Machine Intelligence Research
www.mi-research.net

21(1), February 2024, 197-214
DOI: 10.1007/s11633-023-1415-y

Boosting Multi-modal Ocular Recognition via Spatial Feature Reconstruction and Unsupervised Image Quality Estimation

Zihui Yan[1] Yunlong Wang[1] Kunbo Zhang[1] Zhenan Sun[1] Lingxiao He[2]

[1] Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China

[2] JD AI Research, Beijing 100176, China

Abstract: In the daily application of an iris-recognition-at-a-distance (IAAD) system, many ocular images of low quality are acquired. As the iris part of these images is often not qualified for the recognition requirements, the more accessible periocular regions are a good complement for recognition. To further boost the performance of IAAD systems, a novel end-to-end framework for multi-modal ocular recognition is proposed. The proposed framework mainly consists of iris/periocular feature extraction and matching, unsupervised iris quality assessment, and a score-level adaptive weighted fusion strategy. First, ocular feature reconstruction (OFR) is proposed to sparsely reconstruct each probe image by high-quality gallery images based on proper feature maps. Next, a brand new unsupervised iris quality assessment method based on random multiscale embedding robustness is proposed. Different from the existing iris quality assessment methods, the quality of an iris image is measured by its robustness in the embedding space. At last, the fusion strategy exploits the iris quality score as the fusion weight to coalesce the complementary information from the iris and periocular regions. Extensive experimental results on ocular datasets prove that the proposed method is obviously better than unimodal biometrics, and the fusion strategy can significantly improve the recognition performance.

Keywords: Iris recognition, periocular recognition, spatial feature reconstruction, fully convolutional network, flexible matching, unsupervised iris quality assessment, adaptive weight fusion.

Citation: Z. Yan, Y. Wang, K. Zhang, Z. Sun, L. He. Boosting multi-modal ocular recognition via spatial feature reconstruction and unsupervised image quality estimation. *Machine Intelligence Research*, vol.21, no.1, pp.197–214, 2024. http://doi.org/10.1007/s11633-023-1415-y

1 Introduction

1.1 Motivation

In the disposition of border defence and the release of iris mobile phones, millions of people have registered their iris information as an individual identifying label. Iris recognition has been widely accepted by the public as a safe personal authentication approach. It has become a dependable and user-friendly biometric trait, especially in COVID-19 explosion. Nevertheless, iris recognition requires high user cooperation. This characteristic extremely limits its serviceable range. Therefore, iris recognition-at-a-distance (IAAD) with less user cooperation has gradually become a better choice for identity authentication. Although there have been a few meaningful attempts[1–3] in IAAD, the hardware quality and photographing conditions are the main research directions of these methods. The omnipresent low quality of the iris images acquired from the IAAD system, even the iris quality of some images does not meet the recognition requirements at all. These low-quality images are generally filtered out by strict image quality assessment mechanism in the existing iris recognition systems. Such a mechanism ensures the recognition accuracy while significantly reducing the recognition efficiency and throughput of the recognition system. As shown in Fig. 1, there are different low quality factors in IAAD, (a) occlusions by eyelids, (b) occlusions by eyeglasses, (c) occlusions by flare, (d) off-angle iris, and (e) deformations caused by pupil contraction and dilation. The normalized iris images corresponding to Fig. 1 are shown in Fig. 2. As we can see, these normalized images reveal more visible intra-class differences.

Lately, some impressive deep learning-based iris recognition methods[4–9] have been reported. The purpose of

Research Article
Manuscript received on December 12, 2022; accepted on January 11, 2023
Recommended by Associate Editor Ming-Ming Cheng
Colored figures are available in the online version at https://link.springer.com/journal/11633

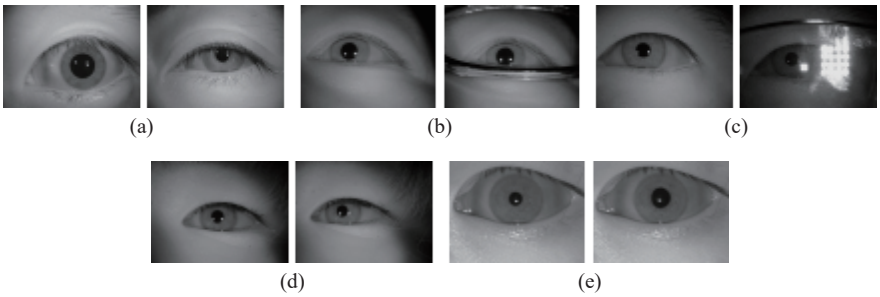


Fig. 1 Example pairs that have significant intra-class differences. Each pair is from the same eye. The left image is of relatively high quality, and the right is of relatively low quality. (a) Occlusions by eyelids; (b) Occlusions by eye-glasses; (c) Occlusions by flare; (d) Off-angle iris; (e) Deformations caused by pupil contraction and dilation.

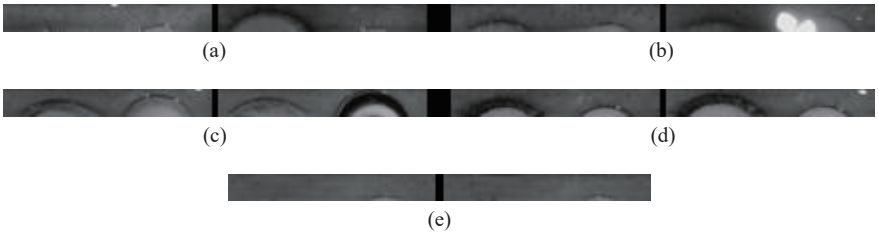


Fig. 2 Normalized images which are corresponding to the sequence number in Fig. 1

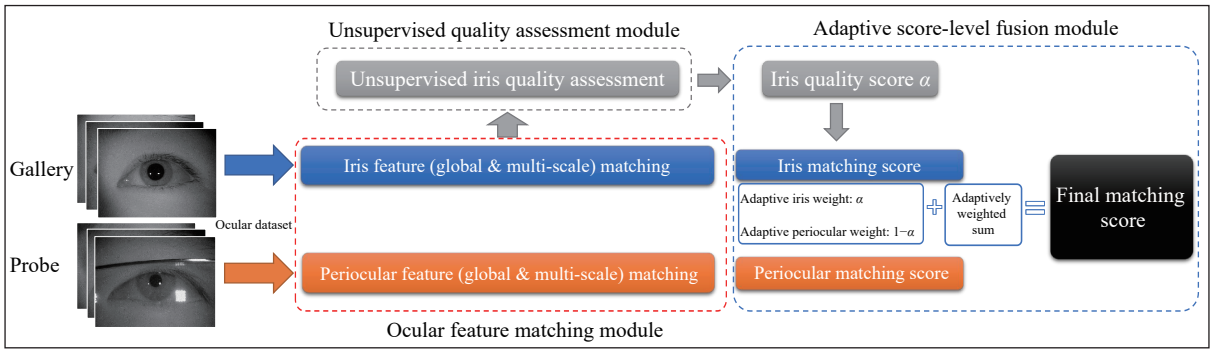
these deep neural networks (DNNs) based methods is to obtain the suitable deep features for intuitionistic pixel-to-pixel or region-to-region iris matching. Nevertheless, there are a lot of challenges, such as occlusions, deformations and misalignment in IAAD systems. The application scope of iris recognition is extremely limited by these existing deficiencies. On this basis, some research depends on the masking template of iris, and is based on pixel-to-pixel or region-to-region matching, starts to resolve iris recognition with occlusion. For instance, MaskNet[7] can calculate the mask of valid iris regions and mitigate the effect to iris recognition of occlusions. The inaccurate mask under occlusions can cause prominent performance degradation. Furthermore, there are some nonlinear deformations of iris texture caused by pupil dilation and contraction in iris images acquired by IAAD system. The reliability of pixel-to-pixel matching strategy is insufficient to these local deformations. How to boost IAAD feature matching performance by making use of strengths in deep learning is still a challenge.

On the other hand, a multi-modal fusion strategy is also a feasible choice to alleviate the performance degradation of IAAD. Periocular regions can be obtained simultaneously with the iris. When the iris part of some images is too low-quality to meet the recognition requirements, the more accessible periocular regions are a good complement for recognition.

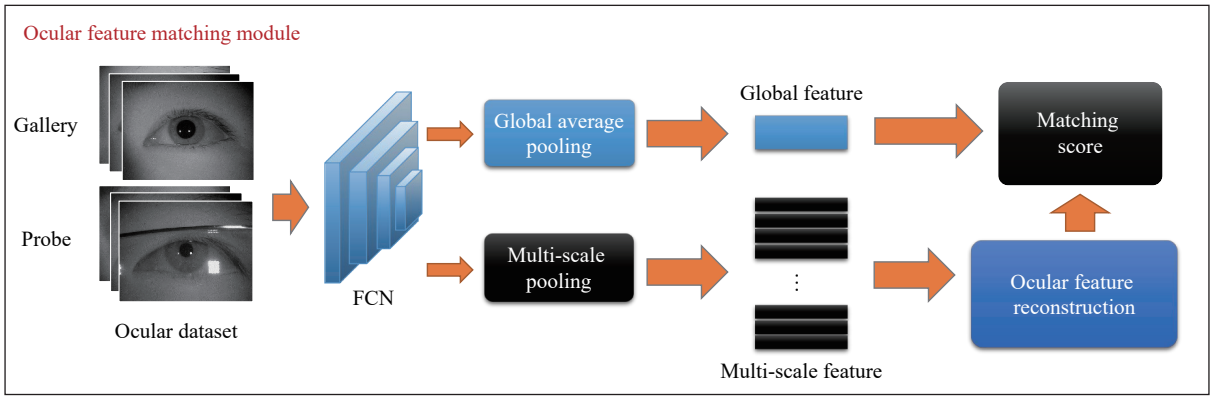
1.2 Contribution

This paper aims to exploit a united and flexible framework to recognize both iris and periocular images acquired in IAAD system. In our method, we elaborately combine deep learning and spatial feature reconstruction which is illuminated by [10]. As shown in Fig. 3, the whole framework is drawn. Such a matching mechanism and feature representation can achieve robustness and flexibility. Meanwhile, the recognition framework is also appropriate for periocular regions. To produce spatial feature maps in different sizes, the iris feature extractor consists of a fully convolutional network (FCN)[7]. Moreover, a triplet objective function is implemented after feature extractor. To obtain the most distinctive spatial features through reconstruction, we develop an end-to-end learning process. Each region in one probe image is anticipated to be sparsely reconstructed from the high-quality gallery images by ocular feature reconstruction (OFR). Thus, the interferences in the IAAD system can be restrained.

