

Iris Recognition on Mobile: Real-Time Feature Extraction and Matching in the Wild



Gleb Odinokikh and Alexey Fartukov

Abstract Methods of biometric recognition are becoming an essential part of various mobile applications. Their usability is determined by the accuracy and the speed of recognition in a highly variable environment. Complex textural features make the human iris one of the most reliable biometric traits. The changing environment and limited computational power of mobile devices give rise to a need for robust and fast feature extraction techniques. A method for iris feature extraction and matching is here proposed. It uses deep and element-wise representations of the discriminative features in combination with characteristics describing the environment. The model outperforms state-of-the-art methods in terms of both accuracy and speed. It has also been tested on a specially collected dataset that contains two-second videos simulating the natural enrollment and verification attempts of the user of the device. The dataset was collected considering the changes in environment and possible behavior of the user. The testing was performed in two scenarios: image-to-image and also video-to-video. A method for iris fusion (both eyes) is also proposed in this paper. Several such methods are studied and compared.

Keywords Iris recognition · Mobile biometrics · Feature extraction · Matching · Multi-instance fusion

1 Introduction

Mobile devices, such as smartphones and tablets, have become an integral part of many people's lives. Nowadays, the transfer and processing of personal information and various financial transactions are carried out using mobile devices. They are personal, which means the presence of an authentication procedure.

G. Odinokikh (✉) · A. Fartukov (✉)
Samsung R&D Institute Russia, Moscow, Russia
e-mail: g.odinokikh@gmail.com; a.fartukov@samsung.com

© Springer Nature Switzerland AG 2020
T. Bourlai et al. (eds.), *Securing Social Identity in Mobile Platforms*, Advanced Sciences and Technologies for Security Applications,
https://doi.org/10.1007/978-3-030-39489-9_11

The methods of biometric authentication have been actively promoted to replace conventional schemes that use keys, personal identification numbers, etc. The growth of interest in biometric technologies is associated mainly with the strengthening of the security requirements of the system and its usability. A lot of attention has been paid to mobile biometrics in recent years [1–3].

The present paper is focused on the human iris as the most reliable biometric modality. The goal of iris recognition is to recognize a human's identity through the textural characteristics of the muscular patterns of the iris. A typical iris recognition system consists of the following stages: iris image acquisition, iris image segmentation, feature extraction, and pattern matching [4]. Iris image acquisition is usually performed using a high resolution camera which is either near-infrared (NIR) or visible-spectrum (VIS), under controlled environmental conditions [5]. The minimal requirements for iris image capturing are summarized in ISO/IEC 19794-6:2011 [6].

Since the mobile market is global, all the possible behavioral- and race-dependent features of the final users must be taken into account. For this reason, in particular, only the NIR spectrum is considered in this paper. The advantages of using NIR images have been well explained in the literature [6–9]. It should be noted that the development and implementation of an iris capturing camera for mobile devices is outside the scope of the present paper. Most of the issues related to the iris capturing device are well summarized in [5, 7].

In the case of a mobile device, it is not always possible to satisfy all the mentioned requirements imposed on the camera. There are two main reasons for this: the camera module should be small enough and not expensive to manufacture. The costs of production play a significant role in the case of a mass market. Another challenge is that capturing the iris image is performed under uncontrolled environmental conditions. These factors greatly affect the quality of the iris image.

This paper describes a method of iris feature extraction and matching that is capable of working in real-time on a mobile device equipped with an NIR camera.

The rest of this paper is organized as follows: the key issues of iris recognition on a mobile device are explained in Sect. 2; Sect. 3 surveys related work; the proposed approach is presented in Sect. 4; and experimental results and conclusions are presented in Sects. 5 and 6 respectively.

2 Problem Statement

A mobile biometric sensor should be able to handle the data under constantly changing environmental conditions and consider user inherent features. In biometric systems that use an image as input data, the factors of the environment are becoming more important. One of them is the ambient illumination, which varies over a range from 10^{-4} at night to 10^5 Lux under direct sunlight. Another is the randomness of the locations of the light sources, along with their unique characteristics, which creates a random distribution of the illuminance in the iris area. These factors lead to a deformation of the iris structure caused by a change in the pupil size,

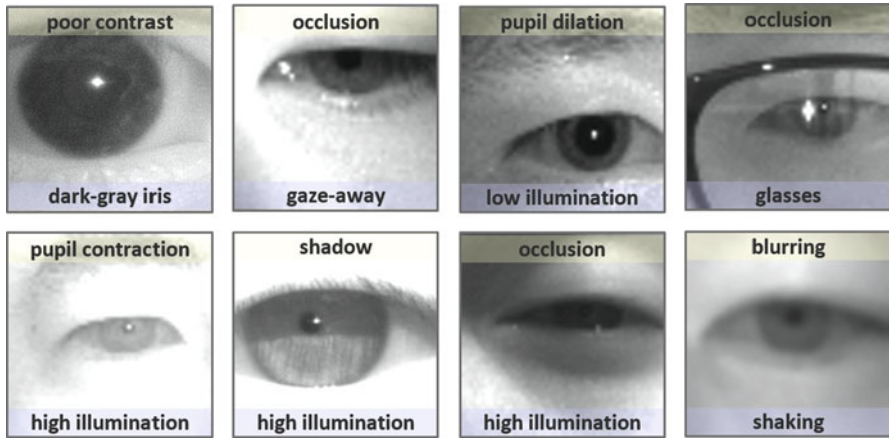


Fig. 1 Examples of iris images captured with a mobile device

making users squint, and degrading the overall image quality. Several examples of iris images are given in Fig. 1. The influence of the environment is well described in the literature [10–13]. Other factors inherent to the user also affect the quality of the output, such as the use of glasses or contact lenses, the existence of a hand tremor, or the mere act of walking, thereby introducing a shaking of the device, the variation in distance to the iris causing the iris to move out of the camera’s depth of field, and occlusion of the iris area by eyelids and eyelashes if the user’s eye is not opened enough [10]. All these and many other factors affect the quality of the input biometric data thus influencing the accuracy of the recognition [14, 15].

