A novel periocular biometrics solution for authentication during Covid-19 pandemic situation

Punam Kumari¹ · K. R. Seeja¹

Received: 16 July 2020 / Accepted: 9 December 2020 / Published online: 3 January 2021
© The Author(s), under exclusive licence to Springer-Verlag GmbH, DE part of Springer Nature 2021

Abstract

The outbreak of novel coronavirus in 2019 has shaken the whole world and it quickly evolved as a global pandemic, placing everyone in a panic situation. Considering its long-term effects on day to day lives, the necessity of wearing face mask and social distancing brings in picture the requirement of a contact less biometric system for all future authentication systems. One of the solutions is to use periocular biometric as it does not need physical contact like fingerprint biometric and is able to identify even people wearing face masks. Since, the periocular region is a small area as compared to face, extraction of required number of features from that small region is the major concern to make the system highly robust. This research proposes a feature fusion approach which combines the handcrafted features HOG, non-handcrafted features extracted using pretrained CNN models and gender related features extracted using a five layer CNN model. The proposed feature fusion approach is evaluated using multiclass SVM classifier with three different benchmark databases, UBIPr, Color FERET and Ethnic Ocular as well as for three non-ideal scenarios i.e. the effect of eyeglasses, effect of eye occlusion and pose variations. The proposed approach shows remarkable improvement in performance over pre-existing approaches.

Keywords Periocular biometrics · Deep convolutional neural network · Handcrafted features · Non-handcrafted features · Gender information · COVID-19

1 Introduction

The present situation of COVID-19 pandemic has challenged the world in all dimensions and no one can predict when the human kind will completely get rid of this virus. Considering its long-term effect, we as a human being must understand and learn to live with this virus. But to combat this situation and make our daily lives safer we can rely on technology and primarily on “contact less” technology. Under COVID 19 crisis, biometric systems which are considered as the spinal bone of any country’s security system is facing a lot of challenges. Almost all the organizations have stopped the use of contact based biometric systems, which can be a major reason to spread corona virus. Finger print biometric authentication systems which are implemented

from small organization to the place where security is on top priority, will no more be considered as a safer option for authentication. Solution is Face or Iris biometric authentication system but even they face certain challenges. The use of face mask that now become an essential part of our daily lives is the major hurdle in face biometric as it can deteriorate the performance of the system. Whereas iris biometric system requires high user cooperation. Under this situation, the use of periocular region as biometric trait is a better solution as compared to other biometric traits. Periocular region refers to the periphery of eyes which contains eye, eyebrow and pre-eye orbital region as illustrated in Fig. 1.

Periocular region works well even if the face of the subject is occluded because of face mask or veil and if the subject is not cooperative. Figure 2 shows some real-life examples where the use of periocular region for authentication is extremely useful.

Periocular region which can be considered as only 20 to 30% of the face area, itself is a small region of interest. Partial occlusion of such small size region may also deteriorate the performance of the system. The reason of the occlusion may be the use of eyeglasses, partial closure of

✉ K. R. Seeja
seeja@igdtuw.ac.in
Punam Kumari
punam_taurus@hotmail.com

¹ Department of Computer Science and Engineering, Indira Gandhi Delhi Technical University for Women, Delhi, India

Fig. 1 First row: Periocular region contains both eye; Second Row: Periocular region contains separate Left/Right eye

Fig. 2 Example of real-life scenarios

eyes or pose variation. Based on the literature, most of the researchers extracted either handcrafted features (Raja et al. 2014; Karahan et al. 2014; Mahalingam and Ricanek 2013; Yadav et al. 2015; Bakshi and Manjhi 2015) or non-handcrafted features (Zhao and Kumar 2016; Kandaswamy et al. 2017) from small periocular region of interest for matching, whereas the use of semantic information of the subject (gender, ethnicity, race etc.) was almost negligible.