Meanwhile, as shown in Fig. 1, the information of the periocular regions is well preserved in these images captured from the iris recognition system. The periocular region has been proved viable for identification[11]. It has a satisfactory complementarity of recognition for iris. In previous research on biometrics fusion, simple score level fusion based on the weighted sum rule is employed in most methods[12, 13]. Mostly, the fusion weights are artificially determined with no optimization. In addition, the fusion methods on feature level are not maturely developed[14]. Adopting the weights of different modalities in different situations is mentioned in few works[15]. In order to obtain better fusion performance according to the different iris quality in various images, a new way to obtain fusion weights is proposed, and the weights are adaptively changing according to different ocular images. The weight is determined by the iris image quality, which is measured by a brand-new unsupervised iris quality assessment method based on random multiscale embedding robustness inspired by [16]. We exploit the iris quality score as the adaptive fusion weight to coalesce the com-



(a)



(b)

Fig. 3 Illustration of the whole framework: (a) Flowchart of the whole framework, which contains 3 main modules; (b) Flowchart of the ocular feature matching module.

plementary information captured from the iris and periocular regions. The current iris quality assessment methods require training data with quality labels judged by human′s perception, or come from reference scores. Such quality measurement methods are subjectively defined. Actually, the best characteristics for the employed iris recognition system may not be fully perceived. The performance of biometric recognition is partially determined by the quality of its samples[17]. The quality of biometric sample is defined as the utility of a sample for the purpose of recognition[17–20]. On the other side, automatic labelling based on comparison scores represents the relative performance of two samples. Therefore, the negative effect may be imposed to quality labels by the relative low-quality sample. Rethinking the shortcomings of these existing methods, the unsupervised iris quality assessment method is proposed to obtain a more accurate quality assessment score for the recognition system. The robustness of an image in the embedding space measures its quality. The representation robustness of the sample and thus, its quality is calculated by utilizing the variations of multiscale embeddings extracted from random subnetworks of the proposed model. This new iris quality assessment method can obtain more objective and reasonable quality scores for our OFR. Therefore, these quality scores can be used as more accurate adaptive score level fusion weight for multi-modal recognition in ocular regions, including iris and periocular.

The main contributions of this work include the following:

1) We propose a united framework, which called OFR, to recognize both iris and periocular images acquired in the IAAD system. It can accommodate the potential applications of IAAD. Extensive experiments show that the proposed OFR achieves satisfactory performance for both iris and periocular recognition. In addition, our method is alignment-free to arbitrarily sized periocular images and flexible to recognize the challenging iris images that contain occlusions and deformations.

2) We propose a novel unsupervised iris quality assessment concept by investigating the robustness of stochastic multiscale embeddings to acquire a more accurate quality assessment score for the recognition system. Our solution measures the quality of an image based on its robustness in the embedding space rather than human perception. It does not depend on the quality label of the training samples and provides a more direct quality assessment mechanism for OFR.

3) We develop a new way to obtain fusion weights. The novel quality assessment methods can acquire more appropriate fusion weights than the existing manual score-level fusion methods. Meanwhile, the weights are adaptively changing according to different ocular images. The score-level fusion strategy can make full use of the recognition advantages of iris and periocular and further improve the IAAD performance. It can not only keep the

high reliability of iris features, but also make use of the accessibility of periocular features. The proposed fusion method can achieve much better performance than traditional fusion methods. Meanwhile, it has outstanding improvements over unimodal recognition methods as a result of the automatic assigning of higher weights to more reasonable modalities.

2 Related work

2.1 Iris recognition

In the development of iris recognition over the last 20 years, there has been tremendous progress. Hand-crafted features are mainly employed in traditional iris recognition methods for iris texture representation. Daugman[21] proposed utilizing Gabor filters for quantization and phase demodulation. The Gabor filter produced 250 bytes IrisCodes and the dissimilarity between two IrisCodes is measured by Hamming distance. On this basis, an advanced methods based on IrisCode is proposed by Masek[22]. It utilized 1D log-Gabor filters. Then, ordinal measures (OM)[23] based framework of binary feature encoding was proposed in 2009.

On the other hand, a range of iris recognition approaches based on deep learning (DL) have also achieved excellent results. DeepIris[4] employed a DL framework to recognize heterogeneous iris images by a pairwise filter group. A DL architecture for iris representation designed for cross-sensor iris recognition was presented by Gangwar and Joshi[5]. Proeça and Neves[6] employed convolutional neural networks (CNNs) to differentiate the corresponding and noncorresponding iris pair. Zhao and Kumar[7] designed spatially corresponding iris feature descriptor to improve the recognition accuracy and generalization capability, especially for the iris images with occlusions. On this basis, dilated residual feature net (DRFNet)[24] was proposed, which utilized residual learning and dilated kernels. In 2018, Maxout CNNs was proposed to solve the difficulties of iris recognition on mobile equipments by Zhang et al.[8] Ren et al.[9] first explored the application of graphical models in iris recognition to learn dynamic graph representations. Neural architecture search (NAS) is first introduced into iris recognition by Nguyen et al.[25] The generalization capability between different sensors, subjects′ cooperation, and image collecting distances is improved by NAS. But the difficulties of recognizing low-quality images are not specifically studied. In 2021, Wei et al.[26] proposed Gabor trident network (GTN) to mitigate the distribution gap challenge and learn more discriminative features for cross-spectral recognition, and spectral adversarial network (SAN) to enhance the adversarial network′s ability to discriminate identity labels. Then in 2022, Wei et al.[27] proposed uncertainty embedding (UE) to generate a discriminative and robust iris representation and further mitigate the impact of uncertain acquisition factors in near-infrared iris recognition, and uncertainty-guided curriculum learning (UGCL) to optimize the model in an easy-to-hard manner for higher recognition accuracy. Nguyen et al.[28] proposed a novel fully complex-valued network to better cater to the unique properties of iris features, explicitly retaining the phase information in the feature representation. The complex-valued network shows a strong correspondence with the classic IrisCode and its phase-based derivatives. More related work is summarized in the recent survey[29]. How to recognize low-quality iris images is still an inevitable challenge. Nevertheless, the low-quality images captured by IAAD systems has not been well resolved. The related work of iris recognition is summarized in Table 1.

Table 1 Overview of the evolution of iris recognition methods

Authors	Feature category	Occlusion handling
Daugman[21]	Hand-crafted	Yes
Masek[22]	Hand-crafted	No
Sun and Tan[23]	Hand-crafted	Yes
Liu et al.[4]	DL-based	No
Gangwar and Joshi[5]	DL-based	No
Zhao and Kumar[7]	DL-based	Yes
Nguyen et al.[28]	DL-based	No

2.2 IAAD system

IAAD systems are always full of challenges. One of the most common is the quality of captured iris images is usually low, thus making for lessened recognition performance. In recent years, there are some meaningful attempts about IAAD. Matey et al.[1] proposed the iris on the move (IOM) system. The system had a camera at fixed position and acquire the iris from a moving subject with high cooperation. Yoon et al.[2] estimated a standing subject within the capture volume by a light stripe projector. Wheeler et al.[3] proposed another IAAD system which utilized two wide-FoV pre-calibrated cameras. It imposed a traditional stop-and-stare requirement. Zhang et al.[30] presented a focus adjustable iris camera by integrating a transmutable lens with a long-focus zoom lens. In 2021, Yan et al.[31] proposed a new method which called iris spatial feature reconstruction (ISFR). The combination of multiscale feature extraction and spatial feature reconstruction was first employed to improve the performance of IAAD. In this work, the iris images with occlusions and deformations are specially studied.

2.3 Periocular recognition

As summarized in [32], most of the existing methods utilized hand-crafted features, such as histogram of oriented gradients (HOG), local binary pattern (LBP) and scale invariant feature transform (SIFT) as feature. Furthermore, it is worth mentioning that there are some ad-

vanced DL based approaches to periocular recognition. Multi-source deep transfer learning which utilized a 5-layer CNN model for cross-sensor periocular recognition was employed in [33]. Zhao and Kumar[34] designed semantic assist based CNNs for less constrained periocular matching. Raja et al.[35] used binarized statistical image features (BSIF), speed up robust feature (SURF) and SIFT to generate periocular features and achieved an independent modular biometric framework on Android device. In their subsequent work[36], the cross-sensor periocular verification was studied. In 2016, there are a few methods based on deeply coupled auto-encoders[37] and deep sparse filter[38] in the *International Conference on Image Processing* (ICIP) 2016 mobile ocular biometric recognition competition[39]. Recently, Yin et al.[40] tried to eliminate the gap between interpretability and biometric recognition by learning feature specific filters that respond to a range of preferred spatial locations. Huang and Li[41] proposed an integrated solution that leverages parts′ discovery as an attention form. Tiong et al.[42] proposed using multi-feature fusion layers for multimodal facial biometrics, thereby leading to significant and informative data learning in dual-stream convolutional neural networks. This network consists of two progressive parts with distinct fusion strategies to aggregate RGB data and texture descriptors for multimodal facial biometrics.

As mentioned above, most of the existing methods utilized hand-crafted features. Moreover, the architectures of existing periocular recognition methods are different from iris recognition methods. So, it is necessary to incorporate two different recognition frameworks in a system that recognizes both iris and periocular, which can significantly reduce the overhead of IAAD systems.

2.4 Fusion of iris and periocular biometrics

It is a viable way that fuse iris and periocular biometrics to achieve better recognition performance. Several fusion approaches have been developed.

In the existing fusion methods, score-level fusion is the most common. Woodard et al.[43] utilized score-level fusion strategy based on the weighted sum rule to improve the recognition performance in non-ideal images, and the weights are determined by searching in the range [0.1, 0.9]. The weighted sum rule score-level fusion strategy was also utilized to improve the recognition performance under relaxed imaging constraints by Tan and Kumar[12]. Raja et al.[44] designed a multimodal authentication system including three layers to fuse iris and periocular information at the score level for mobile cross-sensor applications.

Besides, there were other fusion methods at feature-level and decision-level in recent years. Santos and Hoyle[45] proposed a fusion strategy of iris and periocular at decision-level to enhance the reliability of the unconstrained iris recognition system. Joshi et al.[15] fused iris and periocular at feature-level by concatenating iris and periocular features. Luo et al.[46] proposed an end-to-end deep feature fusion network, where integrated multiple attention mechanisms including self-attention and co-attention mechanisms, for joint iris-periocular recognition. A periocular-assisted dynamic framework for more accurate less-constrained iris recognition was presented by Wang and Kumar[47]. This work consists of two kinds of network architectures to recognize iris and periocular respectively. The related work of fusion by iris and periocular biometrics is summarized in Table 2.

Table 2 Overview of the evolution of fusing iris and periocular biometrics

Authors	Iris feature	Periocular feature	Fusion strategy
Woodard et al.[43]	Gabor	LBP	Score level
Tan and Kumar[12]	Gabor	SIFT, GIST, LBP, HOG, LMF	Score level
Raja et al.[44]	Gabor	SIFT, SURF, BSIF	Score level
Santos and Hogle[45]	Wavelets	LBP, SIFT	Decision level
Joshi et al.[15]	Wavelets	LBP	Feature level
Zhang et al.[8]	DL	DL	Feature level
Wang and Kumar[47]	DL	DL	Score level

In the previous research about iris and periocular fusion, score-level fusion has not been studied intensively in existing work. Because the fusing performance of multimodal mostly depends on the accuracy of the fusion parameters[48], obtaining the most appropriate fusion weight is demanding in a fusion strategy.

3 Methodology

Obtaining the most valid representation for OFR of partial iris/periocular regions, and then fusing iris and periocular matching scores with the adaptive fusion weight determined by iris quality are the main purposes of our method. There are several key elements, the design of FCN based spatially corresponding feature extraction, multiscale feature aggregation, ocular feature reconstruction, the application of loss function in optimization, iris quality assessment and fusion strategy. Next, we will introduce the technical details.