Mobile applications should be simple and convenient in use. In the case of a biometric system on a mobile device, being convenient means providing an easy user interaction and a high recognition speed, which is determined by the computational complexity. Mobile secure systems that process any personal data are even more limited in computational resources. Not many researchers have attached importance to this. These systems typically represent a system-on-a-chip (SoC) completely abstracted from external resources, keeping all the processing inside itself. Such systems were initially developed to carry out simple operations with simple data (PINs, passwords etc.) and did not require huge resources. Neither were they ready for the complex processing of biometric information, but they have continued to be improved. The latter system’s restrictions usually meant even more reduced CPU frequency and limited RAM.

All these problems greatly complicate iris feature extraction, making most of the existing methods unreliable, and promising techniques such as deep neural networks (DNN) inoperable.

There are several commercial mobile iris recognition solutions known to date. The first smartphones enabled with the technology were introduced by Fujitsu [16] and Microsoft [17] in 2015. All Samsung flagship devices were equipped with iris

recognition technology during 2016–2018 [18]. Some B2B and B2G applications of the technology are also known in the market, such as Samsung Tab Iris [19] and IrisGuard EyePay Phone [20]. The scope of the applications of this technology is growing and has brought about a demand for its further improvement.

The present paper is focused on the feature extraction and matching parts of the iris recognition pipeline. Feature extraction, in this case, means a numeric representation of the unique iris features extracted from the preliminarily determined iris area of the image. Matching means calculating a measure of dissimilarity between the two extracted feature vectors.

3 Related Work

Recent achievements in the field of deep learning have allowed a significant leap in the reliability and quality of the research in the field of biometrics and, in particular, in iris recognition. One of the first attempts to explore the capabilities of DNNs was a feasibility analysis of DNN embeddings trained on ImageNet for classification, with the PCA+SVM applied over the VGG embeddings [21] by Minae et al. Furthermore, Gangwar et al. [22] introduced their DeepIrisNet as a model combining all successful deep learning techniques known at the time. They thoroughly investigated the obtained features and produced a strong baseline as a robust foundation for future research. A year later, similar work based on these embeddings was introduced by Tang et al. [23]. At the same time, Proenca et al. [24] presented IRINA. The idea was to use a DNN to find corresponding patches from the examined images, use MRF to perform precise deformable registration, and a SVM to classify genuine and impostor data. They achieved unprecedented robustness to pupil/iris variations and segmentation errors, but the accuracy of the solution was traded off against performance. The proposed design significantly limited the applicability of the method for mobile applications. Another approach with two fully-convolutional networks with a modified triplet loss function has been proposed recently [25]. One of the networks is used for iris template extraction whereas the second produces the accompanying mask. Fuzzy image enhancement combined with simple linear iterative clustering and an SOM neural network were proposed in [26]. Although this method was designed for iris recognition on a mobile device, real-time performance was not achieved. Another recent paper [13] meant to be suitable for the case of a mobile device proposed a two-headed (iris and periocular) CNN with a fusion of the embeddings. Thus, there is no fully optimal solution for iris feature extraction and matching in the published papers.

4 Iris Feature Extraction and Matching

The proposed method represents a CNN designed to use the advantages of the normalized iris image as an invariant, both low and high level discriminative feature representations, and information about the environment. It contains iris feature extraction and matching parts trained together.

4.1 Recognition Pipeline

A common iris recognition pipeline consists of several stages separated by intermediate quality checks. The feature extraction part is preceded by the segmentation stage and followed by the matching. All the input data for the feature extraction (normalized iris and mask images) were obtained automatically by an algorithm developed in our lab. The basic structure of the algorithm was taken from [10] with two modifications: (i) the scheme that contains a special quality buffer was replaced with a straightforward structure as depicted in Fig. 2; (ii) the feature extraction and matching parts were also replaced with the new ones. All the other parts of the algorithm and quality checks were used with no modifications.

4.2 Low-Level Feature Representation

It is known from the previous literature [27, 28] that shallow layers in CNNs are responsible for the extraction of low-level textural information while high-level

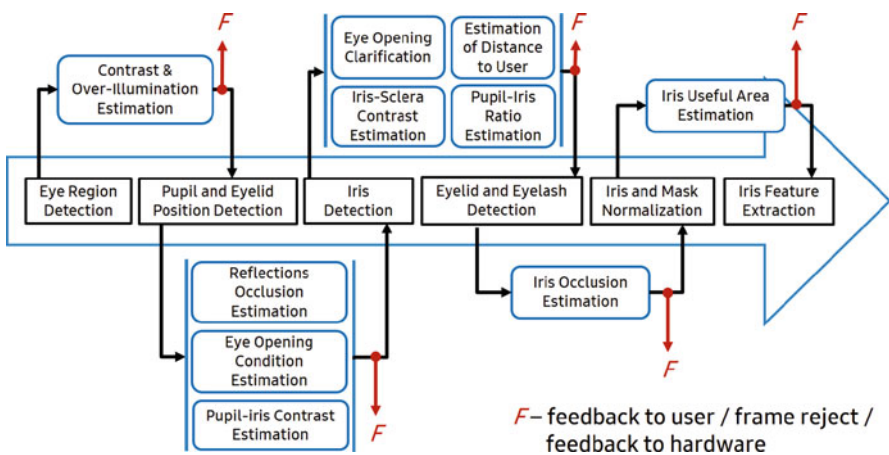


Fig. 2 Quality assessment scheme

representation is achieved with depth. Methods of iris feature extraction based on local texture characteristics, which are calculated by spatially and spectrally local transformations [9, 29] are basically attempts to use low-level description of the texture. These methods have proven their reliability for scenarios with an almost unchanging environment, but are highly sensitive to environmental variations.

A normalized image of the iris allows textural element-wise features to be useful in the case of narrow changes of environmental conditions. They remain well aligned with each other in such a case. For this reason, iris recognition is a good example of a task for which the profitability of using low-level feature representations could be investigated in the context of CNN-based methods and a wide range of environmental changes.

