Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 00 (2020) 000–000

Procedia Computer Science

www.elsevier.com/locate/procedia

24th International Conference on Knowledge-Based and Intelligent Information & Engineering Systems

On the Relationship between Eye Tracking Resolution and Performance of Oculomotoric Biometric Identification

Paul Prasse*, Lena A. Jäger, Silvia Makowski, Moritz Feuerpfeil, Tobias Scheffer

University of Potsdam, Department for Computer Science, August-Bebel-Str. 89, 14482 Potsdam, Germany

Abstract

Distributional properties of fixations and saccades are known to constitute biometric characteristics. Additionally, high-frequency micro-movements of the eyes have recently been found to constitute biometric characteristics that allow for faster and more robust biometric identification than just macro-movements. Micro-movements of the eyes occur on scales that are very close to the precision of currently available eye trackers. This study therefore characterizes the relationship between the temporal and spatial resolution of eye tracking recordings on one hand and the performance of a biometric identification method that processes micro- and macro-movements via a deep convolutional network. We find that that the deteriorating effects of decreasing both, the temporal and spatial resolution are not cumulative. We observe that on low-resolution data, the network reaches performance levels above chance and outperforms statistical approaches.

© 2020 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of KES International.

Keywords: biometrics; eye tracking; deep learning; data quality

1. Introduction

Eye movements exhibit strong individual characteristics that have been demonstrated to be relatively stable over time [2]. Hence, Kasprowski and Ober [15] proposed to use eye movements for biometric identification. Since this seminal work, eye tracking-based identification has been attracting increasing attention and several competitions [14, 13, 16] have shown the promise of eye movements as behavioral biometric characteristic. Until recently, the main drawback was that the time-to-identification of oculomotoric identification was several orders of magnitude above the time needed by competing biometric technologies such as face recognition, fingerprints, or iris recognition.

Recent work has shown that end-to-end trainable deep neural networks that process raw eye tracking signals can exploit micro-movements of the eyes; since these micro-movements occur at high frequency, using them accelerates the

* Corresponding author.
E-mail address: prasse@uni-potsdam.de

1877-0509 © 2020 The Author(s). Published by Elsevier B.V.
This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)
Peer-review under responsibility of KES International.

identification process by two orders of magnitude [12, 19]. A critical limitation of the *DeepEyedentification* method is that micro-movements are barely measurable even with high-frequency, high-precision eye tracking-systems. To assess the applicability of high-frequency, low-amplitude micro-movements as a biometric characteristic in real-world settings, this paper investigates the impact of the temporal and spatial resolution of the eye tracking signal on the performance of oculomotoric biometric identification based on deep learning [12, 19].

Vision research distinguishes three major types of eye movements. During *fixations* of around 250 ms, the eye is relatively still and visual input is obtained. *Saccades*, whose average duration is around 50 ms, are fast relocation movements from one fixation to the next of up to 500 °/s. During a *smooth pursuit*, the eye gaze follows a slowly moving visual target. Besides these macroscopic eye movements, three kinds of ocular micro-movements occur during a fixation. Very slow ocular *drift* away from the intended center of a fixation of around 0.1-0.4°/s is superimposed by high-frequency, low-amplitude *tremor* of around 40-100 Hz and a velocity of up to 0.3°/s. *Microsaccades* are tiny saccades of up to 120°/s that occur occasionally during intended fixation [21, 22, 23, 24, 11].

Prior work on oculomotoric biometric identification typically preprocesses the eye tracking signal into saccades and fixations and subsequently extracts explicit features, such as the duration of fixations, or the amplitude, velocity, or acceleration of saccades. Whereas earlier approaches compare eye movement sequences by first aggregating the features over the relevant eye movement recordings [8, 3, 28], statistical approaches compare the distributions of the features extracted from two eye gaze sequences [25, 7, 26] and generative approaches simulate a user's scanpath (i.e., a sequence of fixations and saccades) using Bayesian Graphical models [30, 29, 17, 1, 18]. *DeepEyedentification*, by contrast, does not apply any preprocessing and feature extraction, but rather operates directly on the raw signal recorded from an eye tracker [12, 19]. This approach is summarized in Section 2.

Prior work on the relationship between the spatial and temporal resolution of the eye tracking signal and the performance of oculomotoric biometric identification [9, 10] investigates the impact of spatial noise and temporal resolution of the eye tracking signal on the performance of the statistical approach proposed by Holland and Komogortsev [7] that uses distributional attributes of saccades and fixations. It is not clear–and does not even appear likely—that this relationship between resolution and accuracy will be the same for the DeepEyedentification method that exploits ocular micro-movements. In our experiments, we use this statistical approach and its successor method of Rigas *et al.* [26] as reference methods.

The remainder of this paper is structured as follows. Section 2 reviews the state-of-the-art of deep learning-based oculomotoric biometrics, Section 3 lays out the problem setting, Section 4 describes the data sets used for the experiments presented in Section 5, Section 6 concludes.