3.1 Network architecture

The pre-trained models on ImageNet[49] include fully connected (FC) layers and convolution layers, such as VGG[50] and ResNet[51]. Comparing with general CNNs, the FCN contains convolutional layers, pooling layers, and activation layers, which efficiently acquire local pixel representation from their bottom map. However, it does not contain FC layers. Therefore, the original spatial information can be maintained in the output map. Previ-

ous iris recognition research mainly utilized block-based operations or small-size filters to acquire iris features. It is pointed out that the little structural information or meaningful hierarchies in iris pattern is not sufficient in [7], and the most discriminative feature is ought to be contained in the local texture information. The local pixel-to-pixel correspondence between input and output can be retained by FCN. We remove the FC layers in ResNet-50 to formulate FCN as a feature encoder. Then in the training process the parameters are fine-tuned. The spatial features are extracted by FCN and post-processed to two kinds of features. We call them global features and multiscale spatial features. In detail, the global average pooling (GAP) layer can extract global features and these features are used to replenish the overall information. A sequence of multiscale pooling layers can extract multiscale spatial features and these features are used to represent the distribution of locally valid features in different sizes. The multiscale spatial features are passed into OFR in the feature matching process. OFR is a size-free reconstruction mechanism based on dictionary learning. Finally, the weighted sum of OFR similarity distances and global matching results are defined as the matching score.

3.2 Feature representation

Spatial features and global features are combined for the description of ocular feature. The spatial features collect the local iris texture or periocular structure in different scales, and they are a batch of features captured from various receptive fields in different scales to perform matching images of arbitrary size better. The global features is a scalar vector which can obtain useful overall information about each image.

Furthermore, it is a challenging problem that recognizing the effective region in various size of iris/periocular images with deformations and occlusions. It is also non-trivial to align arbitrary-size iris/periocular images to a predefined scale. Meanwhile, recognizing the unaligned periocular images is also a problem to be solved. Hence, the effective region in different scales between two images are going to cause degraded performance and mismatch easily. The multiscale spatial features are alignment-free and more robust for deformations and occlusions. Thus, it can mitigate the influence of different effect scales and disalignments.

The multiple average pooling layers in different kernel sizes corresponding to various receptive fields are drawn in Fig. 4. They constitute the multiscale pooling layer. Due to the different feature distribution between iris and periocular, we select different scales of multi-scale pooling layers. During the iris matching process, this pooling layer takes the preceding output of FCN as input and produces multiscale spatial features by average pooling operations at a stride 2. Dense spatial features are generated by the output extracted by the pooling layer of each kernel size. The local features of iris are represented by each spatial feature. Similarly, the periocular region matching process utilize average pooling operations at a stride 3. We concatenate the output of multiscale pooling layers in different kernel sizes as multiscale spatial features, which is represented by $F_M(x)$.

3.3 Ocular feature reconstruction

There is a crucial but challenging puzzler, how to calculate the similarity measure between two images which are with disalignment, varying occlusions and deformations. A spatial feature reconstruction step, which is a reconstruction mechanism based on dictionary learning, is appended to moderate the influence of the cross-scale mismatch. The iris images after normalization are shown in Fig. 5 and their corresponding periocular images are given in Fig. 6. The process of generating the multiscale ocular features is summarized in Fig. 7. A pair of ocular images with considerable differences in quality is employed as input. I denotes an image in low quality and J denotes an image in high quality. Then, the multi-scale pooling layer obtain correspondingly sized spatial feature maps $x = \text{conv}(I, \theta)$ and $y = \text{conv}(J, \theta)$, where θ represents the parameters in FCN. x denotes a vectorized tensor in the size of $(d \times w \times h)$, where d, w and h represent the numbers of the channel, width and height of x respectively. $x_1, \cdots, x_N$ denote the N blocks which are divided from x, and $N = w \times h$. $x_1, \cdots, x_N$ consist of a block set and each of them is in size of $1 \times 1 \times d$. $F_M(x)$ is utilized to represent the concatenated output of multiscale pooling layers in different kernel sizes, namely the multiscale spatial features. The block set is represented by

$$X = F_M(x) = \{x_1, \cdots, x_N\} \in \mathbf{R}^{d \times N} \tag{1}$$

where $x_n \in \mathbf{R}^{D \times 1}$. Similarly, y is divided into M blocks as

$$Y = F_M(y) = \{y_1, \cdots, y_M\} \in \mathbf{R}^{d \times M}. \tag{2}$$

Hence, x_n can be expressed by a linear combination of Y, viz. reconstruct x_n by finding out the similar blocks. We define the sparse coefficients as ω_n, where $\omega_n \in \mathbf{R}^{M \times 1}$, of x_n in relation to Y. Thus, we work out ω_n. ω_n is constrained with L_2 norm. We define the sparse representation equation as

$$\min_{\omega_n} \|x_n - Y\omega_n\|_2^2 + \beta \|\omega_n\|_2 . \tag{3}$$

The similarity between x_n and $Y\omega_n$ is measured by the first part. And the second part can astrict the sparsity of ω_n. Considering that there are N spatial features in X, (3) is overwritten as

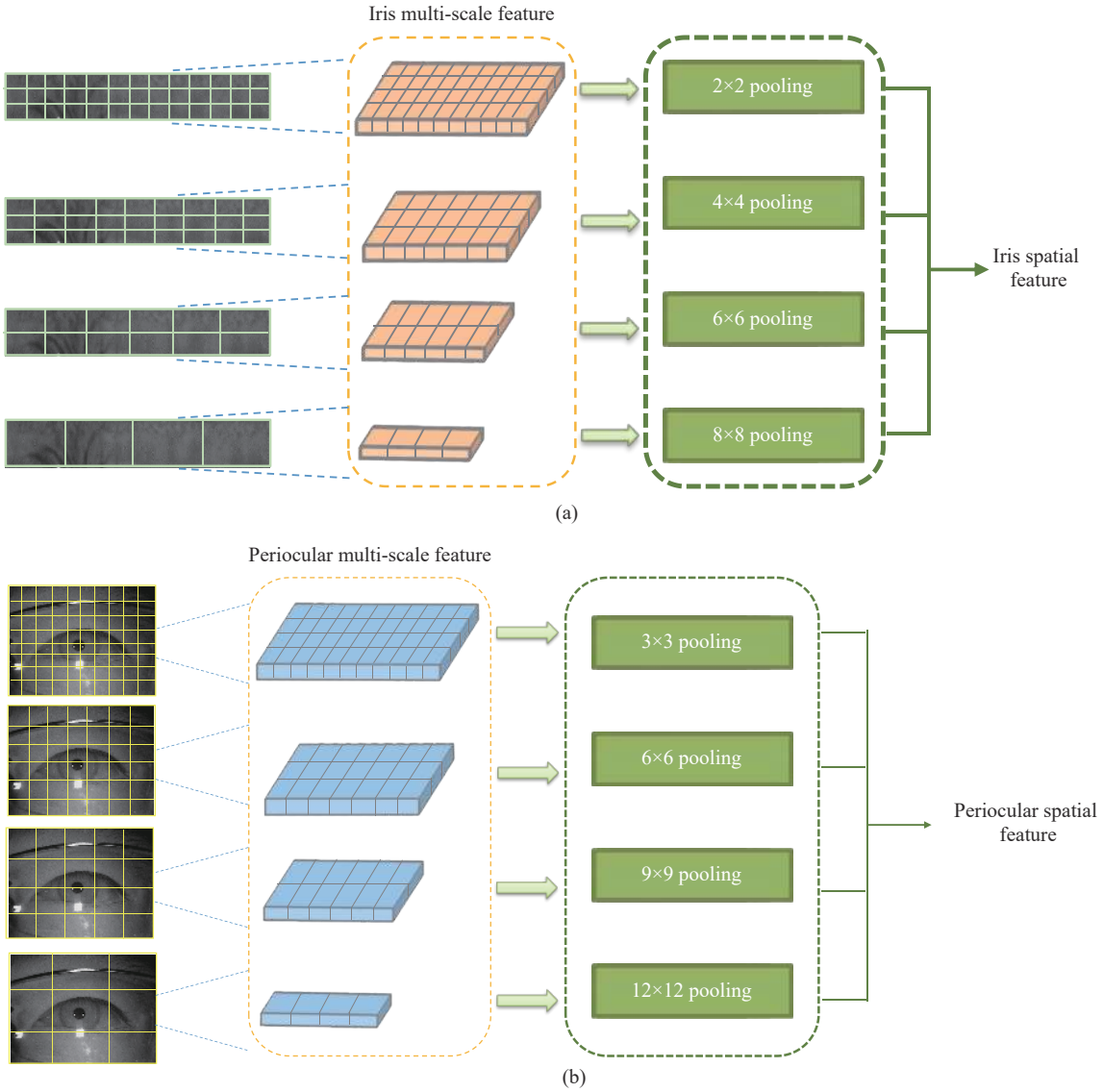
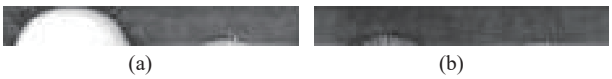
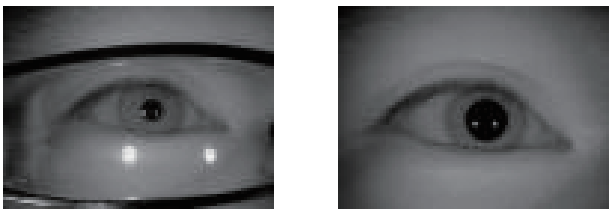


Fig. 4 Multiscale feature representation of (a) iris, (b) periocular



(a) (b)

Fig. 5 Examples of iris images after normalization



(a) (b)

Fig. 6 Examples of periocular images

$$\min_{W} \|X - YW\|_2^2 + \beta\|W\|_F \tag{4}$$

where $W = \omega_1, \cdots, \omega_N \in \mathbf{R}^{M\times N}$ denotes sparse reconstruction coefficient matrix. W is solved by the least-squares algorithm. $W = (Y^{\mathrm{T}}Y + \beta I)^{-1}Y^{\mathrm{T}}X$. We suppose $M = X - YW$, the distance of sparse reconstruction between X and Y is denoted by

$$D_r = \frac{\mathrm{tr}\left(\sqrt{M^{\mathrm{T}}M}\right)}{N}. \tag{5}$$

The OFR strategy is the crux of robustness and flexibility to match iris/periocular images with partial occlusions. Therefore, OFR can use available iris/periocular regions for rational identity verification.

3.4 Loss function

We employ the batch triplet loss function in our method. The purpose of the batch triplet loss function is to insure that an anchor image $x_{a,i}$ of a specific eye is closer to all other positive images $x_{p,i}$ of the same eye than it is to any negative image $x_{n,i}$ of any other eye.