The influence of shallow features in the context of CNNs on recognition performance is studied in this paper. A classic approach [9] using a Hamming-distance based dissimilarity score has been taken as the basis. The vector FV_{sh} of elements x_i is used as a representation of low-level discriminative features:

$$x_i = \frac{\sum |FM_{1,i}^{Sq} - FM_{2,i}^{Sq}| \times M_c}{\sum M_c}, \quad (1)$$

where $FM_{k,i}^{Sq}$ is the i th feature map of the k th iris after normalization to zero mean and unit variance, binarized by sign; M_c is a binary mask representing noise; and M_c is a combination of M_1 and M_2 .

The shallow feature extraction block is depicted in Fig. 3 and the structure of the convolution block #1 is presented in Table 1. Depth-wise separable convolution block structures, first proposed in [30] as memory and computationally efficient, were picked as the basic structural elements for the entire network. Feature maps $FM_{1,i}^{Sq}$ and $FM_{2,i}^{Sq}$ in (1) are obtained after the first convolution layer (Table 1).

After 100 epochs of training, the distributions of the elements of FV_{sh} for genuine and impostor comparisons from the validation set appear as in Fig. 4. Although the filters vary considerably, the distributions look very similar. The shape of the distributions for both classes resemble a Gaussian, therefore d' and EER values were chosen for the evaluation of their separation degree. How the values for each filter were changed during training is presented in Fig. 5. The results presented in Table 3 show that the model using FV_{sh} as an additive factor obtains slightly better results for the baseline model with 3×3 kernels on the first layer. It is also shown that for the larger kernels, the difference in performance becomes more significant (Table 3).

4.3 Deep Feature Representation

Deep (high-level) feature representation is obtained with convolution block #2. The feature maps $FM_{1,i}^{Sq}$ and $FM_{2,i}^{Sq}$ are concatenated after block #1 by channels

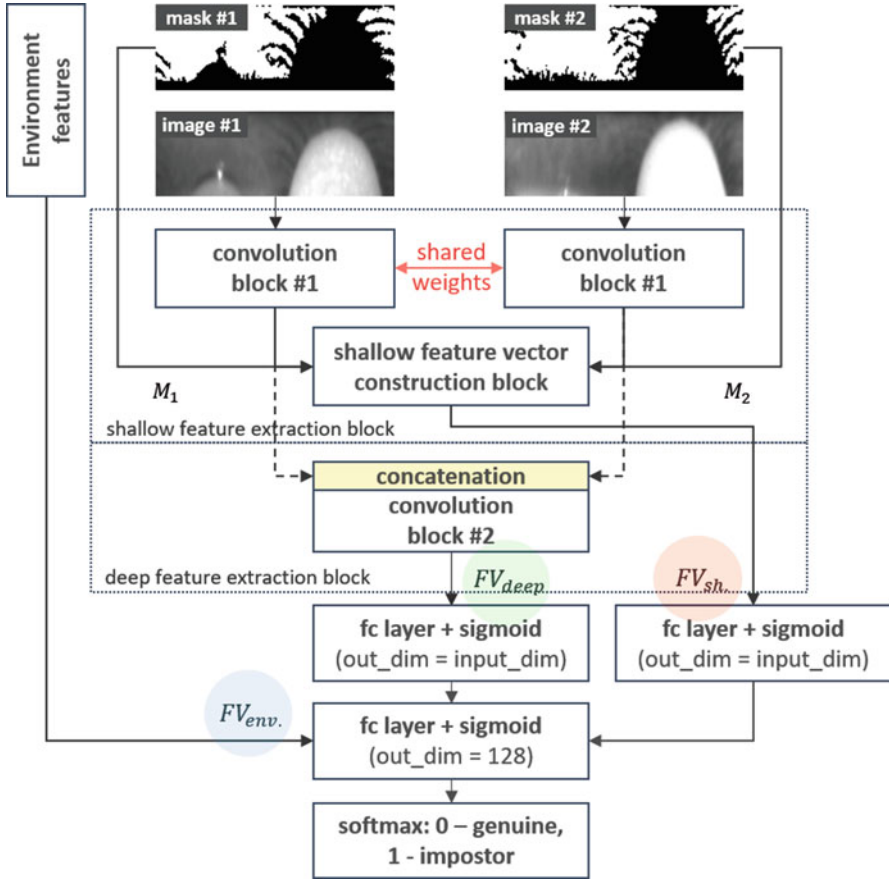


Fig. 3 Proposed model scheme

Table 1 Structure of the convolution blocks

| Layer | Input shape | Convolution block |
|---|--------------------------|-------------------|
| Conv. 3x3 ($s' = 1, act. = tanh$) | $1 \times 49 \times 161$ | #1 |
| Depthwise Sep. Conv. Block ($s' = 2$) | $8 \times 47 \times 159$ | #2 |
| Depthwise Sep. Conv. Block ($s' = 2$) | $32 \times 23 \times 79$ | |
| Depthwise Sep. Conv. Block ($s' = 2$) | $32 \times 11 \times 39$ | |
| Depthwise Sep. Conv. Block ($s' = 1$) | $32 \times 5 \times 19$ | |
| FC layer + BatchNorm + ReLU | 1×1632 | |

and passed through it (Fig. 3). The meaning of the concatenation at this stage is in the invariance property of the normalized iris image. Experiments showed the advantages of this approach in comparison with standard techniques [31] where the feature vectors had highly decreased dimensionality. However, the large sizes of the