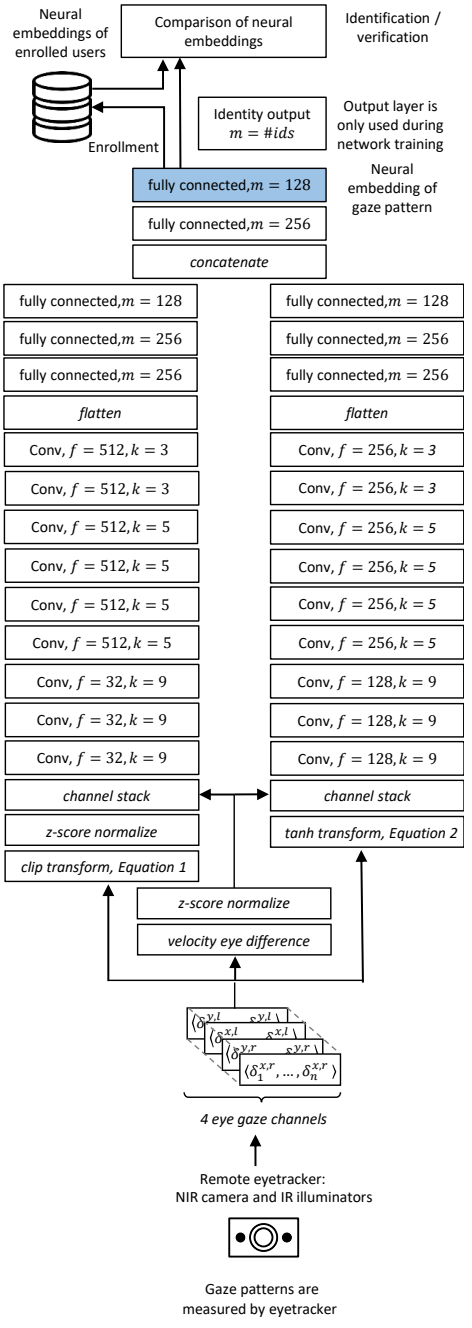


Fig. 1. *DeepEyedentificationLive architecture without liveness detection. Figure adjusted from Makowski* et al. *[19].*

2. Oculomotoric Biometrics Based on Deep Learning

This section summarizes the *DeepEyedentificationLive* architecture [19] which is a binocular extension of the *DeepEyedentification* architecture [12]. DeepEyedentificationLive is also able to process the stimulus and detect presentation attacks; however, we do not use this function in this paper.

2.1. Architecture of the DeepEyedentificationLive network

An eye tracker records binocular gaze sequences of absolute yaw x and pitch gaze angles y of the left l and right eye r at a sampling frequency of ρ, measured in Hz. The binocular DeepEyedentificationLive network (see Figure 1) receives yaw δ_i^x and pitch δ_i^y gaze velocities in $^\circ/s$ as input which are computed from the recorded gaze sequence as $\delta_i^x = \frac{\rho}{2}(x_{i+1} - x_{i-1})$ and $\delta_i^y = \frac{\rho}{2}(y_{i+1} - y_{i-1})$ for the left and the right eye, respectively. This results in a total of four input channels, namely the sequence of yaw $\langle \delta_1^{x,l}, \ldots, \delta_n^{x,l} \rangle$ and pitch angular velocities of the left eye $\langle \delta_1^{y,l}, \ldots, \delta_n^{y,l} \rangle$ and the corresponding yaw $\langle \delta_1^{x,r}, \ldots, \delta_n^{x,r} \rangle$ and pitch angular velocities of the right eye $\langle \delta_1^{y,r}, \ldots, \delta_n^{y,r} \rangle$.

The network processes input sequences of 1,000 time steps corresponding to 1 s of 1000 Hz eye tracking recording. For our experiments, we adjust the length of the input layer such that the network always processes 1 s of eye tracking recording, independently of the sampling frequency at which the data has been recorded.

The key feature of the network's architecture is that the input channels are duplicated and directed into two separate convolutional subnets. The *fast subnet* is designed to process the high angular velocities of (micro-) saccadic eye movements whereas the *slow subnet* is designed to process the slow fixational eye movements (drift and tremor). Each of the subnets is preceded by a transformation layer that applies a transformation to the input to resolve the fast saccadic and slow fixational eye movements, respectively. For the fast subnet, saccadic eye movements are resolved by applying a clipping function that truncates velocities below a threshold ν_{min} to zero and a subsequent z-score normalization (see Equation 1). Based on hyperparameter tuning within a range of psychologically plausible parameters on two independent data sets [12, 19], the velocity threshold ν_{min} is set to $40^\circ/s$. The original DeepEyedentificationLive network also processes the stimulus sequence and provides a liveness-detection output. We train a version of the network without the stimulus input channels and liveness output.

$$t_f(\delta_i^x, \delta_i^y) = \begin{cases} z(0) & \text{if } \sqrt{\delta_i^{x2} + \delta_i^{y2}} < \nu_{min} \\ (z(\delta_i^x), z(\delta_i^y)) & \text{otherwise} \end{cases} \tag{1}$$

The slow fixational eye movements are resolved by applying a sigmoidal function that stretches the slow velocities of drift and tremor approximately within the interval between -0.5 and $+0.5$ and squashes the (micro-) saccadic velocities to the interval between -0.5 and -1 or $+0.5$ and $+1$, depending on their direction (see Equation 2). Independent hyperparameter optimization on two data sets showed that an appropriate value for the scaling factor c of Equation 2 is 0.02 [12, 19].

$$t_s(\delta_i^x, \delta_i^y) = (\tanh(c\delta_i^x), \tanh(c\delta_i^y)) \tag{2}$$

Since binocular alignment is an informative individual characteristics, the four untransformed input velocity channels are also fed into a subtraction layer which computes the yaw $\langle \delta_1^{x,r} - \delta_1^{x,l}, \ldots, \delta_n^{x,r} - \delta_n^{x,l} \rangle$ and pitch velocity differences between the two eyes $\langle \delta_1^{y,r} - \delta_1^{y,l}, \ldots, \delta_n^{y,r} - \delta_n^{y,l} \rangle$. After each of the transformation layers, a stacking layer is inserted which stacks these additional two channels with the input of each of the two subnets.

