

# Person Identification using Fusion of Iris and Periocular Deep Features

Saiyed Umer · Alamgir Sardar · Bibhas Chandra Dhara · Ranjeet Kumar Raout · Hari Mohan Pandey

Received: DD Month YEAR / Accepted: DD Month YEAR

**Abstract** A novel method for person identification based on the fusion of iris and periocular biometrics has been proposed in this paper. The challenges for image acquisition for Near-Infrared or Visual Wavelength lights under constrained and unconstrained environments have been considered here. The proposed system is divided into image preprocessing data augmentation followed by feature learning for classification components. In image preprocessing an annular iris, the portion is segmented out from an eyeball image and then transformed into a fixed-sized image region. The parameters of iris localization have been used to extract the local periocular region. Due to different imaging environments, the images suffer from various noise artifacts which create data insufficiency and complicates the recognition task. To overcome this situation a novel method for data augmentation technique has been introduced here. For features extraction and classification tasks well-known VGG16, ResNet50, and Inception-v3 CNN ar-

chitectures have been employed. The performance due to iris and periocular are fused together to increase the performance of the recognition system. The extensive experimental results have been demonstrated in four benchmark iris databases namely: MMU1, UPOL, CASIA-Iris-distance, and UBIRIS.v2. The comparison with the state-of-the-art methods with respect to these databases shows the robustness and effectiveness of the proposed approach.

**Keywords** Person Identification · Iris recognition · periocular recognition · Data augmentation · Deep learning · Rank-level fusion

## 1 Introduction

Biometric plays a vital role to identify a person based on his/her physiological and behavioral traits belongs to each individual [1]. There are various biometric traits available which are used in different applications for different purposes. Each biometric trait has its own weaknesses and strengths, but the choice of these biometric traits depends on the different purposes of the applications [1]. No particular biometric may successfully meet the requirements for all the applications. Though among the various biometric traits, the iris is unique to every individual and it exhibits more discriminating patterns for which the recognition system has great performance compared to other biometric traits [2]. In the iris recognition system generally, the images are captured either in near-infrared (NIR)/visual wavelength (VW) lighting conditions. According to Proença et al. [1], NIR images captured in constrained environments contain the increasing amount of texture information and it is due to the specular reflections reduced on the iris portion with mostly free from illumination.

---

Saiyed Umer  
Department of Computer Science & Engineering  
Aliah University, Kolkata, India  
E-mail: saiyedumer@gmail.com

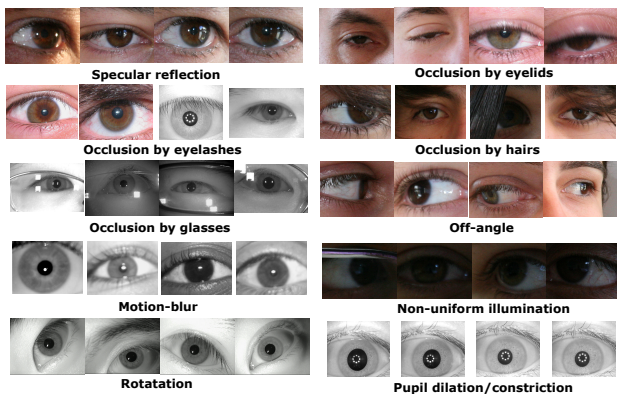
Alamgir Sardar  
Department of Computer Science & Engineering  
Aliah University, Kolkata, India  
E-mail: saiyedumer@gmail.com

Bibhas Chandra Dhara  
Department of Information Technology  
Jadavpur University, Kolkata, India

Ranjeet Kumar Raout  
Department of Computer Science and Engineering  
National Institute of Technology, Srinagar, India

Hari Mohan Pandey  
Department of Computer science  
Edge Hill University, Ormskirk, Lancashire, UK

VW images captured in constrained environments depict higher details of pigmentation on iris pattern than NIR images. The images captured in NIR/VW light in the unconstrained environments (i.e. subjects are at varying distances, subjects are moving), suffer from various noise artifacts such as shadows, specular reflections, occlusion by the eyelid, eyelashes, hair, and eye-glasses, off-angle, motion-blur, rotated and non-uniform illumination [3] [4]. The NIR imaging reduces specular reflections and physical discomfort of illumination while increases the texture information whereas the VW image captures the iris pigmentation in much higher details but also introduces the above-discussed noise artifacts [2]. The images affected by these noises make the recognition system more challenging. Some examples of these noises are demonstrated in Fig 1.



**Fig. 1** Noise artifacts considered for the proposed system.

To detract these issues, considering a small region called periocular [5] for recognition system, achieves better performance than the iris recognition system under the following circumstances: (i) eye blinking and off-angle pose creates noises during image acquisition, (ii) images captured at varying distances while subjects are moving results motion blur and distortion artifacts which makes impossible extraction of discriminating texture information, (iii) capturing the entire face image at a distance results the low-resolution eyeball image and at this situation periocular region plays as a trade-off between the face and iris recognition systems, (iv) acquisition of the periocular region requires less subject cooperation compared to the iris region, (v) Periocular is least invasive among the other eyeball based recognition system. Hence the noises degrade the image quality and make the recognition system more challenging. Some examples of periocular images are shown in Fig.2.



**Fig. 2** Periocular images considered for the proposed system.

The periocular recognition system performs better but its performance is not stable and it is due to the illumination and blurriness artifacts, results in less texture information [6]. Compared to the iris, periocular texture information remains stable because it is internally protected and it contains relevant texture information even it suffers from blurriness and illumination problems. Considering these challenges the researchers had proposed their iris or periocular recognition system by utilizing some reliable segmentation approaches, efficient feature extraction methods and various classifiers or ensemble of classifiers for recognizing the subjects [7].

There are various biometric recognition systems developed by utilizing different sets of features and different classifiers. Many of them used more or less costly methods for iris pre-processing task such as in [8], [9], [10], [11], [12], [13], [14] and [15]. The features extracted from these iris or periocular should contain a set of divergent and competent features and to characterize those features there exist several methods but among them transformed, structural and statistical based approaches are mostly used in practice. Wavelets [15] [11], Gabor filter banks [16], [17], Discrete Fourier Transform (DFT) [18], Discrete Cosine Transform (DCT) coefficients [9] are widely used during feature computation from iris and periocular regions. Structural based approaches such as mathematical morphology [14] and fractal dimensions, are used for feature computation. For statistical-based approaches bag-of-words [19], Sparse Representation [20] [13], Locality-constrained Linear Coding (LLC) [21] and Hierarchical Visual Codebook (HVC) [22] representations have been used for feature computation. A hybridization of transformed, structural and statistical based approaches i.e. deep convolution neural networks (CNNs) [7] [5] have also been employed for building iris and periocular recognition systems. An extensive literature survey for iris/periocular biometrics has been presented in Table 1 with respect to im-

age preprocessing, feature extraction and classification components.

**Table 1** Some existing iris or periocular recognition systems.

Method	Pre-processing	Feature extraction	Classifier
Monoro [9]	Threshold,CHT	DCT	HD
Miyazawa [18]	Threshold	DFT	BLPOC
Hollingsworth [10]	IDO	LogG	HD
Dong [23]	Adaboost	EM,TOF	Weight map
Paucaça [24]	Masek,IDO	Gabor,SIFT	kNN
Reyes-Lopez [25]	LoG,LP	ZM	SVM
Pillai [13]	IDO	Gabor	SR,BF
Costa [15]	Threshold,CHT	DF	ED
Sun [22]	PP,EF	SIFT,HVC	SVM
Tan [26]	RW,CF	LogG,SIFT,LMF,GIST	HD, $\chi^2$
Tan [4]	IDO	logG,ZM	HD
Proenca [1]	EHT	Gabor	Sigmoid, $\ell_2$ -norm
Umer [14]	RCHT	MM	RBF-SVM
Umer [27]	CCHT	OTCM	RBF-SVM
Zhang [28]	CNN	CNN	Fusion
Proenca [29]	CNN	Data Augmentation	Fusion

HD (Hamming distance),CNN (Convolution Neural Network)  
 DF (Dynamic Features), MM (Multiscale morphology)  
 DCT (Discrete cosine transform), EM (Elastic model)  
 ED (Euclidean distance), LogG (Log-Gabor)  
 PP (Pulling-pushing), RW (Random-Walker)  
 LP (Laplacian pyramid),CF (Curve-Fitting)  
 TOF (Trilobe ordinal feature), ZM (Zernike moments)  
 EF (Edge-Fitting), OTCM (Ordered Texture code matrix)  
 LSVMF (Linear SVM+fusion), BF (Bayesian-Fusion)

In real-time situations, the iris portion may not be detected properly due to blurring, occlusion, and poor imaging protocols. So, a multimodal biometric recognition system has been proposed using both iris and periocular biometrics extracted from an eyeball image. This recognition system is able to handle any situations with any challenges as discussed in the literature. The proposed system is expected to process an input image captured in NIR/VW light under constrained/unconstrained environments. The contributions of the proposed system are as follows:

- The image  $\mathcal{E}$  captured either in NIR/VW light with different image qualities suffers from various noise artifacts, so, considering these challenges, some efficient iris localization approaches for NIR and VW images have been employed and then using the parameters of the employed iris localization techniques, the local periocular region  $\mathcal{P}$  has been extracted from  $\mathcal{E}$ .
- The acquisitions of eyeball images are subjects cooperative in nature. Capturing multiple samples for each individual under similar circumstances at a time is difficult. Moreover, the captured samples suffer from noise artifacts that are unable to process further. Hence it incurs the data insufficiency problem for which the classifiers are not trained well and results in lower performance. To overcome these situations we have employed some novel data augmentation [30] techniques applied on training samples.

In the proposed data augmentation techniques, each sample is replicated to several samples by applying filtering and affine transformation techniques and it results in the augmented samples which carry variants of complex texture information.