In the proposed work, we have given equal weightage to all three features (handcrafted, non-handcrafted and semantic information) and combined them to create a single feature vector. The single feature vector extracted from the input images were fed to multiclass support vector machine classifier for classification. The significant contributions of this research are,

1. Proposed a feature fusion approach of handcrafted, non-handcrafted features and semantic information for extracting features from the small periocular region.
2. Created a highly robust contactless biometric system whose performance did not affect even when the person is wearing mask.
3. Evaluated the performance of the proposed system in three non-ideal conditions.

- Subject is wearing glasses.
- Eye portion is masked or hidden.
- Pose variation.

2 Background

Park et al. (2009) proposed periocular region as a biometric trait for person identification and provided a new direction to use periocular region based biometric system as a stand-alone modality or as a supporting biometric trait for other biometric systems. Later, they extended (Park et al. 2011) their proposed concept to examine the effectiveness of periocular region for nonideal scenarios such as pose variation, masking of critical eye components and inclusion of eyebrow in periocular region of interest. The outcome of their experiments provided a strong support to develop an effective periocular biometric system to use in the scenario where other biometric traits, for example, face and iris biometric are not completely useful.

In recent years, continuous research efforts (Alonso-Fernandez and Bigun 2016; Kumari and Seeja 2019; Nigam et al. 2015) were made to establish periocular region as a strong biometric trait. In early days, researchers used hand-crafted features for periocular image matching. Hand crafted features are broadly classified into two categories (1) Global feature descriptors and (2) Local feature descriptors. Global feature descriptor consider image as a whole and create a single feature vector for the whole image whereas local feature descriptor divides the image into patches (group of pixels), create feature vector for every patch and finally combine them to create a single feature vector.

Global feature descriptors can be further classified into three categories (1) texture-based feature descriptor (2) color-based feature descriptor and 3) shape-based feature descriptor. Raja et al. (2014) and Mahalingam et al. (2013) used different texture-based feature descriptor such as Binary Statistical Image Feature (BSIF) and Local Binary Pattern (LBP) respectively for matching. The primary advantage of texture-based feature descriptor is that they are easy to implement with very low computational cost but they are highly sensitive to noise, blurring and rotation in images. Woodard et al. (2010) used Color based feature descriptor for periocular image matching and obtained noticeable recognition accuracy. The primary limitation of color-based feature descriptor is that it works well only if the input image has uniform color distribution. Proenca et al. (2014) and Le et al. (2014) considered shape based feature descriptors such as shape of eye lid and shape of eyebrow respectively for periocular image matching. Shape based features are easy to visualize and implement and they are very helpful when contour of the shape is more important as compare to its inner

content, but these features are highly sensitive to noise and shape changes. One more disadvantage of shape-based features is that some time it is very hard to distinguish shape of the contour with background color such as for dark skinned people it is almost impossible to extract shape of eyebrow from face images.

Another category of handcrafted features is local feature descriptor. Hollingsworth et al. (2011) and Rattani et al. (2017) used Histogram of Oriented Gradients (HOG) whereas Karahan et al. (2014) implemented Scale Invariant Feature Transform (SIFT) and Speeded Up Robust Features (SURF) with their variants as local feature descriptors and obtained remarkable recognition accuracy. Later, it is found that (Alonso-Fernandez and Bigun 2014; Karahan et al. 2016; Chen et al. 2017) the presence of non-ideal images can degrade the overall performance of the system. They analysed various non-ideal cases such as problem of occlusion which may occur because of the presence of glasses, shadow on face or accidental closer of eyelids because of natural eye movement by the subject and concluded that even occlusion of 5 to 10 percent of small area of periocular region can reduce the final performance of the system. To handle the problem of pose variation, some researchers (Park et al. 2011; Castrillón-Santana 2016) used different handcrafted features and obtained noticeable results but also concluded that their methods need further improvement. With the emergence of deep learning concepts, the focus of the researchers (Zhao and Kumar 2016, 2018; Tiong et al. 2019) has been moved to non-handcrafted features and obtained visible improvement in performance of periocular biometric systems.