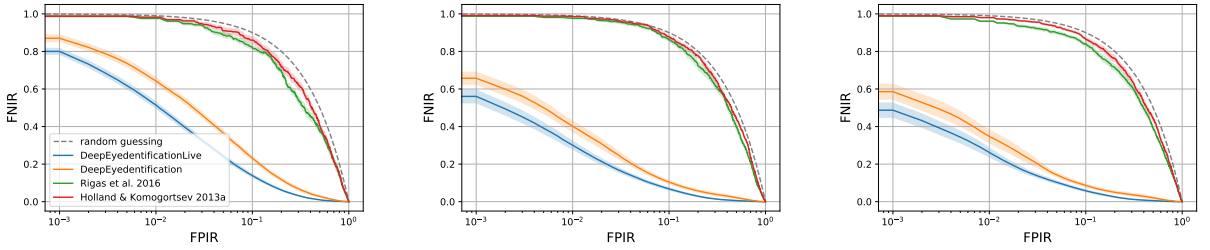
The two subnets share the same number and type of layers. Each of the subnets consists of a series of one-dimensional convolutional layers, where the convolutions are applied to the six input channels over the temporal axis. The number of filters and kernel size of the convolutional layers (f and k in Figure 1), as well as the number of units of the subsequent fully connected layers (m in Figure 1), are allowed to differ between the two subnets. For our experiments, we use the same hyperparameter values as Makowski *et al.* [19] (see Figure 1). After each of the convolutional and fully connected layers, batch normalization and ReLU activation is applied. An average pooling layer with pooling size 2 and stride size 1 is inserted after each convolutional layer.

For training, a softmax output layer with one unit for each user in the training data is added. Using categorical cross-entropy as loss function, the network is trained to predict a viewer's identity from an eye tracking sequence.

Once training is completed, the softmax output layer is removed and the activation of the last fully connected layer (highlighted in blue in Figure 1) is used as neural feature embedding of an input gaze sequence. At application time, the similarity between an enrolment and a test sequence is computed as the cosine similarity of their neural embeddings, averaged over all input windows of 1,000 ms.

2.2. Performance of the DeepEyedentificationLive network

Makowski *et al.* compared the performance of the DeepEyedentification network [12] and its binocular extension DeepEyedentificationLive [19] with the statistical approaches by Holland and Komogortsev (2013) [7] and Rigas *et al.* (2016) on the *JuDo1000* data set (see Section 4). Figure 2 shows the results for the identification of 20 enrolled users in the presence of five impostors after seeing 1, 5, and 10 seconds of eye tracking recording at test time. The deep learning-based methods outperform the statistical state-of-the-art approaches. Moreover, the DeepEyedentification network is outperformed by its binocular extension.



(a) 1 s of test recording, 1000 Hz (b) 5 s of test recording, 1000 Hz (c) 10 s of test recording, 1000 Hz

Fig. 2. Performance of state-of-the-art methods for oculomotoric identification on the *Judo1000* data set. False-Negative Identification-Error Rate (FNIR) over False-Positive Identification-Error (FPIR). Colored bands show the standard error. Figure adapted from Makowski *et al.* [19].

3. Problem setting

We study the influence of spatial and temporal resolution on the performance of the oculomotoric biometric system described in Section 2. At test time, the model observes a sequence of yaw and pitch gaze angles of the left eye $\langle (x_1^l, y_1^l), ..., (x_n^l, y_n^l) \rangle$ and the right eye $\langle (x_1^r, y_1^r), ..., (x_n^r, y_n^r) \rangle$ recorded by an eye tracker. The spatial resolution, or *precision*, of the eye tracker quantifies the reliability of the eye tracker; i.e., how well it is able to reproduce its measurements. It can be estimated by the standard deviation of a set of n samples x_i and sample average $\bar{x}$ with a constant true gaze direction as in Equation 3. It is usually measured with an artificial eye.

$$precision = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}. \tag{3}$$

We study the influence of precision on the system performance by adding Gaussian noise with increasing standard deviation to the input signal. Temporal resolution is varied by downsampling the input signal at different levels. Figure 3 shows an eye trace at different levels of precision and temporal resolution.

We evaluate system performance for biometric identification. The system decides whether or not an observed sequence matches the identity of one out of a set of enrolled users. The input sequence is compared to enrollment sequences of all enrolled users. In the case of a match, the system returns the identity and otherwise classifies the user as an impostor. Varying the decision threshold gives a DET curve of false-negative identification-error rate (FNIR) over false-positive identification-error rate (FPIR). The equal error rate (EER) is the point in the DET curve for which FNIR equals FPIR.

The similarity of two gaze sequences is computed by the cosine similarity of their neural network embeddings. The network is trained on a separate set of training users and is optimized to create similar embeddings for sequences from the same user and dissimilar embeddings for sequences from different users.

4. Data sets

For our experiments, we used the following four eye-tracking data sets (see Table 1). The *JuDo1000* data set [19, 20] consists of binocular eye tracking data (horizontal and vertical gaze coordinates) recorded with an Eyelink Portable Duo eye tracker with a vendor-reported spatial precision of 0.01° at a sampling frequency of 1,000 Hz using a chin rest to stabilize the participant's head. Each of 150 subjects participated in four experimental sessions with a temporal lag of at least one week between any two sessions. In each session, participants viewed a total of 108 trials in each of which a black dot jumped to five random locations. The duration for which the dot remained in one location varied between different trials (250, 500, and 1000 ms). The *JuDo1000* data set can be downloaded from the Open Science framework database [1].