- To investigate more discriminating features from the iris/periocular region, pre-trained deep convolutional neural network (CNN) models have been employed to distinguish dissimilar images from different subjects and to rapport the images from the same subject. For learning the CNN model, pre-trained, transfer learning, and fine-tuning approaches have also been deputed for the proposed system.
- To control the trade-off between errors due to the iris and periocular recognition system and to increase the performance of the proposed recognition system, the individual performance due to iris and periocular biometrics, have been fused together by the rank-level post-classification fusion technique to identify the unknown subject.
- Both NIR and VW images under constrained and unconstrained imaging environments have been considered here and good performance has been achieved on some public domain benchmark databases such as UPOL, MMU1, CASIA-Iris-Distance, and UBIRIS.v2 database.

The rest of this paper is organized as follows: Section 2 describes some basic methodology where the basic convolutional neural network (CNN) concepts and different CNN models have been discussed along with various fusion techniques used in the biometric recognition system. Section 3 provides the framework for the proposed system, with a detailed description of each module. Experimental results and discussions are reported in Section 4. Section 5 concludes this paper.

## 2 Basic Methodology

### 2.1 Convolutional Neural Network (CNN)

The CNN has (a) Convolution layer [31] which is the core building block of CNN that performs most of the computational heavy lifting and the parameters of this layer consist of a set of filter banks (kernels) while extracting features with increasing complexity. In this layer, the input image is convolved with filter banks (kernels) and then compute the dot product between the entries of the filters and the input and produces the feature-maps for that corresponding filter kernels. As a result, the network learns filters that activate when it detects some specific type of feature at the same spatial position in the input. (b) Maxpooling layer is inserted

to reduce the spatial size of the representation and also to reduce the computation overheads by decreasing the number of parameters in the network. This layer operates with filters of size  $2 \times 2$  applied with a stride of 2 downsamples every depth slice in the input feature map by 2 along both width and height by deciding a max value over 4 numbers. This layer also controls the over-fitting problem. (c) A fully connected layer considers all the features to obtain the information about the overall shape of the image and finally, it generates a probability score in the last layer over the number of classes for which the network is trained.

There are some well known CNN models which have gained great success in the field of computer vision research areas. These CNN models are as follows:

- *Very deep convolution network (VGG):-* This model is proposed by Simonyan and Zisserman [32]. During training, the input to this model is a fixed-size  $224 \times 224$  three-channel image. The images are passed through a stack of convolutional (Conv) layers, where small receptive filters of size  $3 \times 3$  (which is the smallest size to capture the notion of left/right, up/down, center) are used. Further  $1 \times 1$  convolution filters are also utilized where a linear transformation of input channels followed by non-linearity is used. To preserve the spatial resolution after convolution, the padding of 1 pixel for  $3 \times 3$  Conv. layers is employed. Max-pooling over  $2 \times 2$  pixel window is performed with stride 2. With different depth in different architectures, a stack of Conv layers followed by three fully-connected (FC) layers have been utilized such as the first two FC layers have 4096 channels and the third FC layer has 1000 channels. The third FC layer performs the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) classification. The final layer of this model is the soft-max layer. In all the architectures of this model, the FC layers are the same. Here VGG16 CNN architecture has been employed. Fig. 3 shows the configuration layers of the VGG16 network.

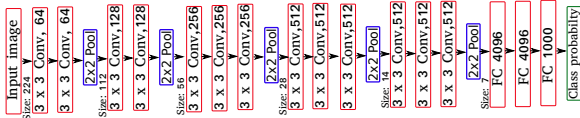


Fig. 3 VGG16 CNN architecture.

- *Deep Residual Network Architectures (ResNet):-* This model is proposed by He et al. [33]. ResNet model is based on the VGG nets. Here also the convolutional layers have  $3 \times 3$  filters and they follow some simple designs such as: (i) for the layers having the same

number of filters have the same output, (ii) the number of filters is doubled if the convolved output size is halved such that the time complexity per layer is preserved. The model ends with an average pooling layer and a 1000-way fully-connected layer with softmax. The number of weighted layers is 50 here. This model has fewer filters and lower complexity than VGG nets. Fig.4 shows the configuration layers of ResNet50 network.

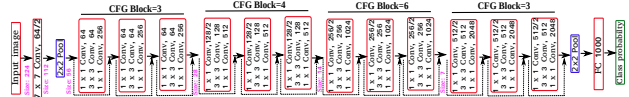


Fig. 4 Resnet50 CNN architecture.

- *Inception-v3 (GoogLeNet):-* This model is introduced by Szegedy et al. [34, 35]. This model uses a multi-scale approach by using multiple classifier structures while combined with multiple sources for back-propagation. Without causing penalties Inception-v3 model increases both the width and depth of the network. In this model, multiple Inception layers are applied in convolution on the input feature map in different scales for allowing to make more complex decisions. Fig. 5 shows the configuration layers of Inception-v3 network.

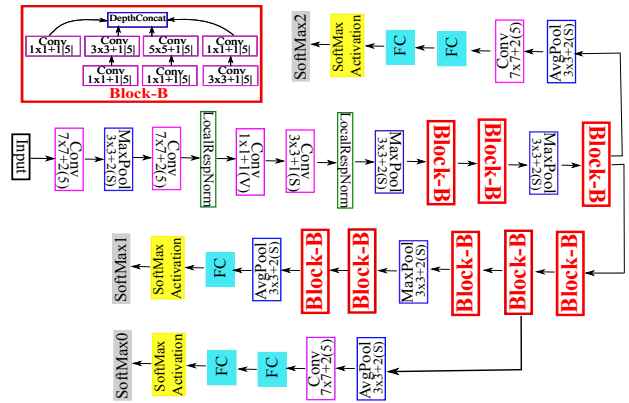


Fig. 5 Inception-v3 CNN architecture.

## 2.2 Biometric Fusion

The information from every single source of biometric is combined together to improve the performance of the biometric recognition system. This combining technique is known as biometric fusion. In the biometric recognition system there are two types of fusion techniques: (a)

pre-classification and (b) post-classification fusion techniques. In the pre-classification fusion technique, the biometric information are fused before the classification task and the fusion techniques under this category are sensor-level and feature-level fusion techniques. The performance in the pre-classification fusion technique suffers from the data redundancy, noise introduction, and multi-environment image acquisition problems. In the post-classification fusion technique, the biometric information is fused after classification task based on the scores obtained by the classifiers or matchers. The performance in the post-classification fusion technique is free from noises and also improves the recognition performance to some extent. For the proposed system we have employed the post-classification fusion techniques where the scores are fused after matching. In the post-classification fusion technique the fusion levels are: (a) Decision-level, (b) Score-level and (c) Rank-level.

- In the decision level fusion technique, a decision in terms of Boolean is the output from every biometric modality which is combined based on the majority voting (or OR-ing technique) to make the final decision for the recognition system. It works only for a biometric verification system.
- In the score level fusion technique, a matching score is obtained based on measuring the similarity between the stored template and the query template from each biometric system. Then the individual scores are combined using *sum-rule* or *product rule* based techniques to derive a final score that undergoes a decision module system for recognizing a subject. It works for both the biometric verification and identification system.
- In the rank level fusion technique, the scores obtained from every biometric system are arranged in decreasing order and then the rank is assigned to each sorted score. The highest score gets the minimum rank. In the multimodal biometric system, the scores from different modalities are converted to ranks. Then rank-level fusion techniques [36] are applied on these ranks to obtain the final identification performance in terms of correct recognition rates (CRR) rates. There are three different rank-level fusion techniques: (i) Highest Ranking (HR), (ii) Borda Count (BC) and (iii) Logistic Regression (LR).

We assume that for an identification system there are  $\mathcal{M}$  different subjects enrolled as templates with  $\mathcal{N}$  different biometric systems or matchers.  $r_{ki}$  be a rank assigned to  $k_{th}$  subject ( $k \in \{1, \dots, \mathcal{M}\}$ ) by the  $i_{th}$  matcher ( $i \in \{1, \dots, \mathcal{N}\}$ ).

1. In the highest-ranking (HR) method, the subject is identified due to the highest rank assigned by the biometric system or matcher. So, according to HR technique the rank for  $k_{th}$  subject by  $i_{th}$  matcher will be given by  $r_k = \min(r_{ki}, r_{kj})$ , where  $r_k$  be the fused ranks which are compared with the actual ranks for *rank 1* position of the  $k_{th}$  subject and the exact number of matching for *rank 1* position derives the correct recognition rate of the biometric system.
2. The Borda Count (BC) is an unsupervised rank fusion technique where the fusion is being performed for  $k_{th}$  subject by  $i_{th}$  matcher, will be given by  $r_k = r_{ki} + r_{kj}$ . Then the fused ranks are rearranged in sorted order and the comparisons have been performed with their actual *rank 1* position. The number of matches for *rank 1* position is the correct recognition of the biometric system.
3. The Logistic Regression (LR) is the weighted BC method where the fusion is being performed for  $k_{th}$  subject by  $i_{th}$  matcher using  $r_k = w_i * r_{ki} + w_j * r_{kj}$ ,  $w_i + w_j = 1$ , while  $w_i$  and  $w_j$  be the weights assigned to the  $i_{th}$  and  $j_{th}$  biometric system respectively.

Here to train these CNN architectures we have employed the technique of fine tuning [35]. For both iris and periocular, the above CNN architectures are trained individually. Then these CNNs are tested for iris and periocular test-dataset accordingly. The network which has better performance for both the biometrics is chosen for the proposed system. Since the biometric samples suffer from various noise artifacts and the quality level also decreases. So, to control the trade-off between the errors due to the iris and periocular recognition systems, the scores from these individual biometric systems are fused together using rank-level fusion techniques to obtain a final decision for person identification.



### 3 PROPOSED METHOD

The proposed biometric recognition system has three components: “*image pre-processing*”, “*data augmentation*” and “*feature learning for classification*”. The data flow diagram of the proposed system, is demonstrated at Fig.6.