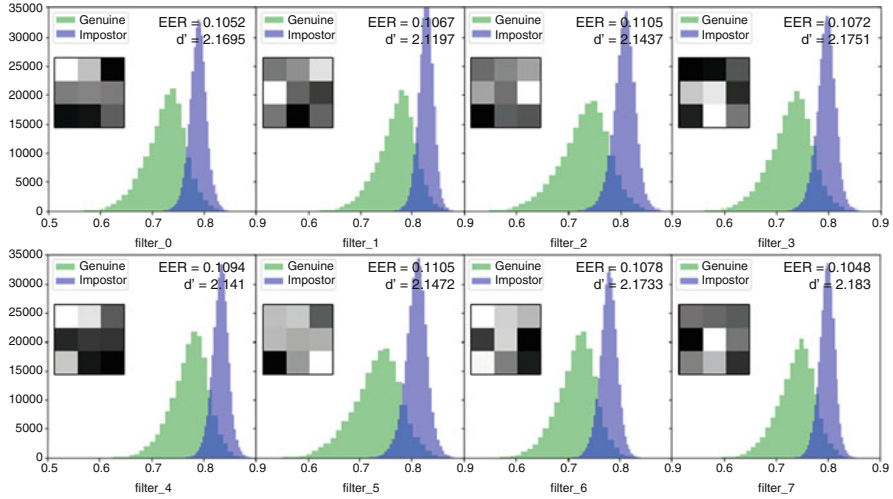


Fig. 4 Distributions of elements of FV_{sh} after 100 epochs

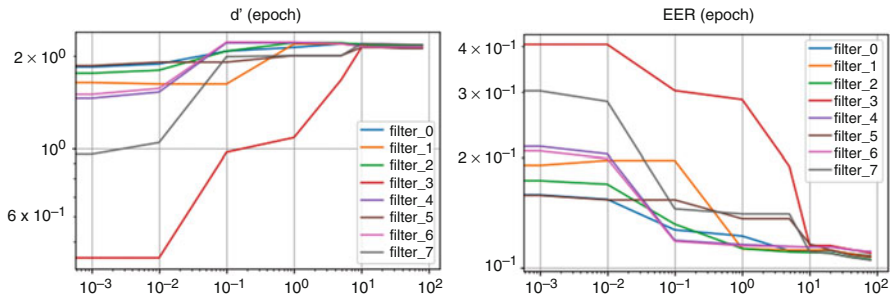


Fig. 5 The dependency of d' (left) and EER (right) for each filter on the number of epochs

vectors and the complexity of the matching procedure are among the drawbacks of this approach. The structure of the block is presented in Table 1. The output vector $FV_{deep} \in R^{128}$ reflects a high-level representation of the discriminative features and is assumed to handle complex non-linear distortions of the iris texture.

4.4 Matching Score Calculation

The analysis of outliers along with the nature of the distributions of the elements of FV_{sh} gave rise to the idea of using a variational inference technique for regularization. What this means is that some vectors are being represented as n -dimensional random variables with a certain shape distribution. In the present paper, the representations of both FV_{sh} and FV_{deep} vectors are described as having multi-

variate normal distributions $FV'_{sh} \sim N(\mu_{sh}, \Sigma_{sh})$ and $FV'_{deep} \sim N(\mu_{deep}, \Sigma_{deep})$ respectively, where μ is the vector of mean values and Σ is the covariance matrix. Variational inference is performed with the so called re-parametrization trick described in [32]. Sampling from the distributions is performed only for training, while only the values of μ are used for inference. A sigmoid activation function is then applied to the result. The same procedure is further performed for the concatenated vectors FV'_{sh} , FV'_{deep} and FV_{env} . Here, FV_{env} reflects environment conditions and contains information about iris area and pupil dilation: $FV_{env} = \{\Delta NPR, AOI\}$, and the area of intersection $AOI = \Sigma M_c / M_c^h \times M_c^w$ with ΔNPR given by

$$\Delta NPR = \left| \frac{R_1^p}{R_1^i} - \frac{R_2^p}{R_2^i} \right| \quad (2)$$

where R^p and R^i are the radii of the pupil and the iris, respectively. The output vector $FV'_d \in R^{128}$ is an input for the last fully-connected layer with two nodes describing the classes. A *SoftMax* classifier is applied to the values from the nodes for probability (matching score) estimation.

According to the obtained results (Table 3), the application of variational inference (VI) improved the recognition performance (VI=No means the replacement of the VI structure with simple fully-connected layers of the same dimensionality and activations), but it is also worth mentioning that it becomes less reasonable with an increasing amount of training data.

4.5 Weighted Loss

A specially designed loss function is another proposed feature. Sometimes two images of the same iris are very different from each other. This can happen for various reasons: the different parts of the iris can be occluded by some noise, one of the images can be badly distorted due to segmentation errors, etc. Thus, it is almost impossible to attribute them to the same class and for this reason a certain part of all genuine comparisons in the training data obstruct the convergence of the model. So, it is reasonable to consider or even completely ignore these comparisons when training. The following algorithm is proposed: (i) calculate the loss function (e.g., cross-entropy) for each comparison in the batch; (ii) apply $weights = \{w_0..w_K\}$ to the top k highest values among the genuine matches; (iii) output the overall sum. In this paper, the values were set to: $weights = 0$ and $k = 10\%$. This approach provided better convergence and achieved a better recognition performance.

4.6 Multi-instance Iris Fusion

The input images may contain both eyes, as depicted in Fig. 6. In this case both irises can be used at the same time [33], which is the obvious way to increase the reliability and convenience of the recognition. It has been observed that at least 40% of the iris area should be visible to achieve the given accuracy level. In other words, the user should open the eyes wider during one-eye recognition, which is not always convenient. Often the iris is significantly occluded by the eyelids, eyelashes, highlights, etc. This happens mainly because of the complex environment, in which the user cannot open the eyes wide enough (bright illumination, windy weather, etc.). It makes the application of the iris multi-instance approach reasonable.

An ideal scenario for matching is when both compared irises are well aligned to each other spatially and the conditions of the capturing are the same in both cases [15, 34]. This is impossible to satisfy in practice, especially in the mobile case. But it is reasonable to use information about the initial relative position of the compared irises before the normalization. A method that performs the fusion of the two irises and uses the relative spatial information and several factors that describe the environment is also considered as an important path of the presented research.