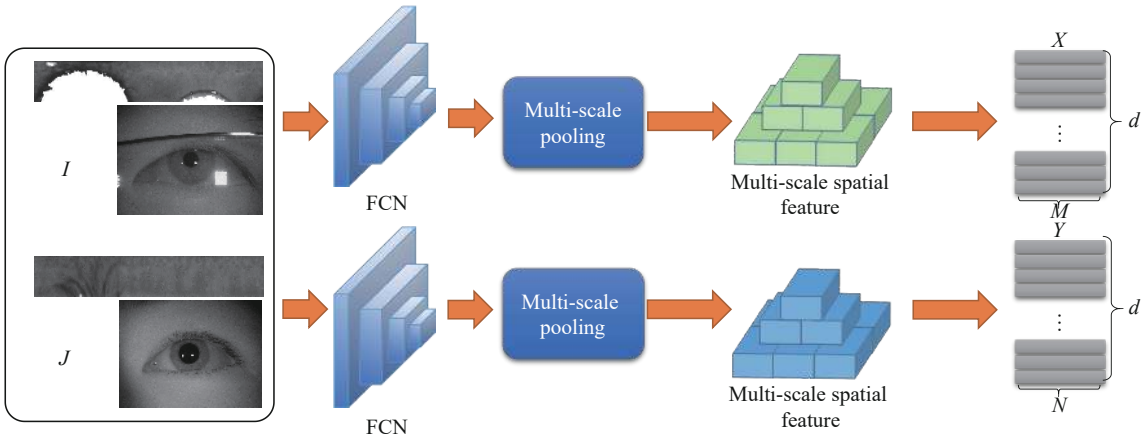


Fig. 7 Schematic of generating multiscale ocular spatial feature, X and Y represent the multiscale feature sets which are generated from image Y and J, respectively.

Hence, we try to make $D(x_{a,i}, x_{n,i}) > D(x_{a,i}, x_{p,i}) + c$, where c enforces the separation between a pair of samples (positive and negative) and D denotes the Euclidean distance between a pair of images. So the loss function is represented by

$$L(\theta) = \sum_{i}^{N} \left[D\left(g_{a,i}, g_{p,i}\right) - D\left(g_{a,i}, g_{n,i}\right) + c \right] \tag{6}$$

where $g_{a,i} = GAP(f_\theta(x_{a,i}))$, $g_{p,i} = GAP(f_\theta(x_{p,i}))$, $g_{n,i} = GAP(f_\theta(x_{n,i}))$.

We randomly choose S subjects. Then, we select K images from each subject to pick out more effective triplet samples. In this way, a batch which contains SK images is obtained. In the batch, a hard example mining technique is adopted for each anchor sample. It means both the hardest positive sample and the hardest negative sample will be chosen. The batch triplet loss function for a mini-batch, viz. a data point $x_{j,i}$, is represented as

$$L_B(\theta) = \sum_{i=1}^{S} \sum_{a=1}^{K} \left[\max_{p=1,\cdots,K} D\left(g_{a,i}, g_{p,i}\right) - \min_{n=1,\cdots,K} D\left(g_{a,i}, g_{n,i}\right) + c \right] \tag{7}$$

where $x_{j,i}$ denotes the j-th image of the i-th subject in the batch. Therefore, SK terms conduce to the loss. The chosen triplets are optimum for learning with the triplet loss and the hardest in a small subset

$$L_B(\theta) = \sum_{i=1}^{S} \sum_{a=1}^{K} \left[\max_{p=1,\cdots,K} \left(D_r(X_{a,i}, X_{p,i}) + D(g_{a,i}, g_{p,i})\right) - \min_{n=1,\cdots,K} \left(D_r\left(X_{a,i}, X_{n,i}\right) + D\left(g_{a,i}, g_{n,i}\right)\right) + c \right] \tag{8}$$

where D_r denotes the reconstruction distance. The similarity distance is expressed in (8). It is composed of two parts: the distance of local feature matching (spatial reconstruction error) and that of global feature matching (Euclidean distance).

Then, an alternating optimization algorithm is employed to solve θ.

Step 1. Optimize $\boldsymbol{W_i^{ap}}, \boldsymbol{W_i^{an}}$ and fix θ.

The sparse reconstruction coefficient matrix is expressed by

$$W_i^{ap} = \left(X_{p,i}^{\mathrm{T}} X_{p,i} + \beta \cdot I\right)^{-1} X_{p,i}^{\mathrm{T}} X_{a,i}$$
$$W_i^{an} = \left(X_{n,i}^{\mathrm{T}} X_{n,i} + \beta \cdot I\right)^{-1} X_{n,i}^{\mathrm{T}} X_{a,i}.$$

To figure out this matrix. First of all, W is optimized and θ is fixed. $\omega_1, \cdots, \omega_N$ are solved respectively when optimizing. As adopted in [52], an optimal ω_N is solved by the feature-sign search algorithm.

Step 2. Fix $\boldsymbol{W_i^{ap}}, \boldsymbol{W_i^{an}}$ and optimize θ.

The gradients of $D_r(X_{a,i}, X_{p,i})$ with respect to $X_{a,i}$ and $X_{p,i}$ and the gradients of $D_r(X_{a,i}, X_{n,i})$ with respect to $X_{a,i}$ and $X_{n,i}$ are expressed by

$$\begin{aligned}
\frac{\partial D_r\left(X_{a,i}, X_{p,i}\right)}{\partial X_{a,i}} &= 2\left(X_{a,i} - X_{p,i} W_i^{ap}\right) \\
\frac{\partial D_r\left(X_{a,i}, X_{p,i}\right)}{\partial X_{p,i}} &= -2\left(X_{a,i} - X_{p,i} W_i^{ap}\right) W_i^{ap\mathrm{T}} \\
\frac{\partial D_r\left(X_{a,i}, X_{n,i}\right)}{\partial X_{a,i}} &= 2\left(X_{a,i} - X_{n,i} W_i^{an}\right) \\
\frac{\partial D_r\left(X_{a,i}, X_{n,i}\right)}{\partial X_{n,i}} &= -2\left(X_{a,i} - X_{n,i} W_i^{an}\right) W_i^{an\mathrm{T}}.
\end{aligned} \tag{9}$$

To calculate $\partial L / \partial \theta$, we utilize these four equations. Consequently, FCN supervised by OFR can employ standard stochastic gradient descent (SGD) to optimize and is end-to-end trainable.

3.5 Dual feature matching

In our proposed method, global features, multiscale spatial features, and the weighted fusion of these two fea-

tures are utilized for iris/periocular feature matching process. We assume global feature and spatial feature obtained from gallery subject c are g_c and S_c, respectively. The global feature set can be denoted as $G = [g_1, g_2, \cdots, g_c]$, where $g_c \in \mathbf{R}^d$. The spatial feature set of the gallery can be denoted as $S = [S_1, S_2, \cdots, S_c]$, where $S_c \in \mathbf{R}^{k_c \times d}$ and k_c represents the amount of spatial features. As for the probe image, the global feature is denoted by p_g and spatial feature is denoted by X.

The similarity of global features between two images is measured by Euclidean distance. The distance is defined as $d_c = \|p_g - g_c\|_2$. Thus, the global similarity distance vector for total C subjects is represented by

$$D = [d_1, d_2, \cdots, d_C]. \tag{10}$$

Except for global features, the spatial feature plays a more important role in our method. It can not only acquire the local texture in iris images, but also perform matching with disalignment. Therefore, the representation of multiscale spatial feature is robust to scale diversities. The similar spatial features can be searched by OFR from the gallery to reconstruct the spatial feature of the probe with minimum error. The similarity distance r_c of spatial reconstruction is denoted by

$$r_c = D_r(X, S_c) = \frac{\mathrm{tr}\left(\sqrt{M^{\mathrm{T}} M}\right)}{k_c}. \tag{11}$$

In (11), $M = X - S_c W_c$. Thereinto, $W_c = (S_c^{\mathrm{T}} S_c + \beta \cdot I)^{-1} S_c^{\mathrm{T}} X$. Hence, the similarity distances for total C subjects form a distance vector, which can be expressed by

$$R = [r_1, r_2, \cdots, r_C]. \tag{12}$$

To realize higher accuracy, the weighted sum of these two vectors forms the final distance. The distance is represented by

$$S_{overall} = \alpha \cdot D + (1 - \alpha) \cdot R. \tag{13}$$

In (13), α is a value which can regulate the weight of spatial reconstruction and global feature matching. In this way, the identity is decided by $c_k = \arg\min_c S_c$, where S_c is the c-th item of S_{overall}.

3.6 Iris quality assessment and fusion strategy

In this part, we will introduce the unsupervised iris quality assessment based on random multiscale embedding robustness and the fusion strategy of iris and periocular.

The purpose of iris quality assessment is to estimate the suitability of an iris image for iris recognition. Actually, the quality of an iris image should reveal its expected recognition performance. Significantly, our iris image quality definition is based on the relative robustness of deeply learned multiscale embeddings of that image. The image quality is defined as the magnitude of variations, which are calculated by the variations of embeddings generated from random subnetworks of cross-scale spatial feature extractor. The illustration of this method is shown in Fig. 8.

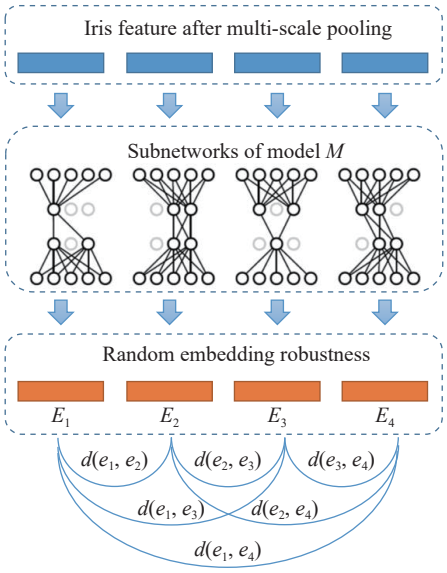


Fig. 8 Illustration of iris quality assessment: An input iris image I is forwarded to different random subnetworks of the used multiscale spatial feature model M. Each subnetwork produces a different stochastic embedding E_I. The variations between these embeddings are calculated by pairwise distances and define the quality of I.

In addition, our method predicts the iris quality $Q(I)$ of a given iris image I using multiscale spatial features. The final layer of the multiscale spatial feature extractor is passed to the dropout operation and the stochastic forward passes are only repeated in the last layers. Each stochastic forward pass, which captures different scale spatial features, applies a different dropout pattern (during prediction) producing a different subnetwork of M. These subnetworks generates different stochastic multiscale iris embeddings. These embeddings are collected in a set $E(I) = \{e_1, \cdots, e_m\}$. As we mentioned above, the probe iris image is represented by a linear combination of the gallery image. The total number of non-zero sparse coefficients in W is N_{eff}. We define the iris quality of image I as

$$Q(I) = \frac{\sum_{N_{eff}} 2\sigma\left(-\frac{2}{m^2}\sum_{i<j} d(e_i, e_j)\right)}{N_{eff}} \tag{14}$$

where $d(e_i, e_j)$ represents the Euclidean distance between a stochastic embeddings pairs. In other words, the iris quality of an image is defined as the sigmoid of the

negative mean Euclidean distance between all stochastic embeddings pairs. The sigmoid function $\sigma(\cdot)$ ensures that the quality score $Q(I) \in [0, 1]$. A greater variation in the stochastic embedding set E indicates a low robustness of the representation and, thus, a lower sample quality Q. Lower variations in E indicate high robustness in the embedding space and is considered as a high sample quality Q.