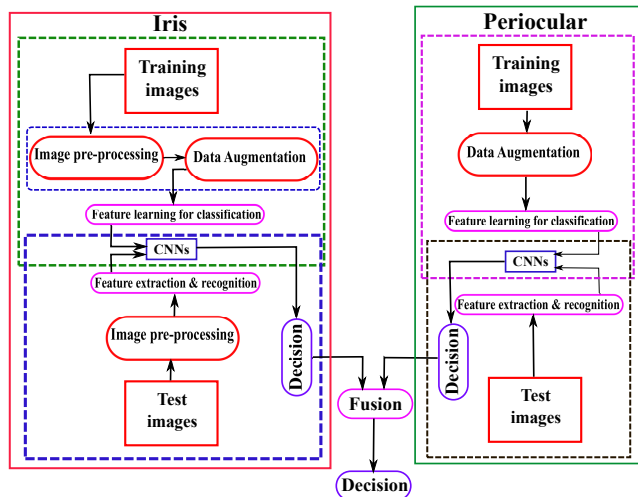


Fig. 6 The block diagram of the proposed system.

#### 3.1 Preprocessing

The acquisition of the eyeball images suffers from various noise artifacts and it is due to the variations in imperfect data acquisition conditions. The noises in the eyeball images are demonstrated in Fig. 1. These noises result from varieties in image quality that leads to inaccurate segmentation of the iris portion within a real-time situation. The accurate iris localization plays a vital role during iris recognition and this iris localization is the process of segmenting an annular iris portion from an eyeball image  $\mathcal{E}$ . The iris normalization is the process of transforming the localized iris portion into a rectangular array of a given size. The combined process of iris localization and normalization is iris pre-processing. Here iris normalization is being performed using Daugman’s rubber sheet model [8].

There are several studies of image acquisition based on the imaging sensors used either in near-infrared (NIR) and/or visible wavelength (VW) lights under constrained and unconstrained environments. The NIR images captured under constrained environment enable the blocking of corneal specular reflection while the spectral radiance of the iris due to its pigmentation levels varies more significantly in VW than in NIR [2]. Hence due

to the different nature of lighting conditions, for NIR images first pupil detection followed by iris detection is performed whereas iris detection followed by pupil detection is employed by VW images. For iris localization we have employed the methods from [37] but most of the parameters have been relaxed here to make the localization process more convenient.

An input eyeball image  $\mathcal{E}$  (Fig. 7(a)) is transformed to gray-scaled image which undergoes to median filtering to suppress the noise (Fig. 7(b)). A histogram analysis from the smoothed image is being performed. The pupil is the darkest region of the eyeball image, the gray-level histogram of this eyeball image corresponds to the peak at the lowest possible gray-level. Select this intensity value as a threshold ‘ $t$ ’ to convert the image  $\mathcal{E}$  to a binary image  $\mathcal{B}$  (Fig.7(c)). Now the morphological boundary extraction algorithm is applied on  $\mathcal{B}$  to get boundary pixels for  $\mathcal{B}$  (Fig.7(d)). To remove the small components from  $\mathcal{B}$ , some morphological area operators have been applied on  $\mathcal{B}$  and then circular Hough transformation (CHT) based technique is applied on boundary pixels of  $\mathcal{B}$  to get pupil center and radius for  $\mathcal{E}$  (Fig.7(e)). Considering this pupil boundary circle be  $C_P$  and its radius be  $r_P$ .

Now for the iris boundary detection process, the intensity variation profile is considered horizontally and then vertically with respect to pupil center  $C_P$  and radius  $r_P$  and obtain the high-intensity variation after pupil region between iris and sclera region. Considering a circle at this high-intensity change between iris and sclera region be  $C_M$  and its radius  $r_M$ . Then  $r_R = \frac{(r_P+r_M)}{2}$ , is the radius for a reference circle (Fig. 7(f)). The inversion transformation technique is applied to the pixels between ‘ $C_R$ ’ and ‘ $C_M$ ’ (Fig. 7(g)) with respect to ‘ $C_R$ ’ centered at  $C_P$ . The transformed pixels (Fig. 7(g)) contain noise which is smoothed by median filtering (Fig. 7(h)). Then morphological boundary extraction algorithms are applied to get boundary pixels (Fig. 7(i)) from where the large connected components are being extracted (Fig. 7(j)), which are removed from Fig. 7(i) to get a residue image (Fig. 7(k)). Now CHT based technique has been applied on boundary pixels of Fig. 7(k) to get a circular boundary Fig. 7(l) which undergoes to the inversion transformation to get circular boundary for the iris region (Fig. 7(m)). Fig. 7(n) and 7(o) show the iris localization and normalization for Fig. 7(a).

In VW image acquisition, the images suffer from various degrading factors such as reflections, off-angle, poor-focused, motion-blurred and improper lighting and due to these reasons the iris region has a slightly less dark region than the pupil region. It is difficult to select a proper threshold to distinguish between the pupil

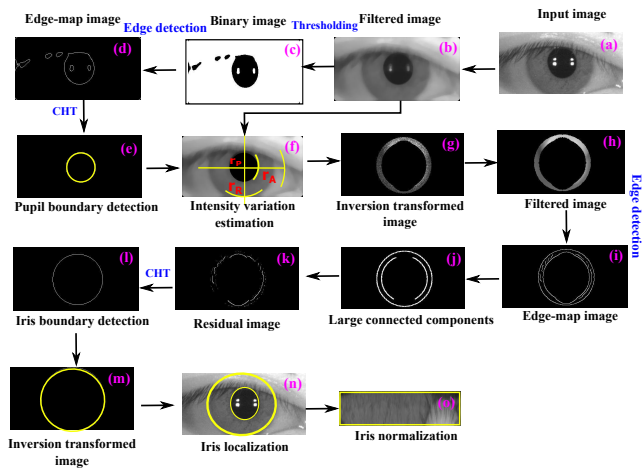


Fig. 7 Iris preprocessing for NIR images.

and the iris region. So, the iris outer boundary detection is easier than the pupil inner boundary detection. Here we convert color eyeball image  $\mathcal{E}$  (Fig. 8(a)) to a gray-scaled image  $\mathcal{E}'$  (Fig. 8(b)). Now similar to NIR iris localization here also a gray-level histogram is analyzed to find the best possible intensity value (i.e. a threshold) to convert  $\mathcal{E}'$  to the binary image  $\mathcal{B}$  (Fig. 8(c)). Here also the morphological boundary extraction algorithms followed by morphological area operators are used to remove small components from  $\mathcal{B}$  to obtain  $\mathcal{B}'$  (Fig. 8(d)). Now CHT technique is applied on the boundary pixels of  $\mathcal{B}'$  to get center and radius  $r_R$  (Fig. 8(e)) for  $\mathcal{E}'$ . The perceptual result of red light is better than green and blue light in VW images and it is due to the value of the wavelength of red light is higher than green and blue light which results in the better contrast between pupil & iris region in the red component image. Here during iris inner boundary detection, we consider the red component image  $\mathcal{E}''$  (Fig. 8(f)) of  $\mathcal{E}$ . Fig. 8(g) shows both the iris and pupil region of  $\mathcal{E}$ . Now for this region also the gray-level histogram is analyzed to get a proper threshold which is used to convert the image Fig. 8(g) to binary image  $\mathcal{B}'$  (Fig. 8(h)). Now canny edge detection algorithm is applied on  $\mathcal{B}'$  to get edge-map image  $\mathcal{B}''$  (Fig. 8(i)) which undergoes to removal the connected components and then CHT is applied on the boundary pixels of  $\mathcal{B}''$  to get pupil inner boundary center and radius  $r_{in}$  (Fig. 8(j)) which is projected on  $\mathcal{E}'$  (Fig. 8(k)). Fig. 8(k) and 8(l) show the iris localization and normalization for VW image  $\mathcal{E}$ .

After iris localization we extract the local periocular region  $\mathcal{P}$  from each  $\mathcal{E}$ . This local periocular region is extracted by considering a rectangular box centered with in iris region having  $\mathcal{P}$  of size  $r_{out} \times 2r_{out}$ , where  $r_{out}$  is the iris outer boundary radius for  $\mathcal{E}$ . Fig. 9 shows some extracted local periocular images  $\mathcal{P}$ s from  $\mathcal{E}$ s.

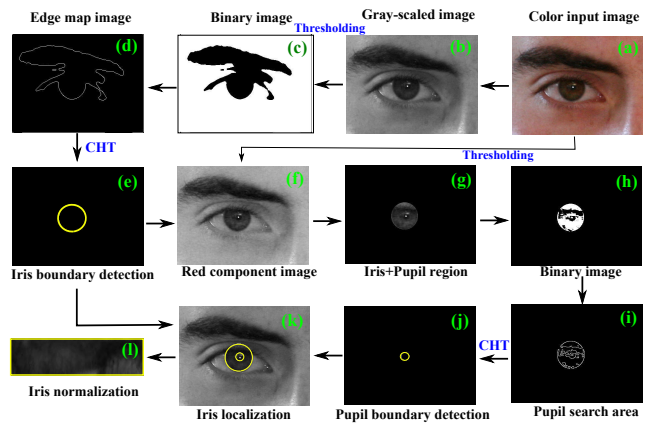


Fig. 8 Iris preprocessing for VW images.

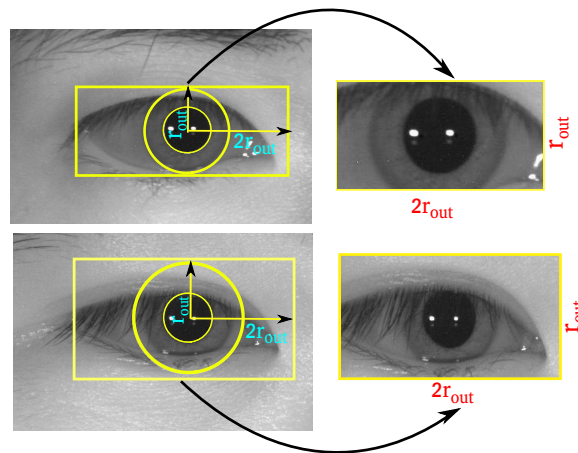


Fig. 9 Local periocular region extraction.