The final dissimilarity score is calculated as a logistic function of the form

$$score = \frac{1}{1 + e^{-\sum w_i \cdot M_i}} \quad (3)$$

where $M \in R^7$ is the set of the following measures:

$$M = \{ \Delta d_0, d_{avg}, AOI_{min}, AOI_{max}, \Delta ND_{min}, \Delta ND_{max}, \Delta PIR_{avg} \} \quad (4)$$



Fig. 6 Examples of the images captured with the mobile device equipped with an NIR camera

where

Δd_0 is the normalized score difference for two pairs of irises,

$$\Delta d_0 = \frac{|d_0^{left} - d_0^{right}|}{d_0^{left} + d_0^{right}} \quad (5)$$

d_{avg} is the average score for the pair,

$$d_{avg} = 0.5 \cdot (d_0^{left} + d_0^{right}) \quad (6)$$

AOI_{min} , AOI_{max} are the minimum and maximum values of the area of intersection between the two binary noise masks M_{prb} and M_{enr} in each pair,

$$AOI = \Sigma M_c / (M_c^h \times M_c^w), \quad M_c = M_{prb} \times M_{enr} \quad (7)$$

ΔND_{min} , ΔND_{max} are the minimum and maximum values of the normalized distance ΔND between the centers of the pupil and the iris,

$$\Delta ND = \sqrt{(NDX_{prb} - NDX_{enr})^2 + (NDY_{prb} - NDY_{enr})^2} \quad (8)$$

$$NDX = \frac{x_P - x_I}{R_I}, \quad NDY = \frac{y_P - y_I}{R_I}, \quad (9)$$

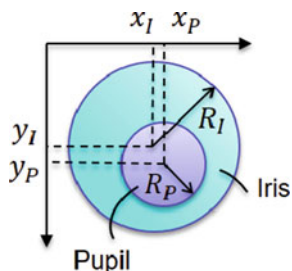
where x_P and y_P are the coordinates of the center of the pupil and R_P is its radius, while x_I and y_I are the coordinates of the center of the iris and R_I is its radius, as depicted in Fig. 7.

The measure ΔPIR_{avg} reflects the difference in pupil dilation during the enrollment and probe using the value of $PIR = R_P/R_I$:

$$\Delta PIR_{avg} = 0.5 \cdot \left(\left| PIR_{enr}^{left} - PIR_{prb}^{left} \right| + \left| PIR_{enr}^{right} - PIR_{prb}^{right} \right| \right) \quad (10)$$

where R_P and R_I are the radii of the pupil and the iris, respectively.

Fig. 7 Parameters of the pupil and iris used for the iris fusion



The weight coefficients for the logistic function were obtained after the training of the classifier on genuine and impostor matches on a small subset of the data. In case only one out of two feature vectors is extracted, all the pairs of values used in the weighted sum are assumed to be equal.

The proposed method helped to increase the recognition accuracy. It is also allowed to decrease the threshold for the visible iris area from 40% to 29% during verification/identification without any loss in the accuracy and performance, which means a decreased overall FRR as a result.

A comparison of the proposed method with well-known consensus and minimum rules was carried out. According to the consensus rule, a matching is considered as successful if both d_0^{left} and d_0^{right} are less than the decision threshold. In the minimum rule, what is required is that the minimum of the two values $\min(d_0^{left}, d_0^{right})$ should be less than the threshold. The testing results are presented in Table 7.

5 Experimental Results

The main objectives of the biometric system performance evaluation include assessing the progress in improving the accuracy during the development of the algorithms and providing an objective reflection of the performance when the system is in operation [35]. To meet these goals, two types of evaluation were conducted: (i) an image-to-image evaluation of the proposed feature extraction and matching method with state-of-the-art methods on several datasets, including publicly available ones; (ii) a video-to-video evaluation to simulate real-world usage of the whole iris recognition solution and test the proposed multi-instance iris fusion approach.

5.1 Image-to-Image Evaluation

Three different datasets were used for the comparison. The following methods were selected as state-of-the-art: (1) FCN+ETL proposed by Zhao and Kumar in [25], which is one of the most cutting edge solutions, with the highest recognition performance; (2) DeepIrisNet [22], representing a classic deep neural net approach as one of the earliest applications of deep learning in the field of iris recognition. A lightweight CNN recently proposed in [13] could also be used for comparison since the results were obtained on the same dataset. Refer to the original paper [13] for the results on the CASIA-Iris-M1-S3 dataset [36].

Many methods were excluded from consideration due to their computational complexity and therefore unsuitability for mobile applications.

Table 2 Datasets details

| Dataset | Images | Irises | Outdoor | Subjects |
|---------|--------|--------|---------|-------------------|
| CMS2 | 7723 | 398 | 0 | Asian |
| CMS3 | 8167 | 720 | 0 | Asian |
| IM | 22966 | 750 | 4933 | Asian & Caucasian |

5.1.1 Dataset Description

The following datasets were used for training and evaluation: CASIA-Iris-M1-S2 (CMS2) [36], CASIA-Iris-M1-S3 (CMS3) [36], and Iris-Mobile (IM). The collection of the last one was performed privately using a mobile device with an embedded NIR camera to simulate real authentication scenarios of the user of a mobile device. The images were captured under a wide range of changes in illumination, both indoors and outdoors, with and without glasses (Table 2). Examples of images are presented in Fig. 1. Images from all the datasets were marked automatically by an algorithm developed in our lab. Examples of iris and mask images are presented in Fig. 3. Each dataset was initially divided into training, validation, and testing subsets in proportions of 70/10/20 (%) respectively. This was so that there would be no images of the same iris in two different subsets.

5.1.2 Training

The number of genuine comparisons N_G was much smaller than the number of impostor comparisons. Therefore all genuine comparisons were used for training and the number of impostor comparisons was fixed as $N_I = 10N_G$. The model that showed the lowest EER on the validation set was selected for evaluation on the testing dataset. All the models were trained for 150 epochs using the Adam optimizer. The training of the proposed model was performed so that one epoch was equivalent to one iteration over all the genuine comparisons whereas the impostors are always randomly selected from the entire set for each batch. The proportion of genuine and impostor comparisons in a batch was set to $N_I^b = 10N_G^b$ and $AOI \geq 0.2$ for all the image pairs.