Our method aims to evaluate the iris image quality from the perspective of employment in recognition tasks, which might be different from accessing the notion of image quality. Given that the recognition network training aims at being robust against intra-identity variations, an image which produces relatively stable identity-related embeddings despite various variations should have high utilization in a recognition task. The agreement between the subnetworks can be used to estimate the embedding robustness. If the m subnetworks have similar outputs (high agreement), and the variations over these random subnetworks (the stochastic embedding set E) are low. Hence, the robustness of this embedding, which is correlated to the quality of the sample, is high. Otherwise, if m subnetworks have dissimilar outputs (low agreement), and the variations over the random subnetworks are high. Consequently, the robustness in the embedding space is low and the quality of the sample can be considered low. In contrast to previous work, our method does not require quality labels for training.

To retain the recognizing advantages of iris and periocular as much as possible, the iris matching score and periocular matching score are weighted and summed up to achieve score-level fusion. The final matching score can be denoted as

$$S_f = Q(I) \times S_{overall}(I) + (1 - Q(I)) \times S_{overall}(P) \quad (15)$$

where $S_{overall}(I)$ represents the iris matching score of an image and $S_{overall}(P)$ represents the periocular matching score of the same image. Furthermore, $Q(I)$ will be set as 0 when the iris part of the probe image fails to match.

3.7 Preprocessing of input

In our method, the segmentation method proposed in [53] is employed. Then, Daugman's rubber sheet model[54] is utilized to normalize the iris region to polar coordinates and the original resolution of normalized images is 540 × 70. Square iris images are employed as input in [5] to retain as much information of short edges as possible. If the input directly utilizes the rectangle images, the filters in tail convolutional layers tend to ignore the information of short edges. In [8], it is proved that the input of square iris images achieves better performance than the input of rectangle. The width of an iris (iris radius minus pupil radius) in iris images is less than 70 pixels. Therefore, the differences between square input images processed by indirect resizing or direct normalization are minor. We preprocess the iris images into square as mentioned in [5]. Then, they are resized to 224 × 224 to input into the feature extractor.

As for periocular images, we have already acquired the the iris centre in iris image preprocessing. The original ocular image size is 640 × 480. Since our method does not have strict requirements for image alignment, we directly utilize the centre of the iris as the centre point to trim a 480 × 480 image from the raw image as the periocular biometric. Then these square images are resized in the same way as the iris images.

4 Experiment

To verify the validity of our method more comprehensively, it is tested on a few public ocular biometric datasets from multiaspects. First of all, to demonstrate the effectiveness of our method for iris recognition, we test on 4 public iris datasets. According to [14], UBIRIS.v2 datasets and CASIA-Iris-Distance can be used to evaluate the effectiveness of IAAD system. Next, we study the influence of different occluded area on iris recognition performance by adding 20%, 30%, 40% and 50% occluded dimension on iris images. The performance is compared with that of several existing approaches on different sizes of occlusion areas. We randomly select the location of occlusion and adopt a rectangular occluded shape. As shown in Fig. 9, the iris images are added random rectangular occlusion. After that, to prove the effectiveness for periocular recognition, we test our method on 2 public periocular datasets. Finally, the performance of the fusion strategy is tested on the same 2 datasets. As for the performance index, we employed the equal error rate (EER), false rejection rate (FRR), false acceptance rate (FAR), and the detection error trade-off (DET) curve.

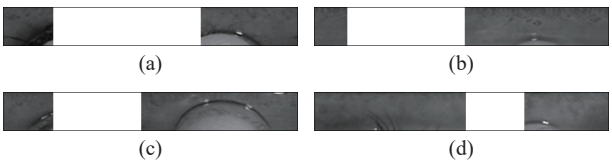
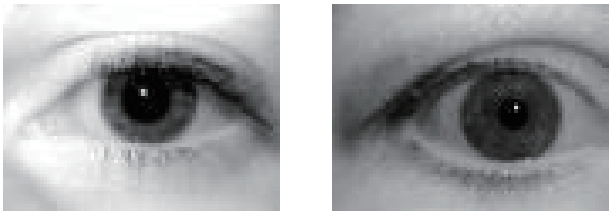


Fig. 9 Examples of iris images with random occlusion

4.1 Iris and periocular datasets

4.1.1 ND-IRIS-0405 iris image dataset

ND-IRIS-0405[55] contains 64 980 images from 356 subjects. There are no subjects which wore glasses during image collection. In this dataset, an LG 2 200 sensor was employed to acquire the images. This sensor utilized near-infrared lumination during collecting and assisted the subject situate their head properly for image collection by audible prompts. As shown in Fig. 10, there are ND-IRIS-0405 examples.



(a) (b)

Fig. 10 Examples of the ND-IRIS-0405 images

4.1.2 CASIA-Iris-Distance

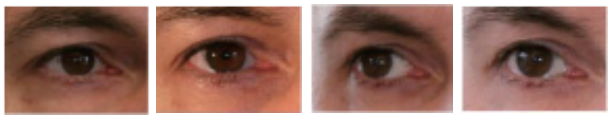
CASIA-Iris-Distance contains 2 572 images from 142 subjects. Each image included both two eyes and was collected from 3 meters away. To crop the eye regions, a detector implemented by OpenCV was employed.

4.1.3 CASIA-Iris-M1-S3

CASIA-Iris-M1-S3 contains 3 600 images from 360 Asian subjects. These images are captured by a mobile phone which installed NIR iris scanner. It has the largest number of subjects among the NIR iris dataset acquired on mobile equipment.

4.1.4 UBIRIS.v2 iris dataset

UBIRIS.v2[56] contains 11, 102 iris images from 261 subjects. The images were acquired under unconstrained conditions with realistic noisy factors. The images in UBIRIS.v2 are collected under visible wavelength lumination rather than near-infrared wavelength. The primary purpose is to evaluate the feasibility of visible wavelength iris recognition under far-from-ideal shooting conditions. There are images with different kinds of occlusions in UBIRIS.v2. This dataset is appropriate for proving the recognizing ability in visible wavelength. There are UBIRIS.v2 examples in Fig. 11.



(a) (b) (c) (d)

Fig. 11 Examples of the UBIRIS.v2 iris dataset images

4.2 General settings

Our project is based on the PyTorch framework. Linux with GTX TITAN X GPU is employed to train and test all models in this paper. The model is trained with 500 epochs. To perform the best performance, we set $c = 0.25$ and $\beta = 0.001$. The set of c and β is referred to [31].

4.3 Experiments on iris recognition

4.3.1 Iris recognition on public datasets

In this section, we focus on demonstrating the effectiveness of the proposed method on public datasets. The comparison with existing methods is more intuitive through the experiments in this section.

To pretrain the FCN, we utilize some images from CASIA-Iris-Thousand. After obtaining the the pretrained FCN, we fine-tune it by different datasets. To fine-tune the pretrained FCN with OFR, five hundred positive pairs and five hundred negative pairs of images are utilized. For each image pair, there are one intact image and one occluded image. As shown in Fig. 2, all the images are normalized to the size of 540 × 70. As introduced in Section 3.6, these normalized images are transformed into the shape of square.

We train and test our method on the first three datasets. In ND-IRIS-0405, the first 25 images of all subjects which belong to left eye are chosen for training, and the first 10 images of all subjects which belong to right eye are chosen for testing. After removing some incorrectly segmented samples, the testing set includes 5 743 231 impostor pairs and 14 780 genuine pairs. Next for the CASIA-Iris-Distance dataset, we choose the images of all subjects which belong to right eye as training set and the images of all subjects which belong to left eye as testing set. The testing set contains 2 969 533 impostor pairs and 20 702 genuine pairs. Then the first 180 subjects of CASIA-Iris-M1-S3 are used for training and the last 180 subjects for testing. The testing set generates 1 611 000 impostor pairs and 8 100 genuine pairs. As for UBIRIS.v2, we select a total of 2 000 images of the first 100 subjects for our experiments. The first 10 images of left eye and the first 10 images of right eye in each subject are maintained. We use the images of all subjects belong to left eye for training, and the images of all subjects belong to right eye for testing. The testing set generates 495 000 impostor pairs and 4 500 genuine pairs. Four DL-based methods are selected for comparison: VGG-16[50], pairwise CNNs[4], ZhaoICCV2017[7] and DRFNet[24]. Because the source code of [24] is lacking, it is compared on only two datasets. In [7], four Caffe models are trained on ND-IRIS-0405 Iris, CASIA-Iris-Distance, CASIA-Iris-M1-S3 and UBIRIS.v2, respectively. A brand new expanding loss function is utilized. Additional supervision of iris region masks is necessary. In the corresponding experiments, the test dataset of CASIA-Iris-Distance and the models trained on ND-IRIS-0405 are directly tested. Furthermore, to make the experiments more comprehensive and convincing, we also select some conventional and representative iris recognition algorithms based on hand-crafted features and traditional local features for comparison, such as the LBP feature-based USIT, Gabor filter-based OSIRIS, and ordinal measures (OMs)[23]. To select the most discriminative features of OMs, AdaBoost algorithm is utilized for comparison. Inbuilt iris segmentation/normalization programs are employed. In addition, we perform single-eye conditions in these experiments.

As shown in Table 3 and Fig. 12, the experimental results and detection error tradeoff (DET) curves are listed

Table 3 Summary of EERs and FRRs at 0.1% FAR for the iris recognition

	ND-IRIS-0405		CASIA.v4-distance		CASIA-M1-S3		UBIRIS.v2	
	FRR at FAR = 0.1 %	EER	FRR at FAR = 0.1 %	EER	FRR at FAR = 0.1 %	EER	FRR at FAR = 0.1 %	EER
OSIRIS	3.73%	1.70%	21.15%	6.52%	41.63%	11.02%	–	–
USIT-LBP	3.31%	1.87%	20.32%	6.45%	40.62%	10.70%	–	–
OMS	3.22%	1.74%	17.05%	6.21%	14.62%	3.06%	–	–
VGG16	2.03%	1.05%	15.17%	6.09%	16.81%	4.13%	16.34%	8.73%
Pairwise-CNN	1.79%	0.92%	14.92%	5.83%	13.01%	3.65%	13.25%	7.12%
UniNet	1.78%	0.99%	11.54%	3.88%	8.76%	3.01%	13.93%	6.67%
DRFNet	1.62%	0.82%	9.21%	2.92%	–	–	–	–
OFR (single-scale)	**1.64%**	**0.80%**	**8.45%**	**2.62%**	**7.02%**	**2.12%**	**12.34%**	**6.49%**
OFR (multi-scale)	**1.46%**	**0.75%**	**7.83%**	**2.17%**	**5.89%**	**1.63%**	**10.67%**	**5.58%**