Tiong et al. (2019) proposed a handcrafted and non-handcrafted feature fusion approach and created a dual stream CNN in which each stream contains eight convolutional layers, two fully connected layers and a late fusion layer with shared weights and parameters. In their fusion approach, they fed a combination of RGB data and OCLBP feature to each stream of CNN for periocular recognition and demonstrated the effectiveness of fusion of multi-features instead of using only raw data as input. The drawback is that this approach is limited to the subject wearing glasses. Raffei et al. (2019) implemented a score fusion approach and Rotation invariant LBP to extract textural features and color moment for color feature extraction. With the use of CHI square and Euclidean distance for matching of templates generated through texture and color feature respectively, they generated two different matching scores. The matching scores were then fused to obtain the final matching accuracy. The drawback of this approach is that in order to use Color Moment, the input images were mapped from RGB to HSV space because of that the images were affected by the problem of non-removable singularities. Moreover, the HSV color space is also low sensitive to small color changes

which may remove micro features of images and degrade the performance of the system. Kumar et al. (2019) used a well-known handcrafted feature known as Local Binary Pattern to match periocular images extracted from images of same subject at different ages. To improve the performance of the system, they have implemented an image enhancement and denoising technique known as self enhancement quotient with discrete wavelet transform. At the first stage, the method looks very simple and result obtained are good but the paper lacks a discussion on the complexity to implement DWT which require more memory and process cycle and can make the whole process computationally very costly. Bakshi et al. (2018) implemented a feature reduction approach to extract Phase Intensive Local Patterns (PILP) with 20 percent reduction in extracted features known as R-PILP. With Nearest Neighbor classifier they obtained remarkable recognition accuracy with less computation time. The proposed method works well but requires very careful approach at the time of feature reduction otherwise wrong feature reduction may lead to high performance degradation in terms of final accuracy.

Periocular biometrics also played a major role in the area of soft biometric classification. Several researchers considered features from periocular region to identify subject's gender (Castrillón-Santana 2016), ethnicity or race (Chen et al. 2017) etc. Zhao and Kumar (2016) proposed a novel idea to use semantic information in recognition. They added an additional CNN branch trained with gender information to an already existing CNN. This model outperforms and obtained noticeable recognition accuracy of ~92% in periocular recognition. As an extension of their work (Zhao and Kumar 2018), they created a novel deep learning architecture which includes attention model concept. They exploited the concept of visual attention to emphasize the critical components of periocular region such as shape of the eyebrow and shape of the eye. Their model performed better and acquired improved recognition accuracy as compared to other works in literature. Based on their work they infer that eyebrow and eye shape are highly feature rich regions and the inclusion of them in periocular region of interest may improve the discriminating power of the model.

Motivating from the above facts (degradation in accuracy because of non-ideal conditions) and by considering the popularity of handcrafted, non-handcrafted and semantic information in the area of biometrics authentication, this paper proposes an ensemble feature vector approach which is a combination of handcrafted features HOG, non-handcrafted features extracted using pretrained CNN models and gender related features extracted using CNN to create a highly robust periocular biometric system for the authentication of the images captured in various non-ideal scenarios.

3 Materials and methods

3.1 Databases

In this study, three different benchmark databases were chosen for assessing the effectiveness of the proposed feature fusion approach. UBIPr database created by Padole and Proenca (2012), Ethnic Ocular database created by Tiong et al. (2019) and Color FERET database created by Philips et al. (2000) has been selected to analyse the effect of glasses, masked eye portion and pose variation. Figure 3 shows the example images from all of the three databases.