There exist various biometric databases in which some subjects have an insufficient number of samples. Moreover, there are high variations between the images from the same subject while the samples from different subjects, have very little dissimilarity and it is due to the noise artifacts (discussed in paragraph 1 at section 1) introduced during image acquisition. To overcome these situations data augmentation [30] techniques have been employed that can well train the deep CNN architecture.

### 3.2 Data augmentation

Data augmentation [30] is a method for constructing iterative optimization by introducing unobserved data. Due to insufficiency data, the data augmentation methods play an important role in machine learning algorithms, image classification, object recognition or detection problems. The composition of image augmentation with machine learning or deep learning techniques is a remediation on-demand where learners are capable of

gaining a better understanding of image classes during training and testing the classifier. Here the image classification algorithms perform better due to data augmentation [38] of images even the image is of lower quality. Some data augmentation techniques are discussed as follows: Perez et al. [38] have proposed Generative Adversarial Networks and pre-trained neural network-based data augmentation techniques for generic image classification problems and tested on some public domain datasets. Zhong et al. [39] have introduced the Random Erasing data augmentation method for learning the CNNs models to solve person re-identification, image classification, and object detection problems. The synthesizing of image samples have also been performed for data augmentation technique using *SamplePairing* technique by Inoue [40] for standard generic image classification problems. Dvornik et al. [41] have proposed the data augmentation techniques and learn the deep neural networks for the visual recognition system for the scene image understanding and segmentation. The data augmentation for the Alignment-Free Facial Attribute Classification Technique has been proposed by Günther [42] et al. to classify the facial attributes without any bounding box and alignment of the facial region. Mikołajczyk & Grochowski has experimentally improved the performance of deep learning in image classification problems due to the Generative Adversarial Net-based data augmentation technique [43]. Moreover, wank et al. [44] have also employed a Generative Adversarial Network-based data augmentation technique for the palmprint recognition system. In this work, we have also employed and investigated some data augmentation techniques which are applied on normalized iris pattern  $\mathcal{I}$  and local periocular region  $\mathcal{P}$ . The proposed data augmentation techniques are as follows:

- *Bilateral filtering (BF)*: It is a non-linear filtering technique which smooths an image while preserving the edges [45]. Mathematically, it is defined [46] as:

$$\mathcal{O}_B = \frac{\sum_{q \in S} \mathcal{G}_{\sigma_s}(\|p - q\|) I_q \mathcal{G}_{\sigma_r}(I_p - I_q)}{\sum_{q \in S} \mathcal{G}_{\sigma_s}(\|pq\|) \mathcal{G}_{\sigma_r}(I_p I_q)} \quad (1)$$

Where  $\sigma_s$  and  $\sigma_r$  measure the amount of filtering for  $I$ .  $\mathcal{G}_{\sigma_s}$  is the spatial Gaussian which decreases the influence of distant pixels whereas  $\mathcal{G}_{\sigma_r}$  is a range Gaussian that decreases the influence of pixels  $q$  with an intensity value different from  $I_p$ . The expression  $\sum_{q \in S} \mathcal{G}_{\sigma_s}(\|p - q\|) \times I_q$  is for linear filtering performing with Gaussian blur  $\mathcal{G}_{\sigma}(x) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2}{2\sigma^2}}$ . Fig. 10(b) and 11(b) show the effect of bilateral filtering on iris  $\mathcal{I}$  and periocular  $\mathcal{P}$  biometrics respectively.

- *Weighted Least Squares filtering (WLSF)*: It is also another image smoothing technique that uses the weighted least squares energy function [47]. Mathematically, it is defined as:

$$\mathcal{O}_W = \sum_p (O_p - I_p)^2 + \lambda \sum_{q \in N(p)} w_{p,q}(G) (O_p - O_q)^2$$

where  $I$ ,  $G$  and  $O$  are input, guidance and output image respectively and  $p = (x, y)$ ,  $I_p$  and  $G_p$  represent grayscale or color value of  $(x, y)$ .  $N(p)$  is a 4-neighbors of  $p$ ,  $\lambda$  controls the balance between two terms such that the increasing value of  $\lambda$  results more smoothing in  $O$ . The expression  $w_{p,q}(G) = e^{-\frac{\|G_p - G_q\|}{\sigma_c}}$  is spatially varying weight function. Fig.10(c) and 11(c) show the effect of WLSF technique applied on  $\mathcal{I}$  and  $\mathcal{P}$  respectively.

- The images filtered by WLSF technique in the Lab color space have different tone maps with coarse details ( $\mathcal{O}_C$ ), medium details ( $\mathcal{O}_M$ ) and fine details ( $\mathcal{O}_F$ ) respect to the variation of parameters used during WLSF technique. Fig. 10(d)-(f) and 11(d)-(f) show the coarse, medium and fine details for  $\mathcal{I}$  and  $\mathcal{P}$  respectively.
- *Image rotation* is a symmetric group of affine transformation on the image  $I$ . It is done by the bilinear interpolation of inverse transformation of every destination pixels of the image. In this work each  $I$  is rotated by an angle  $\theta = \{75^\circ, 150^\circ, 225^\circ\}$  in clockwise direction around its center point. Such that  $\mathcal{O}_T = ROT(I(\theta))$ ,  $\theta \in \{75^\circ, 150^\circ, 225^\circ\}$ . Fig. 10(g)-(i) and 11(g)-(i) show the examples of image rotation applied on  $\mathcal{I}$  and  $\mathcal{P}$  respectively.
- *Image reflection* is also a symmetric group of affine transformation applied on  $I$ . It is a special case of image rotation about  $180^\circ$  rotation corresponding to a given row number (about a line  $l$ ) of  $I$ . In this work each image is reflected about a line  $l = \{75, 150, 225\}$ , where  $l \ll \#$  of rows in  $I$  i.e.  $\mathcal{O}_R = REF(I(l))$ ,  $l \in \{75, 150, 225\}$ . Fig. 10(j)-(l) and 11(j)-(l) show the examples of image reflection for  $\mathcal{I}$  and  $\mathcal{P}$  respectively.

Here during experimentation for each normalized iris training image  $\mathcal{I}$  we apply the above data augmentation techniques and obtain an augmented set  $\mathcal{X} = \{\mathcal{I}\} \cup \{\mathcal{S}_1 = \{\mathcal{O}_B, \mathcal{O}_W, \mathcal{O}_C, \mathcal{O}_M, \mathcal{O}_F\}\} \cup \{\mathcal{S}_2 = \{\mathcal{O}_{T1}, \mathcal{O}_{T2}, \mathcal{O}_{T3}\}\} \cup \{\mathcal{S}_3 = \{\mathcal{O}_{R1}, \mathcal{O}_{R2}, \mathcal{O}_{R3}\}\}$ . Similarly for each periocular training image  $\mathcal{P}$  we apply data augmentation techniques and obtain an augmented set  $\mathcal{Y} = \{\mathcal{P}\} \cup \{\mathcal{S}_1 = \{\mathcal{P}_B, \mathcal{P}_W, \mathcal{P}_C, \mathcal{P}_M, \mathcal{P}_F\}\} \cup \{\mathcal{S}_2 = \{\mathcal{P}_{T1}, \mathcal{P}_{T2}, \mathcal{P}_{T3}\}\} \cup \{\mathcal{S}_3 = \{\mathcal{P}_{R1}, \mathcal{P}_{R2}, \mathcal{P}_{R3}\}\}$ .



The algorithmic sketch to perform data augmentation techniques on normalized iris  $\mathcal{I}$  and local periocular region  $\mathcal{P}$  is given below.

---

**Algorithm for the proposed Data augmentation techniques**

---

**Input:** Normalized iris  $\mathcal{I}$ , Local periocular  $\mathcal{P}$

**Output:** Data augmented set  $\mathcal{X}, \mathcal{Y}$

---

1. For each image  $\mathcal{I}$  and  $\mathcal{P}$ 
    - (a) Apply Eq.(1) on  $\mathcal{I}$  and  $\mathcal{P}$  and obtain  $\mathcal{O}_B$  and  $\mathcal{P}_B$  respectively.
    - (b) Apply Eq.(2) on  $\mathcal{I}$  and  $\mathcal{P}$  and obtain  $\mathcal{O}_W$  and  $\mathcal{P}_W$  respectively.
    - (c) Obtain coarse  $\mathcal{O}_C$ , medium  $\mathcal{O}_M$  and fine  $\mathcal{O}_F$  details for  $\mathcal{I}$ .
    - (d) Obtain coarse  $\mathcal{P}_C$ , medium  $\mathcal{P}_M$  and fine  $\mathcal{P}_F$  details for  $\mathcal{P}$ .
    - (e) Obtain  $\mathcal{S}_1 = \{\mathcal{O}_B, \mathcal{O}_W, \mathcal{O}_C, \mathcal{O}_M, \mathcal{O}_F\}$  for  $\mathcal{I}$
    - (f) Obtain  $\mathcal{S}_1 = \{\mathcal{P}_B, \mathcal{P}_W, \mathcal{P}_C, \mathcal{P}_M, \mathcal{P}_F\}$  for  $\mathcal{P}$
    - (g) Obtain  $\mathcal{S}_2 = \{\mathcal{O}_{T1}, \mathcal{O}_{T2}, \mathcal{O}_{T3}\}$  for  $\mathcal{I}$
    - (h) Obtain  $\mathcal{S}_2 = \{\mathcal{P}_{T1}, \mathcal{P}_{T2}, \mathcal{P}_{T3}\}$  for  $\mathcal{P}$
    - (i) Obtain  $\mathcal{S}_3 = \{\mathcal{O}_{R1}, \mathcal{O}_{R2}, \mathcal{O}_{R3}\}$  for  $\mathcal{I}$
    - (j) Obtain  $\mathcal{S}_3 = \{\mathcal{P}_{R1}, \mathcal{P}_{R2}, \mathcal{P}_{R3}\}$  for  $\mathcal{P}$
    - (k)  $\mathcal{X} = \{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$  for  $\mathcal{I}$
    - (l)  $\mathcal{Y} = \{\mathcal{P}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$  for  $\mathcal{P}$
- 

In this work the novelty of proposed data augmentation techniques are as follows:

1. **The employed image smoothing & image sharpening data augmentation ( $\mathcal{S}_1$ ) techniques** well extract silent sparse features in the iris and periocular biometrics like edges, contours and approximate the edge-sensitive texture patterns in those images and for this Bilateral and weighted least square filtering techniques have been applied. Both these filtering techniques preserve the edges in the texture region and perform multi-scale image decomposition tasks for tone mapping, texel detail enhancement and contrast manipulations in the iris and periocular images during feature learning [47].
2. **During image acquisition in unconstrained imaging environments**, the images suffer from various noises and motion blurring artifacts. To reduce the noises and blurring artifacts with preserving the gradient information, the bilateral and weighted least squares filters separate the image structure at different scales while returning the high contrast fine detailed features and edges with increased texture information. The bilateral filter gives the larger

intensity difference whereas the WLS filter gives the larger magnitude of gradients with more discriminant features [48].