The *CEM-I* data set [9] consists of binocular eye-tracking data recorded with a Tobii TX300 eye tracker with a vendor-reported spatial precision of 0.09° at a sampling frequency of 300 Hz. A set of 22 participants were presented with up to 8 trials of different kinds of visual stimuli, namely a *simple pattern* (SIM), a *complex pattern* (COM), a *cognitive pattern* (COG) and a *textual pattern* (TEX)—see Holland and Komogortsev [9] for details.

The *CEM-II* data set [9] consists of right-eye monocular data recorded at 1,000 Hz using an Eyelink 1000 tracker, which has a vendor-reported spatial precision of 0.01°. A population of 32 participants were presented with four experimental trials in which they read the same textual pattern that was used for the *CEM-I* data set.

The *CEM-III* data set [9] was recorded with a PlayStation Eye camera at 75 Hz. A total of 173 participants were presented with two trials of the same visual stimuli used for *CEM-I*, except for the cognitive pattern which has been replaced by a *random pattern* (RAN).

Table 1. Summary of the data sets showing the eye tracking device, its vendor-reported spatial precision in degrees of visual angle, the sampling frequency of the recording in Hz, whether the data is binocular or monocular, whether participants' heads were stabilized with a chin rest, the number of subjects, the number of experimental sessions, the number of trials per subject, the average trial duration with standard deviation in seconds, and the recording time per subject with standard deviation in seconds.

Data set	Device	Prec.	Freq.	Eye(s)	Chin rest	# subj.	# sess.	# trials	Trial dur.	Rec./subj.
JuDo1000	EyeLink Portable Duo	0.01	1000	both	yes	150	4	432	3 ± 1.6	1260 ± 0
CEM-I (SIM)	Tobii TX300	0.09	300	both	yes	22	1	8	64 ± 47	465 ± 288
CEM-I (COM)	Tobii TX300	0.09	300	both	yes	22	1	2	207 ± 17	386 ± 75
CEM-I (COG)	Tobii TX300	0.09	300	both	yes	22	1	2	99 ± 51	180 ± 95
CEM-I (TEX)	Tobii TX300	0.09	300	both	yes	22	1	4	40 ± 20	148 ± 71
CEM-II (TEX)	EyeLink 1000	0.01	1000	right	yes	32	1	4	51 ± 10	192 ± 38
CEM-III (SIM)	PlayStation Eye	N/A	75	right	yes	173	1	2	89 ± 8	177 ± 18
CEM-III (COM)	PlayStation Eye	N/A	75	right	yes	173	1	2	133 ± 16	264 ± 32
CEM-III (RAN)	PlayStation Eye	N/A	75	right	yes	164	1	2	77 ± 7	154 ± 13
CEM-III (TEX)	PlayStation Eye	N/A	75	right	yes	172	1	2	46 ± 5	91 ± 10

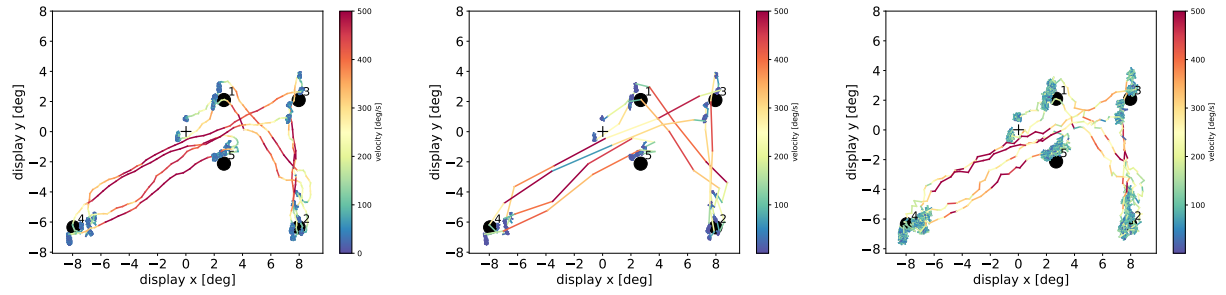
5. Experiments

This section presents the results of our experiments with varied spatial and temporal resolution and varied eye tracking hardware using the data sets described in Section 4 for DeepEyedentificationLive and reference methods.

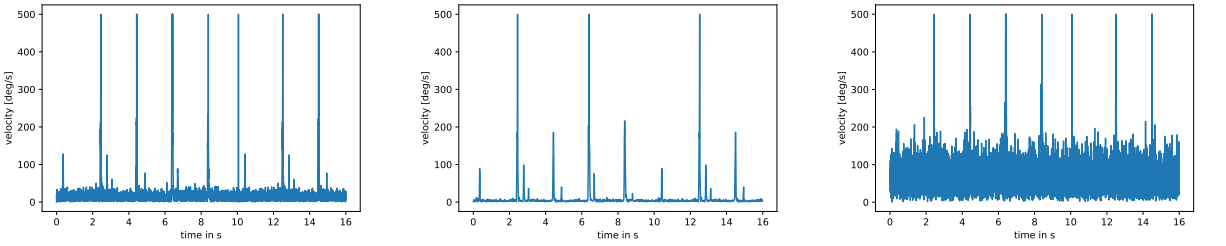
5.1. Varying the Resolution of the Eye Tracking Data

For these experiments, we use the evaluation protocol of Makowski *et al.* [19]: We resample 20 times from the data set, each time selecting 125 users to train an embedding and a disjoint set of 25 users (20 enrolled users, 5 impostors) to evaluate the system. As enrollment data, we randomly select three trials from each of the first three recording sessions. For testing, we use 1, 5, or 10 seconds of eye tracking recording from the fourth session.