3.1.1 UBIPr database

This database contains total 10,252 RGB periocular images from 344 subjects in.bmp format with 501×401 pixels of image resolution. Images in this dataset were captured from 4 m to 8 m distance to include distance variability, but in controlled environment. Dataset also contains metadata files for each image which include coordinates of canthus points, center of the iris, end points and mid points of eyebrow as well as information about gender, level of pigmentation and angle of gaze etc. Some of the images in the dataset suffer from the problem of occlusion because of hair or eye glasses and variation in pose because of tilted head.

3.1.2 Ethnic ocular database

This database consists of two folders namely Train_data and Test_data. Train_data folder contains 70,142 RGB

periocular images from 623 subjects and Test_data folder contains 15,262 RGB periocular images from 411 subjects. All the images are stored in.JPG format with 80×80 pixels image resolution. Folder also contains metadata files which include information about name, nationality, ethnicity, gender, profession and number of pair images available for the subject. All the images were collected in wild with different pose, appearance and illumination variation.

3.1.3 Color FERET database

This database contains total 11,338 RGB periocular images from 994 subjects in.ppm format with 256×384 pixels of image resolution. Dataset also contains metadata files for each image which include coordinates of center of the iris, nose and mouth as well as information about gender, race, expressions etc. Metadata also includes information about occlusion because of hair or eye glasses and variation in pose etc.

3.2 Region of interest (ROI) extraction

The key issue to implement periocular region based biometric authentication system is to extract potential region of interest. The general approach (Park et al. 2009, 2011) is to use the coordinate of iris center as a reference point to extract periocular region of interest. In the proposed study, for the FERET database, same strategy (use of iris center) is implemented to extract periocular region of interest. The coordinates of iris center of images in FERET dataset is provided in its metadata. Considering the center of iris as center of periocular region of interest, a rectangular periocular region of 64×64 size is extracted from the face images for both left and right eye respectively.

Even though the above method is widely used in periocular literature, it does not work well if the eyes of the subjects are fully or partially closed. Hence an algorithm suggested by Liu et al. (2017) is used for the extraction of ROI from images of UBIPr and Ethnic Ocular datasets. This algorithm considers the canthus points as reference points (Fig. 4). The coordinates of canthus points of the images in UBIPr dataset is provided in its metadata but it was not provided in the

Fig. 3 Example images **a** UBIPr database **b** ethnic ocular database **c** color FERET Database

Fig. 4 Illustration of medial/inner and lateral/outer canthus point

metadata of ethnic ocular dataset. Hence, these points were extracted manually from the images in ethnic ocular dataset.

The algorithm is summarized in Algorithm 1.

Algorithm 1 **Input:** (x_1, y_1) are the coordinate for medial canthus points and (x_2, y_2) are the coordinate for lateral canthus points.

Output: Rectangular region of interest

Step 1: Calculate Euclidean distance between medial and lateral canthus points

$$D((x_1, y_1)(x_2, y_2)) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (1)$$

Step 2 Compute $L_p = (L_{px}, L_{py})$, midpoint of the line which connects medial and lateral canthus points.

$$L_{px} = \frac{(x_1 + x_2)}{2} \quad (2)$$

$$L_{py} = \frac{(y_1 + y_2)}{2} \quad (3)$$

Step 3: Compute the top left corner point (x_3, y_3) and bottom right corner (x_4, y_4) of rectangular ROI

$$(x_3, y_3) = (L_{px} - a \times D, L_{py} - b \times D) \quad (4^*)$$

$$(x_4, y_4) = (L_{px} + a \times D, L_{py} + b \times D) \quad (5)$$

Where a and b are constants whose values are chosen to cover the required periocular region

Step 4 Use the calculated points (x_3, y_3) and (x_4, y_4) to extract rectangular ROI.