3. **Here image rotation data augmentation ( $\mathcal{S}_2$ ) technique** with respect to  $\theta = \{75^\circ, 150^\circ, 225^\circ\}$  provides the rotation invariant feature learning to the deep CNN architecture to derive rotation-invariant features for iris and periocular biometrics with more discriminating patterns for the proper image rotation [38].
4. **The image reflection data augmentation ( $\mathcal{S}_3$ ) technique** maintains the gradient information in an image corresponding to the pixels, remain the same in the region of interest. This flipping of images about a line not only increases the number of samples but also increase the effectiveness and robustness of the classification model for an unknown sample [38].
5. **Here the proposed data augmentation techniques** solve the problems of image quality degradation and the insufficiency of training data. These techniques handle the weight decay, dropout, regularization effects and overfitting problems of CNN learning.

### 3.3 Feature learning for classification

The iris and periocular biometrics have more distinctive and unique features for every individual. So, during feature extraction from these biometrics generally three approaches have been employed: (i) Transformed approach which relies on the statistics of filter responses (such as wavelets, Gabor, Contourlet and radon transform), (ii) Structural approach relates to the geometrical representation of texture elements by characterizing the texture primitives with the spatial arrangement of those primitives (such as Mathematical Morphology, Local Binary Pattern and Fractal Dimension) and (iii) Statistical approach relates to the local properties of texture and the statistical distribution of image intensities over those texture region in the image (such as Grey-level Co-occurrence Matrix, Eigen Region, Scale Invariant feature transform, Histogram of oriented gradients, etc.).

The Convolutional neural networks (CNNs) models [49] are used nowadays to solve various problems of texture classification, object recognition, object detection, face recognition, scene understanding and in many more computer vi-

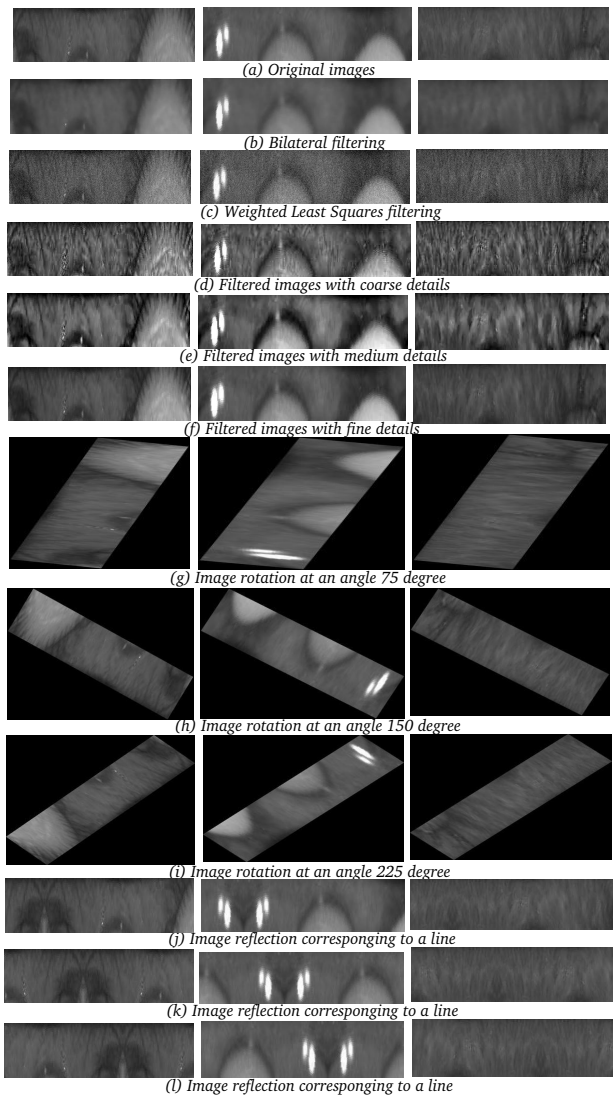


Fig. 10 Data augmentation techniques on iris images  $\mathcal{I}$ .

sion applications and for analyzing and solving these problems the CNN model adopts the machine learning tasks. It requires minimal image preprocessing tasks and learns the network in the feature space based on the weights adjusting during training the network. The overall design of CNN architecture is based on hybridization approaches [50] where both the shapes and textures information is analyzed using various machine learning optimization techniques. The CNN model uses the huge amount of computing power handled by GPUs and provides the multiple levels of representation and abstraction for the complex data patterns and for this pre-trained, transfer learning and fine-tuning methods have been deputed to obtain better performance for the proposed system.

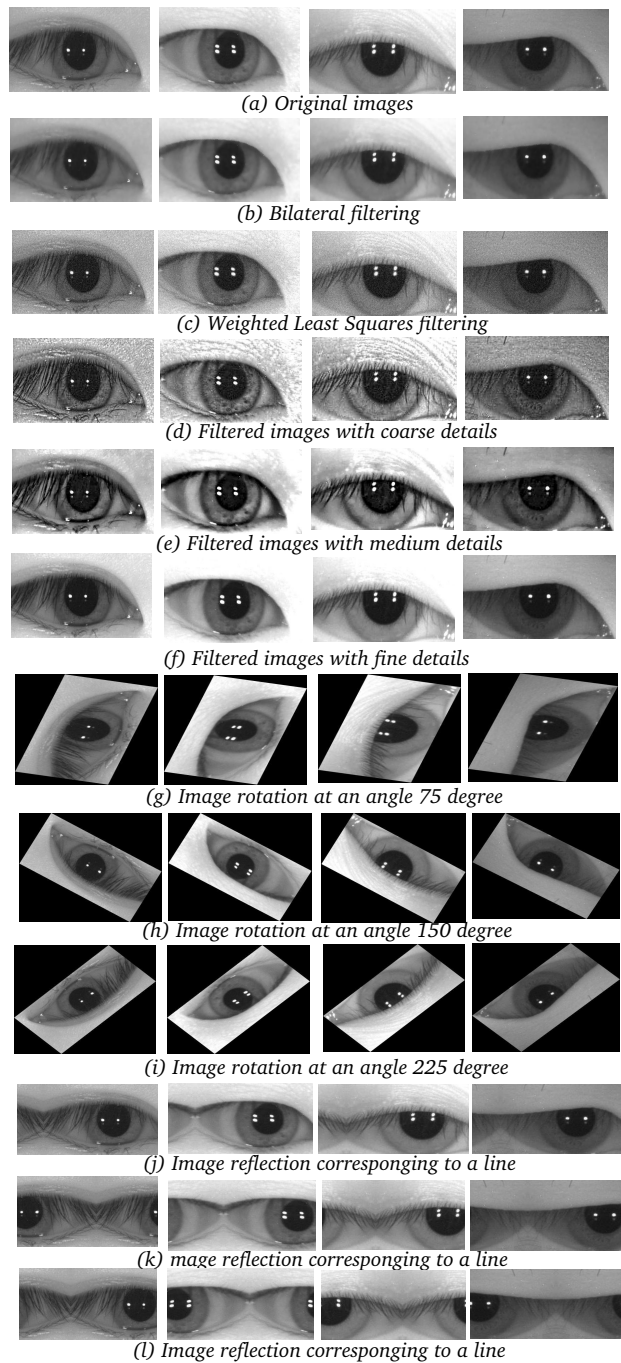
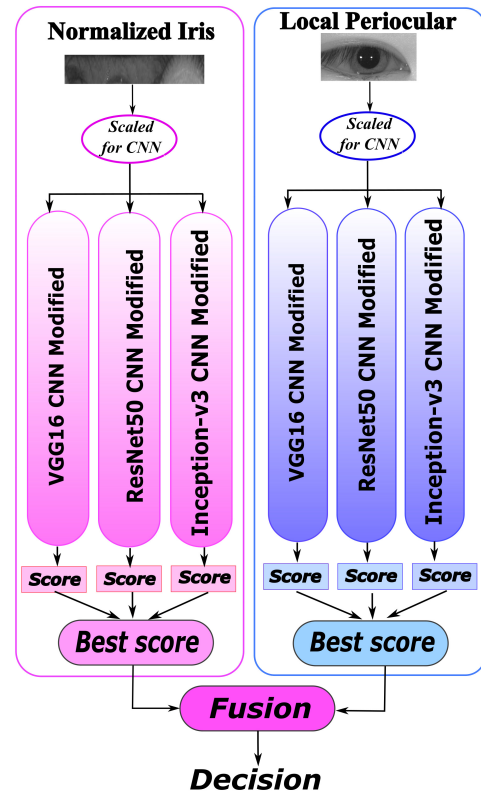


Fig. 11 Data augmentation techniques on periocular images  $\mathcal{P}$ .