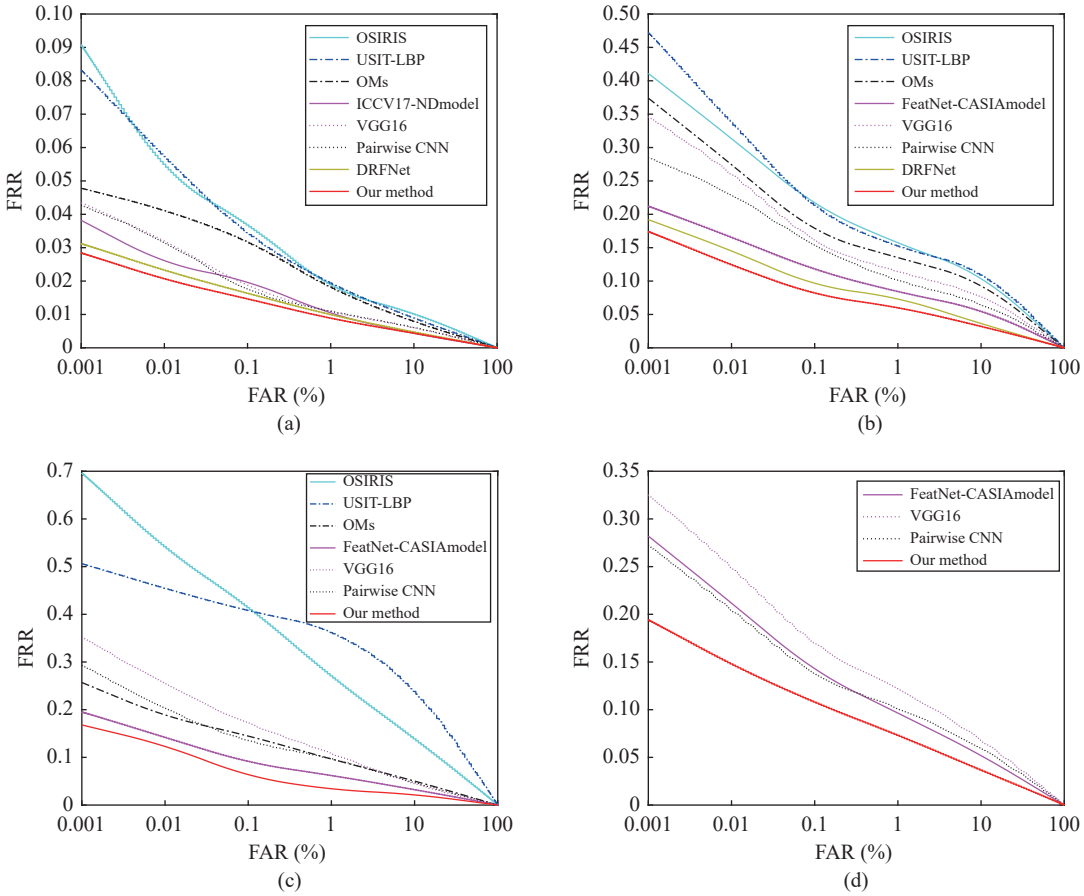


Fig. 12 DET curves on four public datasets for iris recognition. (a) ND-IRIS-0405; (b) CASIA-Iris-Distance; (c) CASIA-Iris-M1-S3; (d) UBIRIS.v2.

more intuitively. We do not only test our method on the datasets collected under NIR wavelength, but also dataset collected under visible wavelength to prove the universality and flexibility. Besides, the ablation experiments of single-scale are conducted to test the efficacy of multiscale features, and the experimental results are listed in Table 3.

As the results demonstrate, it is proved that our proposed method can significantly improve the recognition performance on the mentioned datasets. The application of spatial feature reconstruction and the robustness of multiscale spatial features bring more reliability. The combination of global features and multiscale features is more effective for iris recognition than the features extracted by CNNs. Because of the stronger recognizing ability on the images with occlusions and deformation, our method achieves better performance in these public datasets. Meanwhile, the multiscale pooling has more ro-

bust receptive field than the single-scale pooling. The results prove that the proposed method has much more robustness and universality too. Besides, it demonstrates that our method does not entirely depend on a high-quality gallery of these datasets. More importantly, our method is not only valid on the NIR images but also gets satisfying results on the images in visible wavelength. It is proved that our method has good generalization ability and is extremely valuable for practical applications in IAAD systems. In summary, our method can significantly boost the iris recognizing performance.

4.3.2 Influence of different occlusion area sizes on recognition

In this part, the influence of different occluded sizes on recognition performance is investigated. To find out the results, the iris images from CASIA-Iris-Distance are covered with 50%, 40%, 30%, and 20% occlusions randomly. For the sake of fairness, other methods based on DL are also tested for comparing, including VGG-16[50] and Zhao-ICCV2017[7]. The DET curves of different occluded sizes are listed in Fig. 13.

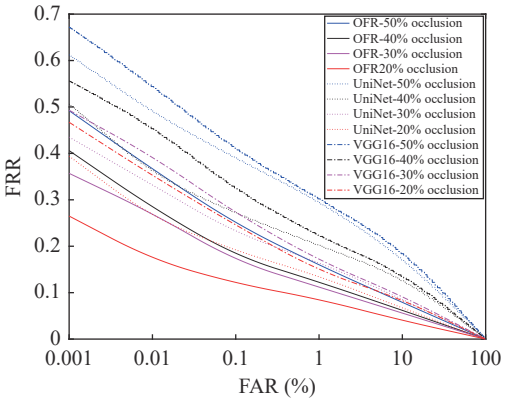


Fig. 13 Comparison of the effect of the occluded size on the recognition performance

After comparing with other DL based methods, our method has much more robustness to the variations in occluded size. Even though the occluded size is 50% of the whole image, our method can achieve satisfactory recognition performance. By comparison, the state of the art (SOTA) methods is not robust enough to the variations in occluded size. The recognition performance degrades prominently. The larger the occlusion area, the lower the accuracy.

4.3.3 Cross-dataset performance evaluation

The generalization is a crucial factor in practical applications of recognition system. Therefore, the cross-dataset performance is a very important indicator to value the generalization. In this part, we evaluate the cross-dataset performance of our method. The model trained by the ND-IRIS-0405 without fine-tuning is used to test on CASIA-Iris-M1-S3 and CASIA.v4-distance. The test dataset is the same as described in Section 4.3. As shown in Table 4, the experimental results are listed.

Table 4 Comparison of EER in the cross dataset performance evaluation

	Cross-dataset performance	
	CASIA.v4-distance	CASIA-Iris-M1-S3
VGG16	10.56%	7.68%
Pairwise CNN	8.34%	6.96%
UniNet	5.61%	4.75%
OFR	**4.32%**	**2.38%**

The experimental results prove that our method has the satisfying generalization ability, which has positive meaning for practical applications.

4.4 Experiments on periocular recognition

In this part, the experiments is to demonstrate the effectiveness of our method on public periocular datasets. The proposed method is tested on 2 public periocular datasets, CASIA-Iris-Distance and CASIA-Iris-M1-S3. In addition, to assess the IAAD performance, the CASIA-Iris-Distance dataset is employed. We follow the same datasets settings and pretraining settings, both for the iris recognition and periocular recognition as described above.

The OFR is compared with 3 DL based methods: AlexNet[57], VGG-16[50], ZhaoTIFS2017[34] and maxout CNNs[8]. As periocular recognition experiments were conducted in [8], the same experimental protocol as the maxout CNNs for AlexNet and VGG-16 is used for fairness. For the method proposed in Zhao-TIFS2017[34], they released NIR and VW frameworks and Caffe models trained on UBIpr and face and ocular challenge series (FOCS) datasets. The model needs additional supervision of left or right eye information and the gender information. The trained NIR Caffe model is directly employed. Furthermore, 2 algorithms based on hand-crafted feature are adopted: Gabor[55] and LBP[58].

As shown in Table 5 and Fig. 14, the comparison results and the DET curves are listed. Moreover, the comparison results with multiscale and single-scale features are shown in the Table 5 to prove the effectiveness of multiscale features.

Our method, which is based on the OFR, achieves superior performance on periocular recognition over other methods based on both deep features and hand-crafted features. In addition, the multiscale spatial feature can further improve the accuracy of recognition.

4.5 Experiments on fusion recognition

The experiments in this part test the performance of bimodal fusion. We also study the contribution of different fusion strategies.

The proposed fusion strategy is compared with the adaptive weights method and the score level fusion method

Table 5 Summary of EERs and FRRs at 0.1% FAR for the periocular recognition

	CASIA-Iris-M1-S3		CASIA.v4-distance	
	FRR at FAR = 0.1 %	EER	FRR at FAR = 0.1 %	EER
LBP	66.22%	29.41%	–	–
Gabor	57.91%	17.62%	49.31%	13.53%
TIFS17-SCNN	40.53%	12.66%	23.62%	6.61%
Maxout CNNs	9.50%	1.89%	9.21%	1.92%
AlexNet	11.52%	2.83%	11.94%	2.73%
Our method (single-scale)	**8.62%**	**2.59%**	**8.81%**	**2.72%**
Our method (multi-scale)	**6.91%**	**1.72%**	**8.01%**	**1.89%**

based on the weighted sum rule. We perform the fixed scorepreset score weights for iris w_i and periocular w_p as comparisons to manual score-level fusion methods. To observe the performance variations with iris and periocular weights changing, w_i is changed by 0.1 in the range of 0.3–0.7, and $w_p = 1 - w_i$. The results of score-level fusion by weighted sum are shown in Fig. 15. For example, 0.3I-0.7P score fusion means the weight of iris is 0.3 and the weight of periocular is 0.7.

Moreover, we also utilize another SOTA DL based iris quality assessment method[59] to obtain adaptive fusion weights and compare them with our proposed iris quality assessment method. The results are also shown in Fig. 15 and Table 6.

Then, we test feature-level fusion by concatenating the global feature set and spatial feature set of iris and periocular respectively. The similarity distance of concatenated feature set is calculated as (13) to decide the identity.

As the results show, weight plays a vital role to fusion result. Fusion performance is worse than iris recognition when the periocular modality has larger weights. Fusion performance is improved by a considerable margin when the iris has larger weights. However, iris recognition is much stricter than periocular recognition for image quality. It means that the larger iris weight is not always more suitable for all images. When the iris quality in the image is poor, giving more weight to periocular can greatly improve the recognition efficiency. In addition, the proposed iris quality assessment can provide a more suitable sum weight than the SOTA DL based iris quality assessment method. The proposed fusion strategy can significantly improve the recognition performance of a single ocular biometric.

5 Conclusions

In this paper, a new ocular biometrics recognition method is proposed, which is occlusion-robust, deformation-aware, and alignment-free for multi-modal ocular recognition. OFR is implanted in FCN to obtain suitable global features and multiscale features. Spatial feature reconstruction is leveraged to ensure the reconstruction error between a pair of images from different eyes is maximum and that between a pair of images from the same eye is minimum. Thus, the local feature at different positions and scales can be matched. The proposed score-level fusion strategy, based on unsupervised iris quality assessment, can enhance the recognition performance than unimodal ocular recognition significantly. The superb performance demonstrates the effectiveness and robustness of our method on both periocular and iris recognition, and it does not depend on high-quality probe images. In addition, excellent performance is achieved on both visible and NIR datasets. It implies that the proposed method can adapt to various application scenarios. Moreover, the generalization ability of our method is proved by the cross-dataset performance evaluation.