[1] https://osf.io/5zpvk/



(a) 1000 Hz, Precision = 0.01° (b) 125 Hz, Precision = 0.01° (c) 1000 Hz, Precision = 0.1°



(d) 1000 Hz, Precision = 0.01° (e) 125 Hz, Precision = 0.01° (f) 1000 Hz, Precision = 0.1°

Fig. 3. Exemplary eye trace and its velocity profile of one trial taken from the *JuDo1000* data set for three different configurations: 1000 Hz original data (a,d), 125 Hz downsampled data (b,e), and the sequence with Gaussian noise added for a resulting precision of 0.1° (c,f). The fixation cross in the center of the screen is displayed before the onset of the trial.

5.1.1. Varied Temporal Resolution

We downsample the *JuDo1000* data to produce data sets of 1000, 500, 250, 125, 62, and 31 Hz. We evaluate the performance of the DeepEyedentificationLive network on these data sets according to the protocol defined in Section 5.1. Figure 4 and Table 2 show the results for one, five and ten seconds of eye tracking recording used for testing. We observe that the EER approximately triples when the sampling rate degrades from 1,000 Hz to 31 Hz. Identification rates for 31 Hz are still above chance level for 20 enrolled individuals and 5 impostors.

Table 2. Varying temporal resolution. Performance of the DeepEyedentificationLive network on the (downsampled) *JuDo1000* data set with original precision (0.01°). EER with standard error for different sampling frequencies and different durations of the input used for testing.

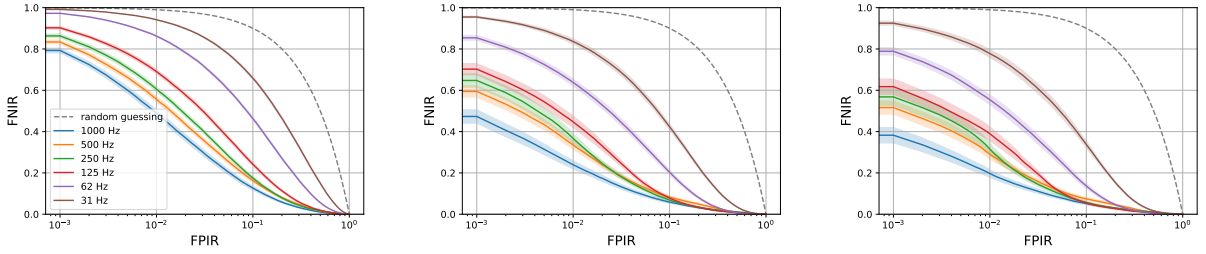
Test recording	1000 Hz	500 Hz	250 Hz	125 Hz	62 Hz	31 Hz
1 s	0.112 ± 0.0043	0.13 ± 0.0042	0.132 ± 0.0041	0.1566 ± 0.0043	0.231 ± 0.0047	0.313 ± 0.0060
5 s	0.073 ± 0.0031	0.091 ± 0.0033	0.082 ± 0.0035	0.09 ± 0.0034	0.138 ± 0.0041	0.211 ± 0.0055
10 s	0.067 ± 0.0029	0.085 ± 0.0032	0.074 ± 0.0035	0.076 ± 0.0032	0.113 ± 0.0039	0.178 ± 0.0050

5.1.2. Varied Spatial Resolution

We vary the precision of the data by adding normally distributed noise with standard deviation s_{noise} 0.0, 0.028, 0.05, 0.1, 0.3 and 0.5°. Given a vendor-reported precision of the eye tracker (measured with an artificial eye) s_x of 0.01°, Equation 4 yields a resulting precision of 0.01, 0.03, 0.05, 0.1, 0.3 and 0.5°, respectively:

$$precision' = \sqrt{precision^2 + s_{noise}^2}. \tag{4}$$

The DeepEyedentificationLive network is evaluated according to the evaluation protocol detailed in Section 5.1 using data with the same precision for training and testing. Figure 5 and Table 3 show the results for one to ten



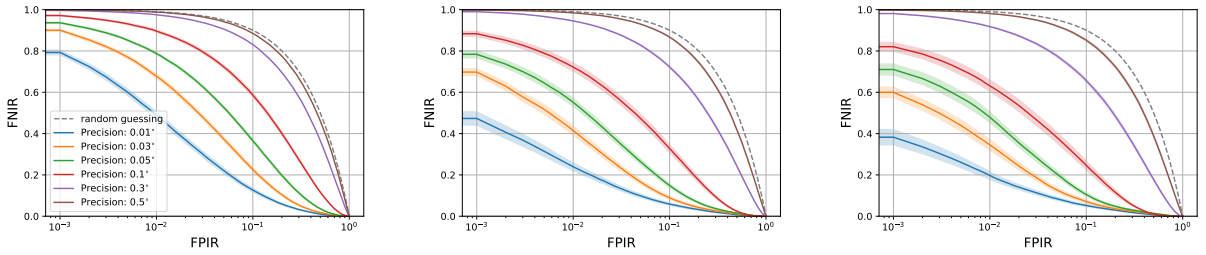
(a) 1 s of test recording (b) 5 s of test recording (c) 10 s of test recording

Fig. 4. Varying levels of temporal resolution. Performance of the DeepEyedentificationLive network on the (downsampled) *Judo1000* data set. False-Negative Identification-Error Rate (FNIR) over False-Positive Identification-Error Rate (FPIR). Colored bands show the standard error.

seconds of input data available at test time. The EER less than triples when the precision increases by one order of magnitude and remains above chance level for 20 enrolled individuals at 0.5°.