In the pre-trained model, the model which has been created earlier, are being reused to solve the similar kinds of the problem instead of deriving the new model from the scratch [51]. The use of a pre-trained model not only saves huge efforts but also adopts better parameter settings for similar kinds of problems. The pre-trained model has already been trained on a large database and the weights are adjusted ac-

cordingly respect to employee database so, these weights and the CNN architecture may be used further for the proposed system followed by the transfer learning approach [51]. Moreover, the parameters of the CNN model have been fine-tuned to the target database [52] for reducing the over-fitting problem of the biometric recognition system. Every CNN model has a Convolutional base (feature extraction) and classifier (image classification) part. In this work, we have used both these parts and have trained the entire pre-trained model for the normalized iris and local periocular region database individually by employing the transfer learning and fine-tuning approaches. For the pre-trained model, we have been adopted the VGG16, ResNet50 and Inception-v3 model (described at Section 2) and then the last two layers of each model have been replaced by two fully connected layers. Then we have applied transfer learning for iris/periocular database and finally, finetune the weights of the pre-trained model with a small learning rate for the number of subjects enrolled for iris/periocular database. Fig 12 demonstrates the feature learning for classification technique of the proposed system.



**Fig. 12** The feature learning technique for classification of the proposed system.

## 4 Experimental results

### 4.1 Databases used

The proposed system is evaluated using four benchmark eyeball databases, namely, MMU1 [53], UPOL [54], CASIA-Iris-Distance [55] (say CASIA-dist) and UBIRIS.v2 [1]. The images of MMU1 and UPOL database were captured in NIR and VW lighting respectively under a constrained environment. The images of CASIA-dist were captured at a distance of 3-10 meters under an unconstrained environment and for this reason, the images suffer from various noise artifacts as mentioned previously. Moreover, the images of CASIA-dist were captured in such a way that both dual eye iris and face patterns are included and for our interest Viola-Jones algorithm is employed to crop the eyeball images from each image of this database. The images of UBIRIS.v2 were captured in conditions when the subjects are moving and also at varying distances and as a result, the images suffer from reflections, off-angle, partial images, poor focused, motion-blur, improper lighting, rotation, and noise [2]. Table 2 summarizes the features of these databases.

During experimentation, for each subject, one sample is considered as a training sample while the remain-

**Table 2** Databases used during experimentation.

Database (Imaging wavelength & environment)	Subject	Samples /Subject	Image size
MMU1 (NIR & Cons.)	90	5	320×280
UPOL (VW & Cons.)	128	3	480×640
CASIA-dist (NIR & Uncons.)	276	8	600×400
UBIRIS.v2 (VW & Uncons.)	422	8	450×350

ing samples are considered as testing samples. Here to increase the number of subjects, the left eye and right eye images for a subject are considered as images from two different subjects. So, UPOL database has 128 subject whereas 128 ( $= 128 \times 1$ ) training samples and 256 ( $= 128 \times 2$ ) testing samples. MMU1 database has 90 subjects whereas 90 ( $= 90 \times 1$ ) training samples and 360 ( $= 90 \times 4$ ) testing samples. CASIA-dist database has 276 subjects whereas 276 ( $= 276 \times 1$ ) training samples and 1932 ( $= 276 \times 7$ ) testing samples. UBIRIS-v2 has 422 subjects whereas 422 ( $= 422 \times 1$ ) training samples and 2954 ( $= 422 \times 7$ ) testing samples.

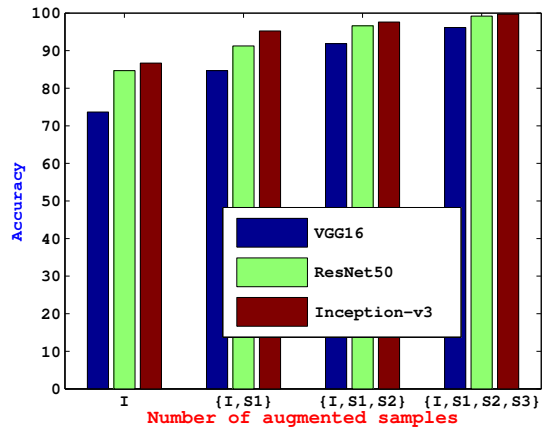
## 4.2 Results and discussions

The proposed biometric recognition system is implemented in Python on Ubuntu 16.04 LTS O/S version with Intel Core i7 processor 3.20 GHz. Here we have used Theano and Keras packages [56] to train the VGG16, ResNet50, and Inception-v3 CNN architectures. During iris localization we extract iris portion from an eyeball image  $\mathcal{E}$  using the methods described either in Fig 7 or Fig 8 for NIR or VW images. Then we extract local periocular region ( $\mathcal{P}$ ) of size  $100 \times 300$  using the parameters of iris localization for  $\mathcal{E}$ . During iris normalization, the localized annular iris region is transformed to an image  $\mathcal{I}$  of size  $100 \times 360$  pixels (Fig. 7(1) and 8(1)). The employed iris localization methods achieve better segmentation accuracy and the consumption of time (in sec.) with the state-of-the-art methods [37].

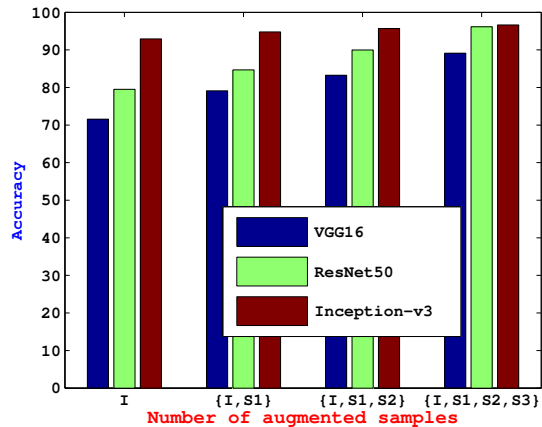
During feature extraction and classification, we train each VGG16, ResNet50, and Inception-v3 CNN architectures accordingly using fine-tuning. Note that for VGG16 and ResNet50, the image size is  $224 \times 224 \times 3$  whereas for Inception-v3 the image size is  $299 \times 299 \times 3$ . Here during feature learning, we resized each normalized iris or local Periocular image to an image of size  $224 \times 224 \times 3$  for VGG16 and ResNet50 while for Inception-v3, these images are resized to  $299 \times 299 \times 3$ . At the time of training, we tune the parameters for the dimensionality of features, number of subjects corresponding to each database, regularization for optimization functions on the output layer respectively. During network training we have employed the mini-batch gradient descend technique [57] with 500 batch size, # of epochs being 100 with a weight decay of 0.95 and 0.001 of learning rates.

To demonstrate the effectiveness of data augmentation techniques employed on training samples, we have drawn a graph (see Fig. 13(a) and (b) with x-axis as # of augmented training samples with accuracy (%) of testing samples as in the y-axis. Here the performance is shown for UPOL and MMU1 database. From the Fig. 13(a) and (b) it is observed that UPOL and MMU1 database have achieved much higher performance for test iris samples due to augmented training samples  $\{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$  for each  $\mathcal{I}$  for VGG16, ResNet50 and Inception-v3 architecture. The images of MMU1 and UPOL database were captured in a constrained environment and the recognition system has achieved better performance even with a very small number of training samples. Hence from Fig. 13(a) and (b) it is observed that for UPOL and MMU1 database, the recognition system has achieved better performance for  $\{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$  data augmented set, i.e., for data augmentation of filter banks on  $\mathcal{I}$ , rotation of  $\mathcal{I}$  with  $\theta = \{75, 150, 225\}$

(experimentally) and reflection of  $\mathcal{I}$  about a line  $l = \{75, 150, 225\}$  (experimentally).



(a)



(b)

**Fig. 13** Effectiveness of data augmentation on (a) UPOL and (b) MMU1 database.

The bilateral and weighted least squares filters (WLS) smooth the image  $I$  while it preserves the edges and separates the image structure at different scales. Moreover, due to these filters, the high contrast fine detailed features and edges are reflected in the image and these increase the texture information. The bilateral filter gives the larger intensity difference while WLS filter gives the larger magnitude of the gradient at the edges and these facilitate more discriminant features [48]. In the set  $\{I, S_1\}$ , the filter response of bilateral and WLS are considered. So, the proposed system gets more improvement from  $\{I\}$  to  $\{I, S_1\}$  and it is also larger than from  $\{I, S_1, S_2\}$  and  $\{I, S_1, S_2, S_3\}$  to  $\{I, S_1\}$ .

The performance of the proposed system for CASIA-dist and UBIRIS-v2 database are shown in Table 3 for both without-augmented training samples (i.e.  $\mathcal{I}$ ) and

with-augmented training samples (i.e.  $\{\mathcal{I}, \mathcal{S}_1\}$ ,  $\{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2\}$ ,  $\{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$ ). Here only Correct Recognition Rate (CRR) corresponding to each database for VGG16, ResNet50 and Inception-v3 CNN features are shown respectively. For CASIA-dist and UBIRIS-v2 database we have also compared the performance with some new CNN architectures such as Inception-ResNet [58]. Inception-ResNet [58] is the ensemble of ResNet and Inception-v3 CNN models and it also takes  $299 \times 299 \times 3$  as input image. From Table 3, it is observed that for both the databases, the proposed system has more or less better performance for Inception-v3 CNN features than ResNet50, VGG16 and Inception-ResNet [58] CNN features with-augmented training samples.

**Table 3** Performance of the proposed iris recognition system.

Database	$\mathcal{I}$	$\{\mathcal{I}, \mathcal{S}_1\}$	$\{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2\}$	$\{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$
	CRR	CRR	CRR	CRR
VGG16				
CASIA-dist	86.39	97.83	98.13	99.64
UBIRIS.v2	85.12	95.39	96.53	97.12
ResNet50				
CASIA-dist	84.44	89.67	94.84	98.92
UBIRIS.v2	89.05	94.87	96.12	98.05
Inception-v3				
CASIA-dist	90.23	94.12	97.01	99.64
UBIRIS.v2	89.71	96.46	97.68	98.76
Inception-ResNet				
CASIA-dist	88.05	94.22	93.56	99.19
UBIRIS.v2	87.12	95.26	94.58	98.34

Hence from the experimentation in Table 3 it is concluded that VGG16, ResNet50, Inception-v3 and Inception-ResNet CNN architectures, have more or less similar performance for  $\{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$  augmented set but still for Inception-v3 the proposed system has achieved better performance. Table 4 shows the performance of the proposed periocular recognition system due to  $\{\mathcal{P}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$  augmented training set using VGG16, ResNet50, Inception-v3 and Inception-ResNet networks. Here also the performance are shown in terms of CRR (correct recognition rate) for both CASIA-dist and UBIRIS-v2 database. From Table 4 we see that for Inception-v3 network, the periocular recognition system has better performance.