There are still some imperfections in the proposed method, e.g., the computational efficiency. These imperfections will be improved predictably in the future. On the whole, we propose a viable way to recognize low-quality images obtained from IAAD systems and will inspire

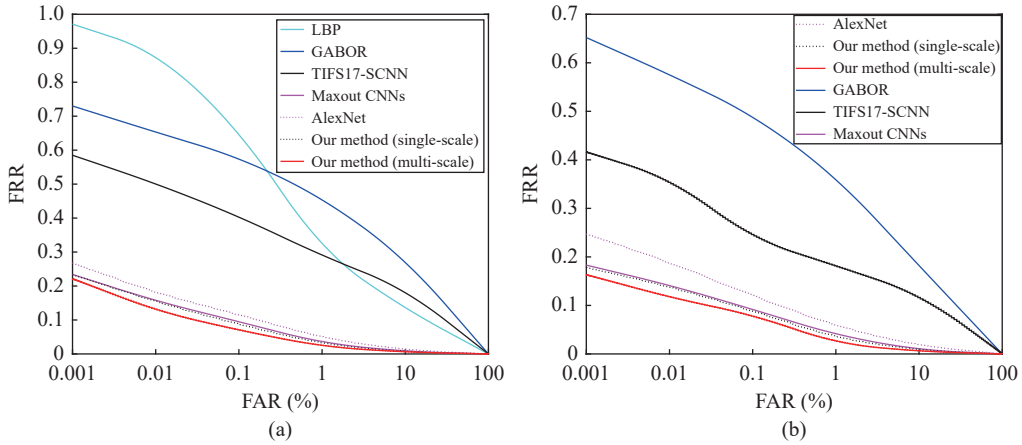


Fig. 14 DET curves on four public datasets for periocular recognition. (a) CASIA-Iris-M1-S3; (b) CASIA-Iris-Distance.

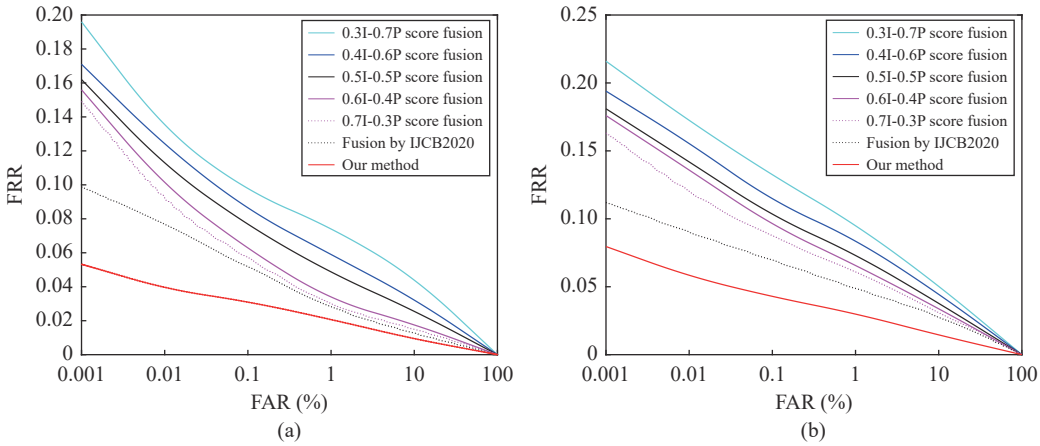


Fig. 15 Comparison of DET curves for fusion recognition. (a) CASIA-Iris-M1-S3; (b) CASIA-Iris-Distance.

Table 6 Summary of EERs and FRRs at 0.1% FAR for the fusion recognition

	CASIA-Iris-M1-S3		CASIA.v4-distance	
	FRR at FAR = 0.1 %	EER	FRR at FAR = 0.1 %	EER
Only iris	5.89%	1.63%	7.83%	2.17%
Only periocular	6.91%	1.72%	8.01%	1.89%
Feature-level fusion	5.32%	1.46%	7.86%	1.86%
Fusion by IJCB2020	5.13%	1.32%	6.92%	1.58%
Fusion by our method	**3.12%**	**0.93%**	**4.26%**	**1.39%**

other investigations on multi-modal ocular recognition in real-world scenarios.

Acknowledgements

This work was supported by National Natural Science Foundation of China (Nos. 62006225, 61906199 and 62071468), the Strategic Priority Research Program of Chinese Academy of Sciences (CAS), China (No. XDA 27040700), and sponsored by The Beijing Nova Program, China (Nos. Z201100006820050 and Z211100002121010).

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

[1] J. R. Matey, O. Naroditsky, K. Hanna, R. Kolczynski, D. J. LoIacono, S. Mangru, M. Tinker, T. M. Zappia, W. Y. Zhao. Iris on the move: Acquisition of images for iris recognition in less constrained environments. *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1936–1947, 2006. DOI: 10.1109/JPROC.2006.884091.

[2] S. Yoon, K. Bae, K. R. Park, J. Kim. Pan-tilt-zoom based Iris image capturing system for unconstrained user environments at a distance. In *Proceedings of the International Conference on Biometrics*, Springer, Seoul, Republic of Korea, pp. 653–662, 2007. DOI: 10.1007/978-3-540-74549-5_69.

[3] F. W. Wheeler, A. G. A. Perera, G. Abramovich, B. Yu, P. H. Tu. Stand-off Iris recognition system. In *Proceedings of the 2nd International Conference on Biometrics: Theory, Applications and Systems*, IEEE, Washington DC, USA, 2008. DOI: 10.1109/BTAS.2008.4699381.

[4] N. F. Liu, M. Zhang, H. Q. Li, Z. N. Sun, T. N. Tan. DeepIris: Learning pairwise filter bank for heterogeneous Iris verification. *Pattern Recognition Letters*, vol. 82, pp. 154–161, 2016. DOI: 10.1016/j.patrec.2015.09.016.

[5] A. Gangwar, A. Joshi. DeepIrisNet: Deep iris representation with applications in iris recognition and cross-sensor Iris recognition. In *Proceedings of IEEE International Conference on Image Processing*, Phoenix, USA, pp. 2301–2305, 2016. DOI: 10.1109/ICIP.2016.7532769.

[6] H. Proença, J. C. Neves. IRINA: Iris recognition (even) in inaccurately segmented data. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, pp. 6747–6756, 2017. DOI: 10.1109/CVPR.2017.714.

[7] Z. J. Zhao, A. Kumar. Towards more accurate iris recognition using deeply learned spatially corresponding features. In *Proceedings of IEEE International Conference on Computer Vision*, Venice, Italy, pp. 3829–3838, 2017. DOI: 10.1109/ICCV.2017.411.

[8] Q. Zhang, H. Q. Li, Z. N. Sun, T. N. Tan. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 11, pp. 2897–2912, 2018. DOI: 10.1109/TIFS.2018.2833033.

[9] M. Ren, Y. L. Wang, Z. N. Sun, T. N. Tan. Dynamic graph representation for occlusion handling in biometrics. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence*, New York, USA, pp. 11940–11947, 2020. DOI: 10.1609/aaai.v34i07.6869.

[10] L. X. He, H. Q. Li, Q. Zhang, Z. N. Sun. Dynamic feature learning for partial face recognition. In *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Salt Lake City, USA, pp. 7054–7063, 2018. DOI: 10.1109/CVPR.2018.00737.

[11] U. Park, A. Ross, A. K. Jain. Periocular biometrics in the visible spectrum: A feasibility study. In *Proceedings of the 3rd International Conference on Biometrics: Theory, Applications, and Systems*, IEEE, Washington DC, USA,

2009. DOI: 10.1109/BTAS.2009.5339068.

[12] C. W. Tan, A. Kumar. Towards online iris and periocular recognition under relaxed imaging constraints. *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 3751–3765, 2013. DOI: 10.1109/TIP.2013.2260165.

[13] L. H. Xiao, Z. N. Sun, T. N. Tan. Fusion of iris and periocular biometrics for cross-sensor identification. In: *Proceedings of the 7th Chinese Conference on Biometric Recognition*, Springer, Guangzhou, China, pp. 202–209, 2012. DOI: 10.1007/978-3-642-35136-5_25.

[14] K. Nguyen, C. Fookes, R. Jillela, S. Sridharan, A. Ross. Long range iris recognition: A survey. *Pattern Recognition*, vol. 72, pp. 123–143, 2017. DOI: 10.1016/j.patcog.2017.05.021.

[15] A. Joshi, A. K. Gangwar, Z. Saquib. Person recognition based on fusion of iris and periocular biometrics. In *Proceedings of the 12th International Conference on Hybrid Intelligent Systems*, IEEE, Pune, India, pp. 57–62, 2012. DOI: 10.1109/HIS.2012.6421309.

[16] P. Terhörst, J. N. Kolf, N. Damer, F. Kirchbuchner, A. Kuijper. SER-FIQ: Unsupervised estimation of face image quality based on stochastic embedding robustness. In: *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Seattle, USA, pp. 5650–5659, 2020. DOI: 10.1109/CVPR42600.2020.00569.

[17] L. Best-Rowden, A. K. Jain. Learning face image quality from human assessments. *IEEE Transactions on Information Forensics and Security*, vol. 13, no. 12, pp. 3064–3077, 2018. DOI: 10.1109/TIFS.2018.2799585.

[18] J. Hernandez-Ortega, J. Galbally, J. Fierrez, R. Haraksim, L. Beslay. FaceQnet: Quality assessment for face recognition based on deep learning. In: *Proceedings of International Conference on Biometrics*, IEEE, Crete, Greece, pp. 1–8, 2019. DOI: 10.1109/ICB45273.2019.8987255.

[19] P. J. Phillips, J. R. Beveridge, D. S. Bolme, B. A. Draper, G. H. Givens, Y. M. Lui, S. Cheng, M. N. Teli, H. Zhang. On the existence of face quality measures. In *Proceedings of the 6th International Conference on Biometrics: Theory, Applications and Systems*, IEEE, Arlington, USA, 2013. DOI: 10.1109/BTAS.2013.6712715.

[20] X. F. Gao, S. Z. Li, R. Liu, P. R. Zhang. Standardization of face image sample quality. In *Proceedings of the International Conference on Biometrics*, Springer, Seoul, Republic of Korea, pp. 242–251, 2007. DOI: 10.1007/978-3-540-74549-5_26.

[21] J. Daugman. How iris recognition works. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 14, no. 1, pp. 21–30, 2004. DOI: 10.1109/TCSVT.2003.818350.

[22] L. Masek. Recognition of Human Iris Patterns for Biometric Identification. Ph. D. dissertation, University of Western Australia, Australia, 2003.

[23] Z. N. Sun, T. N. Tan. Ordinal measures for iris recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 12, pp. 2211–2226, 2009. DOI: 10.1109/TPAMI.2008.240.

[24] K. Wang, A. Kumar. Toward more accurate iris recognition using dilated residual features. *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 12, pp. 3233–3245, 2019. DOI: 10.1109/TIFS.2019.2913234.

[25] K. Nguyen, C. Fookes, S. Sridharan. Constrained design of deep iris networks. *IEEE Transactions on Image Processing*, vol. 29, pp. 7166–7175, 2020. DOI: 10.1109/TIP.2020.2999211.

[26] J. Z. Wei, Y. L. Wang, Y. Li, R. He, Z. N. Sun. Cross-spectral Iris recognition by learning device-specific band. *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 6, pp. 3810–3824, 2022. DOI: 10.1109/TCSVT.2021.3117291.

[27] J. Z. Wei, H. B. Huang, Y. L. Wang, R. He, Z. N. Sun. Towards more discriminative and robust iris recognition by learning uncertain factors. *IEEE Transactions on Information Forensics and Security*, vol. 17, pp. 865–879, 2022. DOI: 10.1109/TIFS.2022.3154240.