3.3 Contrast limited adaptive histogram equalization method (CLAHE)

In order to make the proposed approach capable to handle poor quality images (images with low illumination and contrast or captured in dim lighting condition), the input images are first pre-processed using CLAHE (Zuiderveld 1994). It is an advancement of Adaptive Histogram Equalization (AHE) which was proposed by Pizer et al. (1987). The two reasons to use CLAHE rather than other histogram equalization approach are

1. Simple histogram equalization approach enhances the contrast of images globally whereas CLAHE works by dividing the images into number of tiles i.e. it enhances

-1	0	1
-1	0	1
-1	0	1

Fig. 5 Kernels to calculate horizontal and vertical gradient value in HOG

the local contrast value to highlight the edges and boundaries of the images.

2. Simple histogram equalization methods generally suffer with the problem of over amplification of noise whereas CLAHE limits the over amplification of noise by imposing a clip limit on the image histogram generated by simple Histogram Equalization methods.

To implement CLAHE one has to fine tune two important hyperparameters: Clip limit and Number of tiles.

The CLAHE algorithm is summarised in Algorithm 2.

Algorithm 2 **Step 1:** Divide the input image into $N \times N$ sub images. [N : number of tiles].

Step 2: Calculate the histogram of each sub image and calculate its high peak intensity value.

Step 3: Calculate a threshold to use as clip limit between 0 to high peak intensity value.

Step 4: For each bin in the histogram

if (histogram bin > clip limit)

{clip the histogram,

collect and redistribute all the clipped pixels uniformly to all the bins and obtain renormalized clipped histogram}

Step 5: Obtain the normalized histogram for each sub images with enhanced contrast. Combine all the neighbouring sub images using bilinear interpolation to obtain contrast enhanced image.

3.4 Histogram of gradient (HOG)

In the proposed approach HOG is used to extract the hand-crafted features. HOG is a feature descriptor which is rotation invariance (since it is working on gradients) and its computational complexity is very low. HOG is selected in this study due to its ability to automatically handle some non-ideal situations. Feature extraction process using HOG is summarized in Algorithm 3.

Fig. 6 Proposed CNN architecture to extract gender information

Algorithm 3 **Input:** Image Dataset

Output: HOG feature vectors

Step 1: Consider an image ‘I’ from dataset.

Step 2: Convert the image into grayscale image.

Step 3: Calculate horizontal and vertical gradient value for every pixel of the image using the kernels as shown in Fig. 5.

Step 4: Divide the image into adjacent and non-overlapping cells of $p \times p$ ($p = 4$) pixels, compute the histogram of orientation of gradients and binned them into ‘B’ bins ($B = 9$).

Step 5: There may be few pixels in the image whose orientation value may be close to bin boundary so, they might contribute to different bin also. To handle this situation, HOG uses weighted voting using bilinear interpolation and make fraction of the pixel’s gradient magnitude contribute to two bins.

Step 6: Group the cells into non-overlapping blocks which contained normalized gradient histogram.

Step 7: Normalized the block feature vectors and compute the HOG feature vector for the block.

Step 8: Concatenate all the features obtained from all the blocks to compute HOG feature vector for the image.

Step 9: Repeat step 1 to 8 for all the images in dataset.

3.5 Pretrained convolutional neural network (CNN) models

Deep learning is a technique used to obtain feature sets directly by observing the input images. In this study, deeper layers of seven off the shelf CNN models are used as feature extractors. A brief description of the CNN models used in the proposed study is given below:

AlexNet (Krizhevsky et al. 2012): This CNN is 25 layers deep and consists of 5 convolutional layers.

GoogLeNet (Szegedy et al. 2015): It is a 144 layers deep network, also known as Inception V1. It consists of 9 inception modules and 22 convolutional layers. Inception modules are nothing but a combination of different size filters connected at the same level. Different size filters allow multi-level feature extractions at same level.

ResNet 50 and ResNet 101 (He et al. 2016): Residual-Net50 is 177 layers deep and consists of 50 convolutional layers whereas ResidualNet 101 is 347 layers deep and consists of 101 convolutional layers. It works on the idea of Residual Block which uses ‘skip connection’ by skipping the mid-way layers to avoid the problem of vanishing gradients.