**Table 4** Performance of the proposed periocular recognition system.

Database	VGG16	ResNet50	Inception-v3	Inception-ResNet
	CRR	CRR	CRR	CRR
CASIA-dist	96.79	97.68	98.69	98.43
UBIRIS.v2	97.99	98.12	98.37	98.16

**Table 5** Performance comparison for the proposed system.

Database	Iris		Periocular	
	CASIA-dist	UBIRIS.v2	CASIA-dist	UBIRIS.v2
	CRR	CRR	CRR	CRR
Texture feature models				
LBP [59]	38.71	42.78	41.78	51.47
VQ-SPM [60]	71.79	66.18	65.34	62.32
SRC-SPM [61]	79.12	72.89	69.12	69.51
LLC-SPM [62]	76.82	75.12	71.45	70.12
Gabor [63]	61.21	49.12	51.89	49.12
Tan & Kumar [4]	94.50	83.00	63.00	82.31
Chen et al. [64]	95.05	—	—	—
Proenca et al. [5]	—	—	—	88.07
Proenca et al. [65]	—	—	99.31	88.29
USIT feature models				
Masek [8]	63.12	59.90	61.78	63.71
Ma [66]	58.12	53.19	52.21	48.39
Ko [67]	61.79	65.12	59.24	62.12
Rathgeb [68]	45.19	39.90	49.25	43.11
Rathgeb [69]	51.89	48.12	44.90	57.12
Monro [9]	59.12	46.60	62.12	53.91
Proposed	<b>99.64</b>	<b>98.76</b>	<b>98.69</b>	<b>98.37</b>

Hence from Table 3 and 4 we have derived a conclusion that for further experimentation we have employed the performance due to training augmented set  $\{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$  using Inception-v3 network.

Table 5 demonstrates the comparison of the proposed system with the existing methods for the corresponding databases under the same training-testing protocol. Here we have done a set of experiments using some well-known texture features like Local Binary Pattern (LBP) [59], Vector Quantization followed by Spatial Pyramid Mapping (VQ-SPM) [60], Sparse Representation Coding followed by SPM (SRC-SPM) [61], Locality-constrained Linear Coding followed by SPM (LLC-SPM) [62], Gabor phase feature [63] and Tan & Kumar [4] feature. These features have gained success and popularity in the field of biometric recognition and texture classification problems. During comparison from each  $\mathcal{I}$  and  $\mathcal{P}$ , we extract LBP<sup>1</sup> feature of dimension 531, Gabor<sup>2</sup> with respect to 6 scales and 8 orientations and obtain a feature vector of dimension 1440, each VQ-SPM<sup>3</sup>, SRC-SPM<sup>4</sup> and LLC-SPM<sup>5</sup> are of 5250 dimensional feature vector. We have also compared the performance with some recent articles such as [64], [65] and [5]. Chen et al. [64] obtained the performance for CASIA-dist iris database which is 95.05%. Proenca et al. has obtained the performance for UBIRIS-v2 periocular database 88.07% in [5] and 88.29% in [65].

We have also compared the performance of the proposed method with the methods implemented in the

<sup>1</sup> <http://www.cse.oulu.fi/CMV/Downloads/LBP Matlab>

<sup>2</sup> <http://www.mathworks.com/matlabcentral/fileexchange/44630-gabor-feature-extraction>

<sup>3</sup> <http://slazebni.cs.illinois.edu/>

<sup>4</sup> <http://www.ifp.illinois.edu/~jyang29/ScSPM.htm>

<sup>5</sup> <http://www.ifp.illinois.edu/~jyang29/LLC.htm>



University of Salzburg Iris Toolkit (USIT) [70] software. This software is free to use. In this software, various feature extraction algorithms like Masek et al. [8], Ma et al. [66], Ko et al. [67], Rathgeb et al. [68], Rathgeb et al. [69] and Monro et al. [9], have been implemented. We obtain the performance for each method under the same training-testing protocol as used by the proposed system. The results are reported also in Table 5 and from this table, it is observed that the proposed feature outperforms.

From the experiments described in Table 3 and Table 4, it has been shown that Inception-v3 model has obtained better performance than VGG16, ResNet50 and Inception-ResNet models. So, using Inception-v3 model due to CASIA-dist normalized iris pattern, for every  $k_{th}$  subject,  $\mathcal{M} = 276$  scores are obtained which are arranged in decreasing order. Then the ranks are assigned to the sorted scores. Suppose  $r_{\mathcal{I}_1}^k, r_{\mathcal{I}_2}^k, \dots, r_{\mathcal{I}_{276}}^k$  be the ranks for  $k_{th}$  subject iris image  $\mathcal{I}$ . Similarly due to CASIA-dist local periocular, for every  $k_{th}$  subject  $\mathcal{M} = 276$  scores are obtained and sorted in decreasing order and then ranks are obtained which are  $r_{\mathcal{P}_1}^k, r_{\mathcal{P}_2}^k, \dots, r_{\mathcal{P}_{276}}^k$ . Now according to HR technique the rank of  $j_{th}$  subject can be obtained as  $r_j^k = \min(r_{\mathcal{I}_j}^k, r_{\mathcal{P}_j}^k)$ , using Borda Count (BC) technique, the rank will be  $r_j^k = r_{\mathcal{I}_j}^k + r_{\mathcal{P}_j}^k$  while for Logistic Regression (LR) technique, the rank will be  $r_j^k = \mathcal{W}_{\mathcal{I}} * r_{\mathcal{I}_j}^k + \mathcal{W}_{\mathcal{P}} * r_{\mathcal{P}_j}^k$ ,  $\mathcal{W}_{\mathcal{I}} + \mathcal{W}_{\mathcal{P}} = 1$ ,  $\mathcal{W}_{\mathcal{I}} = 0.6$  and  $\mathcal{W}_{\mathcal{P}} = 0.4$ . Then in each rank-level fusion technique, the comparison of fused ranks is being performed with its actual rank 1 positions and finally, the number of correct matching will be the correct recognition rate of the proposed system. Similarly, the ranks for  $\mathcal{M} = 422$  subjects due to each UBIRIS.v2 normalized iris pattern and local periocular region are obtained and then the corresponding fusion techniques have been applied to obtain the performance of the proposed system for UBIRIS.v2 database. Table 6 demonstrates the fused performance of the proposed system in terms of correct recognition rates (CRR) (in %) due to the fusion of ranks for iris and periocular biometrics.

**Table 6** Fused performance of proposed system in CRR (%).

Database	Iris	Periocular	HR	BC	LR
CASIA-dist	99.64	98.69	99.49	99.91	99.76
UBIRIS.v2	98.76	98.37	98.82	99.37	99.14

From Table 6 it has been observed that the proposed system has attained better performance for BC rank fusion technique. So, for the proposed system we have employed BC rank-level fusion technique to combine the identification performance of iris and periocular biometrics.

## 5 CONCLUSION

A novel method for iris and periocular based biometric recognition system have been proposed in this paper. The proposed system has “image preprocessing”, “data augmentation” and “feature learning for classification” components. The image preprocessing component obtains the iris and periocular regions from an eyeball image under both constrained and unconstrained environments. Similar types of distinctive samples for each training sample, have been generated by the proposed data augmentation technique to increase the number of training samples. The obtained training samples undergo to the feature learning for classification component to fine-tune the employed CNN architecture to obtain the models for iris and periocular recognition system using VGG16, ResNet50, Inception-v3, and Inception - ResNet CNN architecture. Experimentally, it is concluded that the proposed system has obtained better performance due to  $\{\mathcal{I}, \mathcal{S}_1, \mathcal{S}_2, \mathcal{S}_3\}$  augmented iris and periocular training sets. Extensive experimentation has been performed using MMU1, UPOL, CASIA-Iris-Distance, and UBIRIS.v2 database and the performance have been compared with various existing state-of-the-art methods corresponds to each database. From the comparison, it has been shown that the proposed system has outstanding performance and it outperforms other competing methods reported in this paper. Finally, to increase the performance of the recognition system, the scores due to iris and periocular biometrics have been fused using a rank-level fusion technique to obtain the performance for the multimodal biometric recognition system.

## References

1. Hugo Proenca. Iris recognition: What is beyond bit fragility? *Information Forensics and Security, IEEE Transactions on*, 10(2):321–332, 2015.
2. Hugo Proenca. Iris recognition: On the segmentation of degraded images acquired in the visible wavelength. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(8):1502–1516, 2010.