Table 3. Varying spatial resolution. Performance of the DeepEyedentificationLive network on the *JuDo1000* data set with original sampling frequency (1000 Hz) added spatial noise. EER with standard error for different levels of precision in degrees of visual angle and different durations of the input used for testing in seconds.

Test recording	0.01°	0.03°	0.05°	0.1°	0.3°	0.5°
1 s	0.112 ± 0.0043	0.15 ± 0.005	0.12 ± 0.0051	0.287 ± 0.0075	0.433 ± 0.004	0.481 ± 0.0036
5 s	0.073 ± 0.003	0.094 ± 0.0039	0.121 ± 0.004	0.186 ± 0.0075	0.365 ± 0.0074	0.46 ± 0.0076
10 s	0.067 ± 0.0029	0.083 ± 0.0037	0.101 ± 0.0038	0.153 ± 0.0072	0.327 ± 0.009	0.446 ± 0.0102



(a) 1 s of test recording (b) 5 s of test recording (c) 10 s of test recording

Fig. 5. Varying levels of spatial resolution. Performance of the DeepEyedentificationLive network on the *Judo1000* data set with added spatial noise. Colored bands show the standard error.

5.1.3. Reduced Temporal and Varied Spatial Resolution

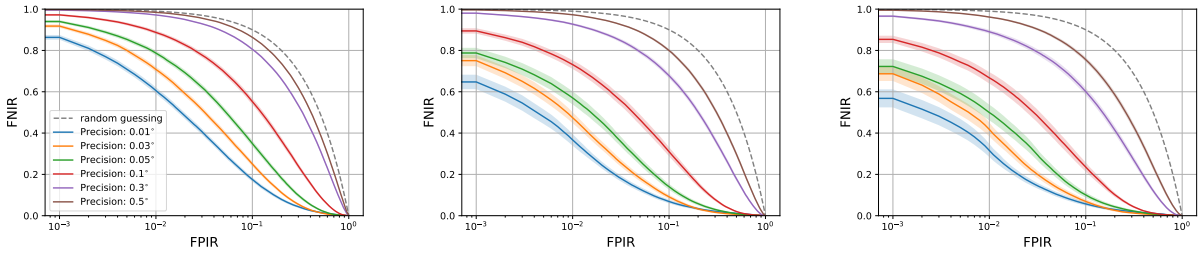
We downsample the *JuDo100* data to 250 Hz and add varying levels of spatial noise to the data as for the experiments described in Section 5.1.2. We evaluate the performance of the DeepEyedentificationLive network on the resulting data sets according to the protocol described in Section 5.1 for varied spatial resolution of the data. Figure 6 and Table 4 show the results for one to ten seconds of test data. We observe that the impact of degraded temporal and spatial resolution are not cumulative; for a precision of 0.5°, the EER at 31 Hz is even lower than at 1,000 Hz.

5.2. Varied Eye Tracking Hardware

We compare the performance of the DeepEyedentificationLive network on the *JuDo1000* data with its performance on the three *CEM* data sets, which were recorded using different eye tracking hardware (see Section 4). As reference

Table 4. Reduced temporal and varying spatial resolution. Performance of the DeepEyedentificationLive network on the *JuDo1000* data set downsampled to 250 Hz with added spatial noise. EER with standard error for different levels of precision in degrees of visual angle and different durations of the input used for testing in seconds.

Test recording	0.01°	0.03°	0.05°	0.1°	0.3°	0.5°
1 s	0.132 ± 0.0041	0.156 ± 0.0039	0.192 ± 0.0054	0.271 ± 0.0074	0.411 ± 0.0051	0.451 ± 0.0042
5 s	0.082 ± 0.003	0.0952 ± 0.0032	0.116 ± 0.0044	0.175 ± 0.006	0.331 ± 0.0083	0.399 ± 0.0083
10 s	0.074 ± 0.0035	0.083 ± 0.0031	0.098 ± 0.0043	0.146 ± 0.0055	0.292 ± 0.0094	0.369 ± 0.0106



(a) 1 s of test recording (b) 5 s of test recording (c) 10 s of test recording

Fig. 6. Reduced temporal and varying spatial resolution. Performance of the DeepEyedentificationLive network on the *Judo1000* data set downsampled to 250 Hz with added spatial noise. False-Negative Identification-Error Rate (FNIR) over False-Positive Identification-Error Rate (FPIR). Colored bands show the standard error.

methods, we re-implement a binocular version of the statistical approaches by Holland and Komogortsev [7] and Rigas *et al.* [26]. We use the same evaluation protocol as Holland and Komogortsev [10]: For each stimulus type separately, we perform 20 iterations of random resampling in each of which we use half of the participants for training and the other half for testing. At application time, we iterate over all trials in the test set, using this trial as actual test instance and all remaining data from the test set for enrollment. Instances for which the similarity between the test trial and enrollment data of the correct identity subject exceeds both the threshold and the similarity between the test trial and all enrolment data from any other subject count as correct identification.

The methods of Holland and Komogortsev and Rigas require the input data to be pre-processed into saccades and fixations. For the data sets collected at 300 Hz or higher, we use a velocity-based saccade detection algorithm [5, 6, 4] with a minimal fixation duration of 20 ms and a minimal (micro-)saccade duration of 6 ms; whereas for the CEM-III data set collected at 75 Hz, we use a dispersion-threshold algorithm for fixation detection [27]. Following Holmqvist *et al.*, we set the dispersion threshold to 2° and the duration threshold to 80 ms [11]. Holland and Komogortsev use a velocity-based algorithm for all three data sets with the minimal fixation duration set to 100 ms [10].