VGG16 and VGG19 (Simonyan and Zisserman 2014): VGG16 is 41 layers deep and consists of 16 convolutional layers whereas VGG19 is 47 layers deep and consists of 19 convolutional layers. The concept of VGG Net comes as an improvement of AlexNet. In VGG Net, large size filters of AlexNet were replaced by small size filters to make VGG Net deeper and to make it capable to learn more complex features from images.

Fig. 7 Proposed methodology

Inception V3 (Szegedy et al. 2016): consists of 48 layers and works on the concept of factorizing convolution. Factorizing convolution is a concept which is used to reduce the parameter without affecting the network efficiency and save the network to become over fitted.

3.6 Proposed CNN model to extract semantic information

In the proposed feature fusion-based approach along with hand crafted and non-handcrafted features, model is explicitly assisted with gender information. To extract gender information, a five-layer CNN classifier model is proposed for binary classification (Male/Female). The detailed layer configuration and their sequencing are shown in Fig. 6.

This CNN is trained using gender information of the subject instead of subject identity. The training labels of UBIPr, Ethnic Ocular dataset and Color FERET are extracted from its metadata as gender information is included in the meta-data. The detailed layer configuration for the proposed CNN architecture is given below:

1. First two Convolutional layers (Conv 1 and Conv 2) each contains 16 filters of size 3×3 with stride of one pixel which is followed by Max Pooling Layer (Max Pooling Layer 1 and Max Pooling Layer 2) with stride of 2 pixels
2. Next two Convolutional layers (Conv 3 and Conv 4) each contains 32 filters of size 3×3 with stride of one pixel which is followed by Max Pooling Layer (Max Pooling Layer 3 and Max Pooling Layer 4) with stride of 2 pixels.
3. The last Convolutional layers (Conv 5) contains 64 filters of size 3×3 with stride of one pixel which is followed by a Max Pooling Layer (Max Pooling Layer 5) with stride of 2 pixels.

With ReLU as activation function, this architecture contains one fully connected layer (with output size 2), one Soft-Max layer and a classification layer to classify images based on the gender (male/female). The features extracted from convolutional layer 5 (Conv 5) is used as gender information.

3.7 Support vector machine (SVM)

In the proposed work, multiclass Support Vector Machine classifier (Zhang 2012) is implemented for classification. The basic idea behind SVM is to consider a hyperplane in n -dimensional space which should be able to separate the data points in their potential classes. The data points which are near to hyperplane are known as support vectors. Up to a greater extent, efficiency of SVM classifier depends on kernel function. The task of kernel function is to map input space into higher dimensional space. For the proposed study, we have used RBF (Radial Basis Function) Kernel.

3.8 Proposed feature fusion approach

The methodology followed in this research is illustrated in Fig. 7. In the proposed work input images from three different databases UBIPr, Ethnic Ocular and Color FERET are used for training as well as testing purposes. The first step of the proposed approach is the extraction of optimum periocular region of interest. UBIPr and Ethnic ocular dataset consists of periocular images whereas Color FERET dataset contains face images. After that CLAHE was used to improve the quality of low illumination and/or low contrast images.

Three different types of features, handcrafted features (HOG), non-handcrafted features (using pretrained CNN) and gender related features (using a 5-layer CNN), are extracted from periocular region. To combine the features

Table 1 Database statistics

Database Name	TrainingDataset	ValidationData-set	TestDataset
Experiment 1: When subjects are wearing glasses			
UBIPr	7527	1720	1005
Ethnic Ocular	8000	2000	500
Color FERET	8000	2772	800
Experiment 2: When eye portion is manually masked			
UBIPr	5025	2152	3075
Ethnic Ocular	12,800	3200	4000
Color FERET	7405	1852	2315
Experiment 3: Images with pose variation			
UBIPr	3000	1252	3000 (30 degree pose variation) 3000 (-30 degree pose variation)
Color FERET	5000	572	1800 (15 degree pose variation) 3000 (22.5 degree pose variation) 1200 (45 degree pose variation)
Experiment 4: Recognition accuracy on complete database			
UBIPr	6025	1152	3075
Ethnic Ocular	14,800	1200	4000
Color FERET	7405	1852	2315

