mu}_m, \tag{1}$$

where $\tilde{\boldsymbol{z}}_m$ is the posterior-weighted mean of the training sample. The adaptation coefficient is defined as $\alpha_m = n_m/(n_m + r)$, where n_m is the soft count of vectors assigned to the m^{th} Gaussian and $r > 0$ is a fixed *relevance factor*. Variances and weights can also be adapted. For small amount of data, the adapted vector is interpolation between the data and the UBM, whereas for large number of data the effect of UBM is reduced. Given a sequence of independent test vectors $\{\boldsymbol{z}_1, \ldots, \boldsymbol{z}_T\}$, the match score is computed as the difference of the target person and the UBM average log-likelihoods:

$$\text{score} = \frac{1}{T}\sum_{t=1}^{T}\{\log p(\boldsymbol{z}_t|\boldsymbol{\Theta}_{\text{target}}) - \log p(\boldsymbol{z}_t|\boldsymbol{\Theta}_{\text{UBM}})\}. \tag{2}$$

4 Experiments

4.1 Experimental Setup

For the experiments, we collected a database of $N = 17$ users. The initial number of participants was higher (23), but we left out the recordings with low quality data - e.g. due to a frequent loss of eye. Participants were naive of the actual purpose of the study. The recordings were conducted in a quiet usability laboratory with constant lighting. After a calibration, the target stimuli consisted of a two sub-tasks. First, the display showed an introductory text with instructions related to the content and duration of the task. Participants were also asked to keep their head as stable as possible. The following stimulus was a 25 minutes long video of a section of *Absolutely fabulous* - a comedy series produced by BBC. The video contained original sound and subtitles in Finnish. The screen resolution of the video was 800 x 600 pixels and it was shown on a 20.1 inch LCD flat panel Viewsonic VP201b (true native resolution 1600 x 1200, anti-glare surface, contrast ration 800:1, response time 16 ms, viewing distance approximately 1 m). A Tobii X120 (sampling rate 120 Hz, accuracy 0.5 degree) eye-tracker was employed in the study and the stimuli were presented using the Tobii Studio analysis software v. 1.5.2.

We use one training file and one test file per person and do cross-matching of all file pairs, which leads to 17 genuine access trials and 272 impostor access trials. We measure the accuracy by *equal error rate* (EER) which corresponds to the operating point for which the false acceptance (accepting an impostor) and false rejection (rejecting a legitimate user) are equal. The UBM training data, person-dependent training data and test data are all disjoint. The UBM is trained by a deterministic splitting method followed by 7 k-means iterations and 2 EM iterations. In creating the user-dependent adapted Gaussian mixture models, we adapt both the mean and variance vectors with relevance factor $r = 16$.

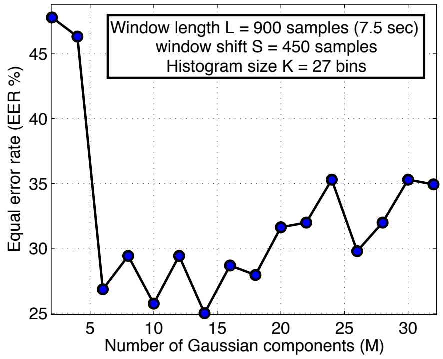


Figure 4: *Effect of model order.*

4.2 Results

We first optimize the feature extractor and classifier parameters. The UBM is trained from 140 minutes of data whereas the training and test segments have approximate durations of five and one minute, respectively. We fix the number of Gaussians to $M = 12$ and vary the length of the data window (L) and the number of histogram bins (K). The error rates presented in Fig. 3 suggest that increasing the window size improves accuracy, and the number of bins does not have to be too high $(20 \leq K \leq 30)$. We fix $K = 27$ and $L = 900$ and further fine-tune the number of Gaussians in Fig. 4. The accuracy improves when increasing the number of Gaussians, and achieves optimum at approximately $6 \leq M \leq 18$. In the following we set $M = 16$.