5.1.3 Performance Evaluation

The recognition accuracy results are presented in Table 4 and Fig. 8. The proposed feature extraction and matching method outperforms the chosen state-of-the-art ones on all the datasets. Since the number of comparisons for the CMS2 and CMS3 testing sets did not exceed 10 million after the division into subsets, it was impossible to estimate the FNMR at $FMR = 10^{-7}$. Another experiment was used to estimate the performance of the proposed model on those datasets without training on them. The model trained on IM was evaluated on the entire CMS2 and CMS3

Table 3 Recognition performance results for several model modifications on IM dataset

| Conv1 | VI | FV_{sh} | EER | FNMR | d' |
|-----------------------|-----|-----------|--------|--------|--------|
| $8 \times 3 \times 3$ | Yes | Yes | 0.0116 | 0.1925 | 4.3155 |
| $8 \times 3 \times 3$ | No | Yes | 0.0120 | 0.2027 | 4.2048 |
| $8 \times 3 \times 3$ | Yes | No | 0.0125 | 0.2085 | 4.1253 |
| $8 \times 9 \times 9$ | Yes | Yes | 0.0134 | 0.1566 | 4.3034 |
| $8 \times 9 \times 9$ | Yes | No | 0.0172 | 0.1694 | 3.9850 |

Table 4 Recognition performance evaluation results

| Method | EER | | | Testing | FPS |
|------------------|--------|--------|--------|----------|-----|
| | CMS2 | CMS3 | IM | | |
| DeepIrisNet [22] | 0.0709 | 0.1199 | 0.1371 | WithinDB | 11 |
| FCN+ETL [25] | 0.0093 | 0.0301 | 0.0607 | WithinDB | 12 |
| Proposed | 0.0014 | 0.0190 | 0.0116 | WithinDB | 250 |
| | 0.0003 | 0.0086 | – | CrossDB | |

datasets in order to obtain FNMR at $FMR = 10^{-7}$ (CrossDB). The results presented in Table 4 and Fig. 8 demonstrate the high generalization ability of the model. However, it is fair to note that the IM dataset contains much more data than the other two.

A mobile device equipped with Qualcomm Snapdragon 835 CPU was used for estimating the overall execution time for these iris feature extraction and matching methods. It should be noted that a single core of CPU was used. The results are summarized in Table 4.

5.2 Video-to-Video Evaluation

In fact, both the registration and verification procedures involve the processing of not one, but a sequence of images. The video format gives more information about the possible behavior of the user and the environment. Unfortunately, there are no such publicly available iris datasets. So, in order to test the recognition performance on data that would be close to real-world scenarios, an additional dataset was collected privately. It is a set of two-second video sequences, each of which is a real enrollment/verification attempt.

5.2.1 Dataset Description

The dataset was collected using a mobile device with an embedded NIR camera. It contains videos captured in different environment: (i) indoors (IN) and outdoors (OT); (ii) with and without glasses; (iii) at different distances. The conditions of illumination during the capturing were set as: (i) three levels for the indoor samples

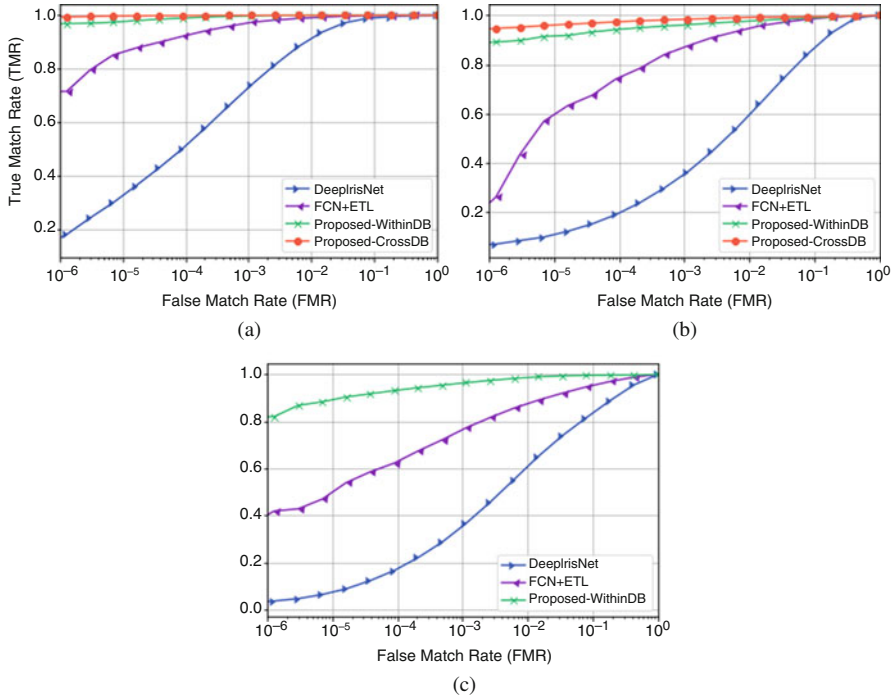


Fig. 8 ROC curves obtained for comparison with state-of-the-art methods on different datasets: (a) CASIA-Iris-M1-S2 (CMS2), (b) CASIA-Iris-M1-S3 (CMS3) and (c) Iris Mobile (IM)

Table 5 Dataset specification

| Dataset | Non-glasses | Glasses |
|--------------------|-------------------|------------|
| Users in dataset | 476 | 224 |
| Max comparisons | 22 075 902 | 10 605 643 |
| Race | Asian & Caucasian | |
| Eyes on video | Two | |
| Videos per user | 10 ± 2 | |
| Video length | 30 frames | |
| Capturing distance | 25–40 cm | |
| Camera resolution | 1920 × 1920 | |

(0–30, 30–300 and 300–1000 Lux); (ii) a random value in the range 1–100 K Lux (data was collected on a sunny day); Different arrangements of the device relative to the sun were also considered during the capturing. A detailed description of the dataset is presented in Table 5. The Iris Mobile (IM) dataset used for the image-to-image evaluation was randomly sampled from, as well. Examples of the pictures from the videos are depicted in Fig. 6.