Table 5 shows the identification accuracy of the different methods for each of the *CEM* data sets. The diverging results of our own evaluation and the one reported by Holland and Komogortsev [10] might be due to differences in the preprocessing algorithms and the threshold parameters (see above). The main difference is that we label microsaccades as saccades, whereas Holland and Komogortsev treat them as part of a fixation. Besides the preprocessing and possible differences in the implementational details, the fusion metrics used to combine the different fixational and saccadic features differs between our implementation and the one used by Holland and Komogortsev for the evaluation of their model [7] on the *CEM* data sets [10]: Whereas the latter train a linear model to weight the different features, we apply the simple mean metrics as used by Rigas *et al.* [26] in their main evaluation of their model and the method of Holland and Komogortsev [7].

Second, we evaluate DeepEyedentifcationLive network on the CEM-III data (75 Hz). Since the CEM-III data is monocular, we duplicate the data from the right eye as input for the left-eye channels. Note that it is not possible to apply this evaluation setting on the CEM-I and CEM-II data sets because of their limited number of subjects.

Figure 7 shows the false-negative identification rate over the false-positive identification rate for one to ten seconds of eye tracking recording available at test time. The EER is 0.289, 0.249 and 0.24 for one, five and ten seconds of test recording, respectively.

Table 5. Identification accuracies ± standard error (in %) on the CEM data sets with different stimulus types. The first row presents the average duration of one subject's recorded data. Whereas the third and fourth rows present our re-implementation of Rigas et al. (2016) and Holland and Komogortsev (2013), the bottom row shows the numbers reported by Holland and Komogortsev [10, Table 4].

Method	Experiment / Stimulus Combination								
	CEM-I 300 Hz, N=22				CEM-II 1 kHz, N=32	CEM-III 75 Hz, N=173			
	SIM	COM	COG	TEX	TEX	SIM	COM	RAN	TEX
Avg. recording/subj. in s	465	386	180	148	192	177	264	154	91
DeepEyedentificationLive	63 ± 2	**67 ± 3**	**56 ± 2**	**75 ± 2**	**80 ± 2**	**34 ± 1**	**37 ± 1**	**43 ± 1**	**35 ± 1**
Rigas et al., 2016 [26] (ours)	67 ± 15	37 ± 8	55 ± 12	41 ± 9	78 ± 17	8 ± 2	8 ± 2	5 ± 1	6 ± 1
H & K, 2013 [7] (ours)	**68 ± 15**	31 ± 7	48 ± 11	33 ± 7	71 ± 16	8 ± 2	7 ± 2	5 ± 1	5 ± 1
H & K, 2013 [7]	53	22	19	31	38	7	5	5	4

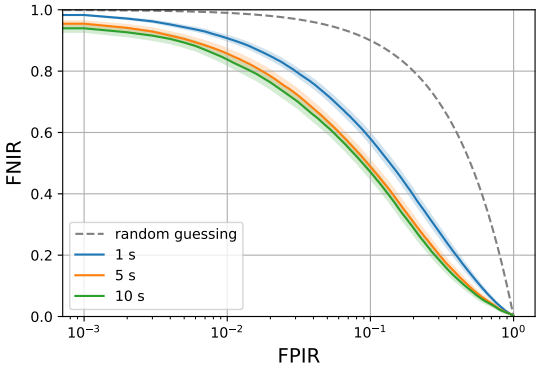


Fig. 7. Performance of the DeepEyedentificationLive network on the *CEM-III* data set (75 Hz) for 1, 5 and 10 seconds of eye tracking recording available at test time. Colored bands show the standard error.

6. Conclusion

We observe that the performance of oculomotoric biometric identification based on deep learning depends on the temporal and spatial resolution at which the eye-gaze is tracked. The performance degrades remarkably gently with decreasing sampling rate and decreasing precision: For 5 seconds of test data, the EER increases from 0.073 at 1,000 Hz to 0.082 at 250 Hz and 0.09 at 125 Hz. Even at only 31 Hz, the EER of 0.211 has less than tripled. The EER decreases from 0.073 for 0.01° to 0.121 for 0.05°, and still remains above chance for 0.5°.

Moreover, we can conclude that decreasing both, the temporal and spatial resolution, does not have additive deteriorating effects—on data with lower sampling frequency, the model even appears to be more robust against spatial noise. For example, at 250 Hz, the network still reaches an EER smaller than 0.1 at a spatial precision of 0.03°.

The experiments on the *CEM* data sets show that deep learning-based oculomotoric identification outperforms statistical approaches independently of the specific hardware used and for various kinds of visual stimuli. The experiments on the *CEM* data sets furthermore show that the DeepEyedentificationLive network benefits from larger amounts of training data. The comparatively good results on the *CEM-III* data set, which has a large number of participants but very few data per participant, indicates that the number of training users is critical for the network to learn an informative feature embedding.

We conclude that deep-learning-based oculomotoric biometric identification is most accurate when high-resolution data is available, but even on low-resolution data, it still reaches accuracy levels far beyond chance. This opens the possibility to integrate oculomotoric biometrics into face recognition or iris scanning systems with relatively low-grade cameras, such as cameras of mobile devices.

Acknowledgments

This work was partially funded by the German Science Foundation under grant SFB 1294.