3. Tieniu Tan, Xiaobo Zhang, Zhenan Sun, and Hui Zhang. Noisy iris image matching by using multiple cues. *Pattern Recognition Letters*, 33(8):970–977, 2012.
4. Chun-Wei Tan and Ajit Kumar. Accurate iris recognition at a distance using stabilized iris encoding and zernike moments phase features. *Image Processing, IEEE Transactions on*, 23(9):3962–3974, 2014.
5. Hugo Proença and João C Neves. Deep-prwis: Periocular recognition without the iris and sclera using deep learning frameworks. *IEEE Transactions on Information Forensics and Security*, 13(4):888–896, 2018.
6. Unsang Park, Raghavender Reddy Jillela, Arun Ross, and Anil K Jain. Periocular biometrics in the visible spectrum. *IEEE Transactions on Information Forensics and Security*, 6(1):96–106, 2011.
7. Zijing Zhao and Ajay Kumar. Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE Transactions on Information Forensics and Security*, 12(5):1017–1030, 2017.
8. Libor Masek and Peter Kovesi. Matlab source code for a biometric identification system based on iris patterns. *The School of Computer Science and Software Engineering, The University of Western Australia*, 26, 2003.
9. Donald M Monro, Soumyadip Rakshit, and Dexin Zhang. Det-based iris recognition. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(4):586–595, 2007.
10. Karen P Hollingsworth, Kevin W Bowyer, and Patrick J Flynn. The best bits in an iris code. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 31(6):964–973, 2009.
11. Amol D Rahulkar and Raghunath S Holambe. Half-iris feature extraction and recognition using a new class of biorthogonal triplet half-band filter bank and flexible k-out-of-n: A postclassifier. *IFS, IEEE Trans. on*, 7(1):230–240, 2012.
12. Damon L Woodard, Shrinivas Pundlik, Philip Miller, Raghavender Jillela, and Arun Ross. On the fusion of periocular and iris biometrics in non-ideal imagery. In *Pattern Recognition (ICPR), 2010 20th International Conference on*, pages 201–204. IEEE, 2010.
13. Jaishanker K Pillai, Vishal M Patel, Rama Chellappa, and Nalini K Ratha. Secure and robust iris recognition using random projections and sparse representations. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(9):1877–1893, 2011.
14. Saiyed Umer, Bibhas Chandra Dhara, and Bhabatosh Chanda. Iris recognition using multiscale morphologic features. *Pattern Recognition Letters*, 65:67–74, 2015.
15. Ronaldo Martins da Costa and Adilson Gonzaga. Dynamic features for iris recognition. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, 42(4):1072–1082, 2012.
16. Somnath Dey and Debasis Samanta. Iris data indexing method using gabor energy features. *IFS, IEEE Trans. on*, 7(4):1192–1203, 2012.
17. Chun-Wei Tan and Ajit Kumar. Efficient and accurate at-a-distance iris recognition using geometric key-based iris encoding. *Information Forensics and Security, IEEE Transactions on*, 9(9):1518–1526, 2014.
18. Kazuyuki Miyazawa, Koichi Ito, Takafumi Aoki, Koji Kobayashi, and Hiroshi Nakajima. An effective approach for iris recognition using phase-based image matching. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 30(10):1741–1756, 2008.
19. Saiyed Umer, Bibhas Chandra Dhara, and Bhabatosh Chanda. A novel cancelable iris recognition system based on feature learning techniques. *Information Sciences*, 406:102–118, 2017.
20. Ajay Kumar and Tak-Shing Chan. Iris recognition using quaternionic sparse orientation code (qsoc). In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, pages 59–64. IEEE, 2012.
21. Hui Zhang, Zhenan Sun, Tieniu Tan, and Jianyu Wang. Ethnic classification based on iris images. In *Biometric Recognition*, pages 82–90. Springer, 2011.
22. Zhenan Sun, Hui Zhang, Tieniu Tan, and Jianyu Wang. Iris image classification based on hierarchical visual codebook. 2013.
23. Wenbo Dong, Zhenan Sun, and Tieniu Tan. Iris matching based on personalized weight map. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(9):1744–1757, 2011.
24. V Paúl Pauca, Michael Forkin, Xiao Xu, Robert Plemmons, and Arun A Ross. Challenging ocular image recognition. In *SPIE Defense, Security, and Sensing*, pages 80291V–80291V. International Society for Optics and Photonics, 2011.
25. Juan Reyes-Lopez, Sergio Campos, Hector Allende, and Rodolfo Salas. Zernike’s feature descriptors for iris recognition with svm. In *Computer Science Society (SCCC), 2011 30th International Conference of the Chilean*, pages 283–288. IEEE, 2011.

26. Chun-Wei Tan and Ajay Kumar. Towards online iris and periocular recognition under relaxed imaging constraints. *Image Processing, IEEE Transactions on*, 22(10):3751–3765, 2013.
27. Saiyed Umer, Bibhas Chandra Dhara, and Bhabatosh Chanda. Texture code matrix-based multi-instance iris recognition. *Pattern Analysis and Applications*, pages 1–13, 2015.
28. Qi Zhang, Haiqing Li, Zhenan Sun, and Tieniu Tan. Deep feature fusion for iris and periocular biometrics on mobile devices. *IEEE Transactions on Information Forensics and Security*, 13(11):2897–2912, 2018.
29. Hugo Proença and João C Neves. Deep-prwis: Periocular recognition without the iris and sclera using deep learning frameworks. *IEEE Transactions on Information Forensics and Security*, 13(4):888–896, 2017.
30. David A Van Dyk and Xiao-Li Meng. The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10(1):1–50, 2001.
31. Yann LeCun et al. Lenet-5, convolutional neural networks. URL: <http://yann.lecun.com/exdb/lenet>, 2015.
32. Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
33. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
34. Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
35. Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2818–2826, 2016.
36. Anil Jain, Karthik Nandakumar, and Arun Ross. Score normalization in multimodal biometric systems. *Pattern recognition*, 38(12):2270–2285, 2005.
37. Saiyed Umer, Bibhas Chandra Dhara, and Bhabatosh Chanda. Nir and vw iris image recognition using ensemble of patch statistics features. *The Visual Computer*, pages 1–18.
38. Luis Perez and Jason Wang. The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*, 2017.
39. Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. *arXiv preprint arXiv:1708.04896*, 2017.
40. Hiroshi Inoue. Data augmentation by pairing samples for images classification. *arXiv preprint arXiv:1801.02929*, 2018.
41. Nikita Dvornik, Julien Mairal, and Cordelia Schmid. On the importance of visual context for data augmentation in scene understanding. *arXiv preprint arXiv:1809.02492*, 2018.
42. Manuel Günther, Andras Rozsa, and Terranee E Boul. Affact: Alignment-free facial attribute classification technique. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pages 90–99. IEEE, 2017.
43. Agnieszka Mikołajczyk and Michał Grochowski. Data augmentation for improving deep learning in image classification problem. In *2018 international interdisciplinary PhD workshop (IIPhDW)*, pages 117–122. IEEE, 2018.
44. Gengxing Wang, Wenxiong Kang, Qiuxia Wu, Zhiyong Wang, and Junbin Gao. Generative adversarial network (gan) based data augmentation for palmprint recognition. In *2018 Digital Image Computing: Techniques and Applications (DICTA)*, pages 1–7. IEEE, 2018.
45. Sylvain Paris, Pierre Kornprobst, Jack Tumblin, Frédo Durand, et al. Bilateral filtering: Theory and applications. *Foundations and Trends® in Computer Graphics and Vision*, 4(1):1–73, 2009.
46. Sylvain Paris, Pierre Kornprobst, Jack Tumblin, and Frédo Durand. A gentle introduction to bilateral filtering and its applications. In *ACM SIGGRAPH 2007 courses*, page 1. ACM, 2007.
47. Zeev Farbman, Raanan Fattal, Dani Lischinski, and Richard Szeliski. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)*, volume 27, page 67. ACM, 2008.
48. Kartic Subr, Cyril Soler, and Frédo Durand. Edge-preserving multiscale image decomposition based on local extrema. In *ACM Transactions on Graphics (TOG)*, volume 28, page 147. ACM, 2009.
49. Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
50. Vincent Andrearczyk and Paul F Whelan. Using filter banks in convolutional neural networks for texture classification. *Pattern Recognition Letters*,

- 84:63–69, 2016.
51. Ioannis Athanasiadis, Panagiotis Mousoulitis, and Loukas Petrou. A framework of transfer learning in object detection for embedded systems. *arXiv preprint arXiv:1811.04863*, 2018.
  52. Wanli Ouyang, Xiaogang Wang, Cong Zhang, and Xiaokang Yang. Factors in finetuning deep model for object detection with long-tail distribution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 864–873, 2016.
  53. Multimedia university iris database [online]. Available:<http://pesona.mmu.edu.my/ccteo/>.
  54. Michal Dobes and Libor Machala. Iris database. *Palacky University in Olomouc, Czech Republic*, <http://www.inf.upol.cz/iris>, 2007.
  55. Casia iris image databases service team. *CAS Institute of Automation*, <http://biometrics.idealtest.org/>, 2009.
  56. François Chollet et al. Keras: Deep learning library for theano and tensorflow. URL: <https://keras.io/k>, 2015.
  57. Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*, 2015.
  58. Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.
  59. Guoying Zhao and Matti Pietikainen. Dynamic texture recognition using local binary patterns with an application to facial expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(6):915–928, 2007.
  60. Svetlana Lazebnik, Cordelia Schmid, and Jean Ponce. Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, volume 2, pages 2169–2178. IEEE, 2006.
  61. Jianchao Yang, Kai Yu, Yihong Gong, and Thomas Huang. Linear spatial pyramid matching using sparse coding for image classification. In *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, pages 1794–1801. IEEE, 2009.
  62. Jinjun Wang, Jianchao Yang, Kai Yu, Fengjun Lv, Thomas Huang, and Yihong Gong. Locality-constrained linear coding for image classification. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 3360–3367. IEEE, 2010.
  63. John Daugman. How iris recognition works. *CSVT, IEEE Trans. on*, 14(1):21–30, 2004.
  64. Jianxu Chen, Feng Shen, Danny Ziyi Chen, and Patrick J Flynn. Iris recognition based on human-interpretable features. *IEEE Transactions on Information Forensics and Security*, 11(7):1476–1485, 2016.
  65. Hugo Proença and João C Neves. A reminiscence of mastermind: Iris/periocular biometrics by in-set cnn iterative analysis. *IEEE Transactions on Information Forensics and Security*, 2018.
  66. Li Ma, Tieniu Tan, Yunhong Wang, and Dexin Zhang. Personal identification based on iris texture analysis. *PAMI, IEEE Trans. on*, 25(12):1519–1533, 2003.
  67. Jong-Gook Ko, Youn-Hee Gil, Jang-Hee Yoo, and Kyo-IL Chung. A novel and efficient feature extraction method for iris recognition. *ETRI journal*, 29(3):399–401, 2007.
  68. Christian Rathgeb and Andreas Uhl. Secure iris recognition based on local intensity variations. In *Image Analysis and Recognition*, pages 266–275. Springer, 2010.
  69. Christian Rathgeb and Andreas Uhl. Context-based biometric key generation for iris. *IET computer vision*, 5(6):389–397, 2011.
  70. Christian Rathgeb, Andreas Uhl, and Peter Wild. Iris recognition: from segmentation to template security. *Advances in Information Security*, 59, 2012.