5.2.2 Testing Procedure

All the video sequences were used for simulating both the enrollment and verification transactions (attempts) in the non-glasses (NG) case. The sequences captured for users wearing glasses (G) were used for simulation of the verification attempts only. Each video sequence is considered as a single attempt. The extracted probe/enrollment template is the result of a successful attempt and may contain a maximum of 60 (30 frames \times 2 eyes) iris feature vectors. The successful construction of the feature vector means passing all the intermediate quality checks in the recognition pipeline.

The testing procedure consists of the following steps:

1. Passing of all the videos that satisfy the condition IN&NG through the feature extraction to produce the enrollment template: the template is considered as successfully created if the following requirements are met:
 - a. At least 5 feature vectors were constructed for each eye;
 - b. At least 20 out of 30 frames were processed.
2. Passing of all the videos through the feature extraction to produce the probe template, which is considered as successfully created in the case of at least 1 feature vector being constructed;
3. Creating of a pair-wise matching table of the dissimilarity scores for all the comparisons: each probe template is compared with all enrollment templates except the ones generated from the same video;
4. Calculating of the measures: FTE, FTA, FNMR(FMR) and FRR(FAR).

One important thing that makes the enrollment and verification different are the values of the following thresholds: (i) the normalized eye opening (NEO) value, described in [10], was set as 0.5 for the enrollment and 0.2 for the verification; (ii) the non-masked area of the iris (not occluded by any noise) was set as 0.4 and 0.29 for the enrollment and probe, respectively.

5.2.3 Performance Evaluation

The recognition accuracy results are presented in Table 6. The proposed feature extraction and matching method is compared with the one proposed in [10] as a part of the whole pipeline. The compared method is based on Gabor wavelets with an adaptive phase quantization technique (Gabor+AQ). Both methods were tested in three different verification environments: indoors without glasses (IN&NG), indoors with glasses (IN&G), and outdoors without glasses (OT&NG). The enrollment was always carried out only indoors without glasses and, for this reason, the value of $FTE=3.15$ is the same for all the cases. The target $FMR=10^{-7}$ was set in every experiment.

Applying different matching rules was also investigated. The proposed multi-instance fusion showed advantages over the other compared rules (Table 7).

Table 6 Recognition accuracy in different verification conditions

| Error rate, % | Method | Verification condition | | |
|---------------|-----------------|------------------------|------|-------|
| | | IN&NG | IN&G | OT&NG |
| EER | Proposed | 0.01 | 0.09 | 0.42 |
| | Gabor + AQ [10] | 0.10 | 0.35 | 3.15 |
| FNMR | Proposed | 0.48 | 5.52 | 10.1 |
| | Gabor+AQ [10] | 1.07 | 8.94 | 32.5 |
| FTA | – | 0.21 | 4.52 | 0.59 |

Table 7 Recognition accuracy for different matching rules

| Error rate, % | Method | Fusion | Minimum | Consensus |
|---------------|---------------|--------|---------|-----------|
| EER | Proposed | 0.01 | 0.21 | 0.21 |
| | Gabor+AQ [10] | 0.10 | 1.31 | 1.31 |
| FNMR | Proposed | 0.48 | 0.92 | 1.25 |
| | Gabor+AQ [10] | 1.07 | 3.17 | 4.20 |

The overall execution time for the whole pipeline was measured on a single core of Qualcomm Snapdragon 835 CPU and was 55 milliseconds, which is about 18 FPS real-time performance.

6 Conclusion

A novel approach to iris feature extraction and matching was proposed in this paper. It showed robustness to the high variability of the iris representation caused by changes in the environment and physiological features of the iris itself. The profitability of using shallow textural features, feature fusion, and variational inference as a regularization technique, was also investigated in the context of the iris recognition task. One more feature of the proposed solution is its multi-instance iris fusion, which helps to increase the performance in case the input image contains both eyes at the same time. The proposed solution was tested in the video-to-video scenario and showed its ability to work in real-time in an uncontrolled environment. Although it showed high accuracy indoors, the outdoor recognition is still challenging.

Acknowledgements This research was also supported in part by a research grant 19-07-01231 of Russian Foundation of Basic Research.