References

[1] Abdelwahab, A., Kliegl, R., Landwehr, N., 2016. A semiparametric model for Bayesian reader identification, in: EMNLP 2016, pp. 585–594.
[2] Bargary, G., Bosten, J.M., Goodbourn, P.T., Lawrance-Owen, A.J., Hogg, R.E., Mollon, J., 2017. Individual differences in human eye movements: An oculomotor signature? Vision Research 141, 157–169.
[3] Cuong, N., Dinh, V., Ho, L.S.T., 2012. Mel-frequency cepstral coefficients for eye movement identification, in: ICTAI 2012, pp. 253–260.
[4] Engbert, R., 2006. Microsaccades: A microcosm for research on oculomotor control, attention, and visual perception. Progress in Brain Research 154, 177–192.
[5] Engbert, R., Kliegl, R., 2003. Microsaccades uncover the orientation of covert attention. Vision Research 43, 1035–1045.
[6] Engbert, R., Mergenthaler, K., 2006. Microsaccades are triggered by low retinal image slip. Proceedings of the National Academy of Sciences of the U.S.A. 103, 7192–7197.
[7] Holland, C., Komogortsev, O., 2013a. Complex eye movement pattern biometrics: Analyzing fixations and saccades, in: ICB 2013.
[8] Holland, C., Komogortsev, O.V., 2011. Biometric identification via eye movement scanpaths in reading, in: IJCB 2011, pp. 1–8.
[9] Holland, C.D., Komogortsev, O.V., 2012. Biometric verification via complex eye movements: The effects of environment and stimulus, in: BTAS 2012, pp. 39–46.
[10] Holland, C.D., Komogortsev, O.V., 2013b. Complex eye movement pattern biometrics: The effects of environment and stimulus. IEEE Transactions on Information Forensics and Security 8, 2115–2126.
[11] Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., Van de Weijer, J., 2011. Eye tracking: A comprehensive guide to methods and measures. Oxford University Press, Oxford.
[12] Jäger, L.A., Makowski, S., Prasse, P., Liehr, S., Seidler, M., Scheffer, T., 2020. Deep Eyedentification: Biometric identification using micro-movements of the eye, in: ECML PKDD 2019.
[13] Kasprowski, P., Harkeżlak, K., 2014. The second eye movements verification and identification competition, in: Proceedings of the International Joint Conference on Biometrics.
[14] Kasprowski, P., Komogortsev, O.V., Karpov, A., 2012. First eye movement verification and identification competition at BTAS 2012, in: BTAS 2012, pp. 195–202.
[15] Kasprowski, P., Ober, J., 2004. Eye movements in biometrics, in: International Workshop on Biometric Authentication, pp. 248–258.
[16] Komogortsev, O.V., Karpov, A., Holland, C.D., 2015. Attack of mechanical replicas: Liveness detection with eye movements. IEEE Transactions on Information Forensics and Security 10, 716–725.
[17] Landwehr, N., Arzt, S., Scheffer, T., Kliegl, R., 2014. A model of individual differences in gaze control during reading, in: EMNLP 2014, pp. 1810–1815.
[18] Makowski, S., Jäger, L.A., Abdelwahab, A., Landwehr, N., Scheffer, T., 2019. A discriminative model for identifying readers and assessing text comprehension from eye movements, in: ECML PKDD 2018, pp. 209–225.
[19] Makowski, S., Jäger, L.A., Prasse, P., Scheffer, T., 2020. Biometric identification and presentation-attack detection using micro-movements of the eyes, in: 2020 IEEE International Joint Conference on Biometrics (IJCB).
[20] Makowski, S., Jäger, L.A., Scheffer, T., . Judo1000 eye tracking data set. https://osf.io/5zpvk/. doi:10.17605/OSF.IO/5ZPVK.
[21] Martinez-Conde, S., Macknik, S.L., Hubel, D.H., 2004. The role of fixational eye movements in visual perception. Nature Reviews Neuroscience 5, 229–240.
[22] Martinez-Conde, S., Macknik, S.L., Troncoso, X.G., Dyar, T.A., 2006. Microsaccades counteract visual fading during fixation. Neuron 49, 297–305.
[23] Martinez-Conde, S., Macknik, S.L., Troncoso, X.G., Hubel, D.H., 2009. Microsaccades: A neurophysiological analysis. Trends in Neurosciences 32, 463–475.
[24] Otero-Millan, J., Troncoso, X.G., Macknik, S.L., Serrano-Pedraza, I., Martinez-Conde, S., 2008. Saccades and microsaccades during visual fixation, exploration, and search: Foundations for a common saccadic generator. Journal of Vision 8, 21–21.
[25] Rigas, I., Economou, G., Fotopoulos, S., 2012. Biometric identification based on the eye movements and graph matching techniques. Pattern Recogntion Letters 33, 786–792.
[26] Rigas, I., Komogortsev, O., Shadmehr, R., 2016. Biometric recognition via eye movements: Saccadic vigor and acceleration cues. ACM Transactions on Applied Perception 13, 6.
[27] Salvucci, D.D., Goldberg, J.H., 2000. Identifying fixations and saccades in eye-tracking protocols, in: Proceedings of the 2000 Symposium on Eye tracking Research & Applications (ETRA 2000), pp. 71–78.
[28] Silver, D.L., Biggs, A., 2006. Keystroke and eye-tracking biometrics for user identification, in: ICAI 2006, pp. 344–348.
[29] Yoon, H., Carmichael, T., Tourassi, G., 2015. Temporal stability of visual search-driven biometrics, in: SPIE Medical Imaging: Image Perception, Obserformance, and Technology Assessment.
[30] Yoon, H.J., Carmichael, T.R., Tourassi, G., 2014. Gaze as a biometric, in: SPIE Medical Imaging Conference: Image Perception, Observer Performance, and Technology Assessment.