[28] K. Nguyen, C. Fookes, S. Sridharan, A. Ross. Complex-valued Iris recognition network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 1, pp. 182–196, 2023. DOI: 10.1109/TPAMI.2022.3152857.

[29] K. Nguyen, H. Proença, F. Alonso-Fernandez. Deep learning for Iris recognition: A survey, [Online], Available: https://arxiv.org/abs/2210.05866, 2022.

[30] K. B. Zhang, Z. T. Shen, Y. L. Wang, Z. N. Sun. All-in-focus Iris camera with a great capture volume. In *Proceedings of IEEE International Joint Conference on Biometrics*, Houston, USA, 2020. DOI: 10.1109/IJCB48548.2020.9304932.

[31] Z. H. Yan, L. X. He, Y. L. Wang, Z. N. Sun, T. N. Tan. Flexible iris matching based on spatial feature reconstruction. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, to be published. DOI: 10.1109/TBIOM.2021.3108559.

[32] F. Alonso-Fernandez, J. Bigun. A survey on periocular biometrics research. *Pattern Recognition Letters*, vol. 82, pp. 92–105, 2016. DOI: 10.1016/j.patrec.2015.08.026.

[33] C. Kandaswamy, J. C. Monteiro, L. M. Silva, J. S. Cardoso. Multi-source deep transfer learning for cross-sensor biometrics. *Neural Computing and Applications*, vol. 28, no. 9, pp. 2461–2475, 2017. DOI: 10.1007/s00521-016-2325-5.

[34] Z. J. Zhao, A. Kumar. Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 5, pp. 1017–1030, 2017. DOI: 10.1109/TIFS.2016.2636093.

[35] K. B. Raja, R. Raghavendra, C. Busch. Smartphone based robust iris recognition in visible spectrum using clustered K-means features. In *Proceedings of IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications*, Rome, Italy, pp. 15–21, 2014. DOI: 10.1109/BIOMS.2014.6951530.

[36] K. B. Raja, R. Raghavendra, C. Busch. Dynamic scale selected Laplacian decomposed frequency response for cross-smartphone periocular verification in visible spectrum. In *Proceedings of the 19th International Conference on Information Fusion*, IEEE, Heidelberg, Germany, pp. 2206–2212, 2016.

[37] R. Raghavendra, C. Busch. Learning deeply coupled autoencoders for smartphone based robust periocular verification. In *Proceedings of IEEE International Conference on Image Processing*, Phoenix, USA, pp. 325–329, 2016. DOI: 10.1109/ICIP.2016.7532372.

[38] K. B. Raja, R. Raghavendra, C. Busch. Collaborative representation of deep sparse filtered features for robust verification of smartphone periocular images. In *Proceedings of IEEE International Conference on Image Processing*, Phoenix, USA, pp. 330–334, 2016. DOI: 10.1109/ICIP.2016.7532373.

[39] A. Rattani, R. Derakhshani, S. K. Saripalle, V. Gottemukkula. ICIP 2016 competition on mobile ocular biometric recognition. In *Proceedings of IEEE International Conference on Image Processing*, Phoenix, USA, pp. 320–324, 2016. DOI: 10.1109/ICIP.2016.7532371.

[40] B. J. Yin, L. Tran, H. X. Li, X. H. Shen, X. M. Liu. Towards interpretable face recognition. In *Proceedings of IEEE/CVF International Conference on Computer Vision*, IEEE, Seoul, Republic of Korea, pp. 9347–9356, 2019. DOI: 10.1109/ICCV.2019.00944.

[41] Z. X. Huang, Y. Li. Interpretable and accurate fine-grained recognition via region grouping. In: *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition*, IEEE, Seattle, USA, pp. 8659–8669, 2020. DOI: 10.1109/CVPR42600.2020.00869.

[42] L. C. O. Tiong, S. T. Kim, Y. M. Ro. Multimodal facial biometrics recognition: Dual-stream convolutional neural networks with multi-feature fusion layers. *Image and Vision Computing*, vol. 102, Article number 103977, 2020. DOI: 10.1016/j.imavis.2020.103977.

[43] D. L. Woodard, S. Pundlik, P. Miller, R. Jillela, A. Ross. On the fusion of periocular and Iris biometrics in non-ideal imagery. In *Proceedings of the 20th International Conference on Pattern Recognition*, IEEE, Istanbul, Turkey, pp. 201–204, 2010. DOI: 10.1109/ICPR.2010.58.

[44] K. B. Raja, R. Raghavendra, M. Stokkenes, C. Busch. Multi-modal authentication system for smartphones using face, Iris and periocular. In *Proceedings of International Conference on Biometrics*, IEEE, Phuket, Thailand, pp. 143–150, 2015. DOI: 10.1109/ICB.2015.7139044.

[45] G. Santos, E. Hoyle. A fusion approach to unconstrained Iris recognition. *Pattern Recognition Letters*, vol. 33, no. 8, pp. 984–990, 2012. DOI: 10.1016/j.patrec.2011.08.017.

[46] Z. D. Luo, J. T. Li, Y. S. Zhu. A deep feature fusion network based on multiple attention mechanisms for joint Iris-periocular biometric recognition. *IEEE Signal Processing Letters*, vol. 28, pp. 1060–1064, 2021. DOI: 10.1109/LSP.2021.3079850.

[47] K. Wang, A. Kumar. Periocular-assisted multi-feature collaboration for dynamic iris recognition. *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 866–879, 2021. DOI: 10.1109/TIFS.2020.3023289.

[48] K. Veeramachaneni, L. A. Osadciw, P. K. Varshney. An adaptive multimodal biometric management algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 35, no. 3, pp. 344–356, 2005. DOI: 10.1109/TSMCC.2005.848191.

[49] J. Deng, W. Dong, R. Socher, L. J. Li, K. Li, L. Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Miami, USA, pp. 248–255, 2009. DOI: 10.1109/CVPR.2009.5206848.

[50] K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition, [Online], Available: https://arxiv.org/abs/1409.1556, 2014.

[51] K. M. He, X. Y. Zhang, S. Q. Ren, J. Sun. Deep residual learning for image recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 770–778, 2016. DOI: 10.1109/CVPR.2016.90.

[52] H. Lee, A. Battle, R. Raina, A. Y. Ng. Efficient sparse coding algorithms. In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, Vancouver, Canada, pp. 801–808, 2006.

[53] C. Y. Wang, J. Muhammad, Y. L. Wang, Z. F. He, Z. N. Sun. Towards complete and accurate Iris segmentation using deep multi-task attention network for non-cooperative Iris recognition. *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2944–2959, 2020. DOI: 10.1109/TIFS.2020.2980791.

[54] J. G. Daugman. High confidence visual recognition of persons by a test of statistical independence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 11, pp. 1148–1161, 1993. DOI: 10.1109/34.244676.

[55] K. W. Bowyer, P. J. Flynn. The ND-IRIS-0405 Iris image dataset, [Online], Available: https://arxiv.org/abs/1606.04853, 2016.

[56] H. Proenca, S. Filipe, R. Santos, J. Oliveira, L. A. Alexandre. The UBIRIS.v2: A database of visible wavelength Iris images captured on-the-move and at-a-distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 8, pp. 1529–1535, 2010. DOI: 10.1109/TPAMI.2009.66.

[57] A. Krizhevsky, I. Sutskever, G. E. Hinton. ImageNet classification with deep convolutional neural networks. In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Lake Tahoe, USA, vol. 1, pp. 1097–1105, 2012.

[58] Y. Hu, K. Sirlantzis, G. Howells. Optimal generation of iris codes for Iris recognition. *IEEE Transactions on Information Forensics and Security*, vol. 12, no. 1, pp. 157–171, 2017. DOI: 10.1109/TIFS.2016.2606083.

[59] L. Y. Wang, K. B. Zhang, M. Ren, Y. L. Wang, Z. N. Sun. Recognition oriented iris image quality assessment in the feature space. In *Proceedings of IEEE International Joint Conference on Biometrics*, Houston, USA, 2020. DOI: 10.1109/IJCB48548.2020.9304896.

Zihui Yan received the B. Eng. degree in automation from Tsinghua University, China in 2015. He is currently the Ph. D. degree candidate in engineering at School of Artificial Intelligence, University of Chinese Academy of Sciences, China, and Center for Research on Intelligent Perception and Computing, National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, China.

His research interests include biometrics and deep learning.

E-mail: zihui.yan@cripac.ia.ac.cn

ORCID iD: 0000-0002-6276-6189

Yunlong Wang received the B. Eng. and M. Sc. degrees in automation from University of Science and Technology of China, China. He is currently an associate professor with Center for Research on Intelligent Perception and Computing, (CRIPAC), National Laboratory of Pattern Recognition, (NLPR), Institate of Avtomation, Chinese Acadeny of Sciences (CASIA), China.

His research interests include pattern recognition, machine learning, light field photography and biometrics.

E-mail: yunlong.wang@cripac.ia.ac.cn

ORCID iD: 0000-0002-3535-308X

Kunbo Zhang received the B. Eng. degree in automation from Beijing Institute of Technology, China in 2006, the M. Sc. and Ph. D. degrees in mechanical engineering from State University of New York at Stony Brook, USA in 2008 and 2011, respectively. From 2011 to 2016, he worked as the advanced manufacturing engineer in Nexteer Automotive. He is currently an associate professor with CRIPAC, NLPR, CASIA, China.

His research interests include computational photography, biometric recognition and intelligent systems.

E-mail: kunbo.zhang@cripac.ia.ac.cn

Zhenan Sun received the B. Eng. degree in industrial automation from Dalian University of Technology, China, in 1999, the M. Sc. degree in system engineering from Huazhong University of Science and Technology, China, in 2002, and the Ph. D. degree in pattern recognition and intelligent systems from Institute of Automation, Chinese Academy of Sciences, China in 2006. He is currently a professor in CRIPAC, NLPR, CASIA, China and CAS Center for Excellence in Brain Science and Intelligence Technology School of Artificial Intelligence, University of Chinese Academy of Sciences, China. He is an Associate Editor of IEEE TIFS and *IEEE Biometrics Compendium*. He is a Fellow of IEEE IAPR.

His research interests include biometrics and pattern recognition.

E-mail: znsun@nlpr.ia.ac.cn (Corresponding author)
ORCID iD: 0000-0003-4029-9935

Lingxiao He received the B. Eng. degree in information engineering from Chengdu University of Technology, China in 2014, and the Ph. D. degree from National Laboratory of Pattern Recognition, Center for Research on Intelligent Perception and Computing, Institute of Automation, Chinese Academy of Sciences, China in 2019. He is currently a research scientist with JD AI Research.

His research interests include biometrics, computer vision, machine learning, and deep learning, robotics, feedback control systems and control theory.

E-mail: helingxiao3@jd.com

Machine Intelligence Research

Citation: Z. Yan, Y. Wang, K. Zhang, Z. Sun, L. He. Boosting multi–modal ocular recognition via spatial feature reconstruction and unsupervised image quality estimation. *Machine Inteligence Research*, vol.21, no.1, pp.197–214, 2024. https://doi.org/10.1007/s11633–023–1415–y