References

1. Meng W, Wong D, Furnell S, Zhou J (2015) Surveying the development of biometric user authentication on mobile phones. *IEEE Commun Surv Tutor* 17(3):1268–12937
2. Rui Z, Yan Z (2018) Survey on biometric authentication: toward secure and privacy-preserving identification. *IEEE Access* 7:5994–6009
3. Das A, Galdi C, Han H, Ramachandra R, Dugelay J, Dantcheva A (2018) Recent advances in biometric technology for mobile devices. In: 2018 IEEE 9th international conference on biometrics theory, applications and systems (BTAS), pp 1–11. <https://doi.org/10.1109/BTAS.2018.8698587>
4. Li YH, Savvides M (2009) Iris recognition, overview. *Encyclopedia of biometrics*. Springer, New York, pp 810–819
5. Prabhakar S, Ivanisov A, Jain AK (2011) Biometric recognition: sensor characteristics and image quality. *IEEE Instrum Meas Soc Mag* 14(3):10–16
6. Information technology (2011) biometric data interchange formats – part 6: iris image data, annex b. ISO/IEC 19794-6:2011
7. Corcoran P, Bigioi P, Thavalengal S (2014) Feasibility and design considerations for an iris acquisition system for smartphones. In: 2014 IEEE fourth international conference on consumer electronics Berlin (ICCE-Berlin), pp 164–167. <https://doi.org/10.1109/ICCE-Berlin.2014.7034328>
8. Bowyer KW, Hollingsworth K, Flynn PJ (2008) Image understanding for iris biometrics: a survey. *Comput Vis Image Underst* 110(2):281–307. <https://doi.org/10.1016/j.cviu.2007.08.005>
9. Daugman J (2004) How iris recognition works. *IEEE Trans Circuits Syst Video Technol* 14(1):21–30. <https://doi.org/10.1109/TCSVT.2003.818350>
10. Odinokikh GA, Fartukov AM, Ereemeev VA, Gnatyuk VS, Korobkin MV, Rychagov MN (2018) High-performance iris recognition for mobile platforms. *Pattern Recognit Image Anal* 28(3):516–524. <https://doi.org/10.1134/S105466181803015X>
11. Thavalengal S, Corcoran P (2016) User authentication on smartphones: focusing on iris biometrics. *IEEE Consum Electron Mag* 5(2):87–93. <https://doi.org/10.1109/MCE.2016.2522018>
12. Zhang M, Zhang Q, Sun Z, Zhou S, Ahmed NU (2016) The BTAS competition on mobile iris recognition. In: 2016 IEEE 8th international conference on biometrics theory, applications and systems (BTAS), pp 1–7. <https://doi.org/10.1109/BTAS.2016.7791191>
13. Zhang Q, Li H, Sun Z, Tan T (2018) Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Trans Inf Forensics Secur* 13(11):2897–2912. <https://doi.org/10.1109/TIFS.2018.2833033>
14. Tabassi E (2011) Large scale iris image quality evaluation. In: Proceedings of international conference of the biometrics special interest group (BIOSIG), pp 173–184
15. Matveev I, Novik V, Litvinchev I (2018) Influence of degrading factors on the optimal spatial and spectral features of biometric templates. *J Comput Sci* 25:419–424
16. Fujitsu develops prototype smartphone with iris authentication. Press Release (2015). <https://www.fujitsu.com/global/about/resources/news/press-releases/2015/0302-03.html>
17. Callahan J (2015) Microsoft Lumia 950 and Lumia 950 XL smartphones officially announced. Windows Central. Mobile Nations. <https://www.windowscentral.com/microsoft-lumia-950-and-lumia-950-xl-smartphones-officially-announced>
18. How does the iris scanner work on Galaxy S9, Galaxy S9+, and Galaxy Note9? (2019) <https://www.samsung.com/global/galaxy/what-is/iris-scanning/>
19. Galaxy tab iris (sm-t116izkrins) specification (2019) <https://www.samsung.com/in/business/tablets/galaxy-tab-iris-7-0-t116ir/sm-t116izkrins/>
20. Irisguard EyePay Phone (IG-EP100) specification (2019) <https://www.irisguard.com/node/57>
21. Minaee S, Abdolrashidi A, Wang Y (2016) An experimental study of deep convolutional features for iris recognition. In: 2016 IEEE signal processing in medicine and biology symposium (SPMB), pp 1–6

22. Gangwar AK, Joshi A (2016) DeepIrisNet: deep iris representation with applications in iris recognition and cross-sensor iris recognition. In: ICIP 2016, Phoenix, 25–28 Sept 2016, pp 2301–2305. <https://doi.org/10.1109/ICIP.2016.7532769>
23. Tang X, Xie J, Li P (2017) Deep convolutional features for iris recognition. In: Chinese conference on biometric recognition, pp 391–400. Springer
24. Proença H, Neves JC (2017) IRINA: iris recognition (even) in inaccurately segmented data. In: 2017 IEEE conference on computer vision and pattern recognition, CVPR 2017, Honolulu, 21–26 July 2017, pp 6747–6756. <https://doi.org/10.1109/CVPR.2017.714>
25. Zhao Z, Kumar A (2017) Towards more accurate iris recognition using deeply learned spatially corresponding features. In: IEEE international conference on computer vision, ICCV 2017, Venice, 22–29 Oct 2017, pp 3829–3838. <https://doi.org/10.1109/ICCV.2017.411>
26. Abate AF, Barra S, D’Aniello F, Narducci F (2017) Two-tier image features clustering for iris recognition on mobile. In: Petrosino A, Loia V, Pedrycz W (eds) Fuzzy Logic and Soft Computing Applications. Lecture Notes in Artificial Intelligence, vol 10147, pp 260–269
27. Zeiler MD, Fergus R (2014) Visualizing and understanding convolutional networks. In: Proceedings of ECCV
28. Harley AW (2015) An interactive node-link visualization of convolutional neural networks. In: Proceedings of ISVC
29. Pavelyeva E, Krylov A (2010) An adaptive algorithm of iris image key points detection. In: 20th international conference on computer graphics and vision, GraphiCon’2010, pp 320–323
30. Howard A, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Hartwig A (2017) Mobilenets: efficient convolutional neural networks for mobile vision applications. CoRR abs/1704.04861 . <http://arxiv.org/abs/1704.04861>
31. Koch G, Zemel R, Salakhutdinov R (2015) Siamese neural networks for one-shot image recognition. In: Proceedings of the 32th international conference on machine learning
32. Kingma DP, Salimans T, Welling M (2015) Variational dropout and the local reparameterization trick. In: Cortes C, Lawrence ND, Lee DD, Sugiyama M, Garnett R (eds) Advances in neural information processing systems, vol 28, pp 2575–2583. Curran Associates, Inc. <http://papers.nips.cc/paper/5666-variational-dropout-and-the-local-reparameterization-trick.pdf>
33. Ross A, Jain A, Nandakumar K (2006) Handbook of multibiometrics. Springer, New York
34. Matveev I, Novik V (2019) Using optimal circular path method to match piecewise iris templates. Pattern Recognit Image Anal 29(1):194–202
35. Dunstone T, Yager N (2009) Biometric system and data analysis: design, evaluation, and data mining. Springer, Boston
36. Chinese Academy of Sciences, Institute of Automation (CASIA), CASIA-Iris-Mobile-V1.0 (2015) <http://biometrics.idealtest.org/>