Finally, we display the error rates for varying lengths of training and test data in Table 1. Here we reduced the amount of UBM training data to 70 minutes so as to allow using longer training and test segments for the target persons. Increasing the amount of data improves accuracy, saturating to accuracy around 30 % EER. Although the accuracy is not sufficient for a realistic application, it is clearly below the chance level (50 % EER), suggesting that there is individual information in the eye movements which can be modeled. The error rates are clearly higher than in [Kasprowski 2004; Bednarik et al. 2005]. However, those studies used fixed-dimensional templates that were carefully aligned whereas our system does not use any explicit temporal alignment.

5 Conclusions

We approached the problem of task-independent eye-movement biometric authentication from bottom-up without any prior signal model of how the resulting eye-movement signal and features are generated by the oculomotor plant. Instead, we had an intuitive guess that the way the ocular muscles operate the eye is individual and there is some independence on the underlying task. Our results indicate that this is a feasible conception. The error rates are too high to be useful in a realistic security system, but significantly lower than the chance level (50 %) which warrants for further studies using larger number of participants and wider range of tasks and features. Several improvements could be done on the data processing side as well: using complementary features from pupil diameter, using feature and score normalization and discriminative training [Kinnunen and Li 2010].

Table 1: *Effect of training and test data durations.*

Training data	Test data	Equal error rate (EER %)
10 sec	10 sec	47.1
1 min	10 sec	47.1
3 min	10 sec	35.3
6 min	10 sec	29.8
9 min	10 sec	28.7
1 min	1 min	41.9
2 min	2 min	41.2
4 min	4 min	29.4
5 min	3 min	29.4
6 min	2 min	29.4
7 min	1 min	29.8

References

Bednarik, R., Kinnunen, T., Mihaila, A., and Fränti, P. 2005. Eye-movements as a biometric. In *Proc. 14th Scandinavian Conference on Image Analysis (SCIA 2005)*. 780-789.

Bishop, C. 2006. *Pattern Recognition and Machine Learning*. Springer Science+Business Media, LLC, New York.

Bulacu, M., and Schomaker, L. 2007. Text-independent writer identification and verification using textual and allographic features. *IEEE Trans. on Pattern Analysis and Machine Intelligence 29*, 4 (April), 19–41.

Cogain. Open source gaze tracking, freeware and low cost eye tracking. WWW page, `http://www.cogain.org/eyetrackers/low-cost-eye-trackers`.

Kasprowski, P. 2004. *Human Identification Using Eye Movements*. PhD thesis, Silesian University of Technology, Institute of Computer Science, Gliwice, Poland. `http://www.kasprowski.pl/phd/PhD_Kasprowski.pdf`.

Kinnunen, T., and Li, H. 2010. An overview of text-independent speaker recognition: from features to supervectors. *Speech Communication 52*, 1 (January), 12–40.

Kumar, M., Garfinkel, T., Boneh, D., and Winograd, T. 2007. Reducing shoulder-surfing by using gaze-based password entry. In *Proc. of the 3rd symposium on Usable privacy and security table of contents*, ACM, 13–19.

Luca, A. D., Weiss, R., and Drewes, H. 2007. Evaluation of eye-gaze interaction methods for security enhanced PIN-entry. In *Proc. of the 19th Australasian conf. on Computer-Human Interaction: Entertaining User Interfaces*, ACM, 199–202.

Passfaces. Passfaces: Two factor authentication for the enterprise. WWW page, `http://www.passfaces.com/`.

Reynolds, D., Quatieri, T., and Dunn, R. 2000. Speaker verification using adapted gaussian mixture models. *Digital Signal Processing 10*, 1 (January), 19–41.

Salvucci, D., and Goldberg, J. 2000. Identifying fixations and saccades in eye-tracking protocols. In *ETRA '00: Proceedings of the 2000 symposium on Eye tracking research & applications*, ACM, New York, NY, USA, 71–78.