extracted using three different methods, all the three feature vectors are concatenated to create an ensemble feature vector. Let F_h , F_n and F_g denote feature vectors corresponds to handcrafted feature, non-handcrafted feature and gender feature. The ensemble feature vector obtained by concatenating all the three feature vectors is,

$$F_{\text{ensemble feature vector}} = \text{concat}[F_h, F_n, F_g] \quad (6)$$

Final feature vector was fed into support vector machine classifier for classification of images and to predict the class labels.

3.9 Implementation

This work is implemented on MATLAB r2019a. The system specification used in this work is Intel Core i7-8750H GPU Processor @ 2.2 GHz, 8 GB DDR4 RAM with Windows 10 operating system. Three different experiments were conducted on three different datasets. The dataset statistics is provided in Table 1.

The periocular ROIs are extracted by implementing the algorithms mentioned in Sect. 3.2. The values of constants a and b in algorithm 1 are calculated through empirical evaluation so that the selected periocular region covers all relevant periocular information and ignore all irrelevant information such as nose, forehead etc. For the ethnic ocular dataset, $a=1.2$ and $b=0.8$ and for UBIPr dataset, $a=0.9$ and $b=0.4$. Then the images are preprocessed by implementing the CLAHE algorithm. To implement CLAHE two important hyperparameters need to be carefully tuned: Clip

limit and Number of tiles. For this study, the number of tiles used for all the three databases are 4×4 whereas clip limit used for UBIPr and Ethnic ocular database is 0.05 and for COLOR FERET database is 0.04 (calculated through empirical evaluation). Figure 8 illustrate a comparison of CLAHE and simple AHE on example images from the three different databases. As shown in the Fig. 8, the application of CLAHE prominently highlighted the micro features of images which can automatically enhance the discrimination ability of the proposed approach.

After the preprocessing, three different types features were extracted. For hand crafted feature extraction, HOG is implemented. After implementation of HOG, a handcrafted feature vector of size 8100 components was obtained for each input image. For non-handcrafted feature extraction, seven pre-trained CNNs are used. After analyzing features extracted from different initial, middle and last layers of the pretrained networks, the non-handcrafted features extracted from penultimate layers are selected. It is found that the penultimate layers contain maximum relevant information to distinguish class labels. Table 2 shows non-handcrafted feature vector size obtained using different pretrained networks.

For semantic information or gender related feature extraction, a five-layer CNN proposed in Sect. 3.6 has been implemented. The proposed CNN is trained on images extracted from three different databases. After exploiting features from different layers of the proposed CNN architecture, the features extracted from convolutional layer 5 (Conv 5) is used as gender information for the proposed feature fusion work. This feature vector contains 1024 components for each input image. Performance of proposed CNN architecture for

Fig. 8 Illustration of CLAHE vs AHE **a** UBIPr database **b** ethnic ocular database and **c** color FERET database

Table 2 Size of non-handcrafted feature vectors

Pretrained CNN models	Name of the Layer (from which non-handcrafted features were extracted)	Size of final non-handcrafted feature vector
AlexNet	Fc8	4096
GoogleNet	Loss3-Classifier	1024
ResNet50	Fc1000	2048
ResNet101	Fc1000	2048
VGG16	Fc8	4096
VGG19	Fc8	4096
InceptionV3	Predictions	2048

gender classification on three different databases is provided in Table 3.

After evaluating a lot of different set of hyper parameters such as different learning rates (0.0001, 0.0003, 0.0005, and 0.0008), minibatch size (10–64) and epoch (starting from 2 to 10) the final parameter specification used to train the proposed CNN model for gender related feature extraction and the hyperparameters used to train the pretrained CNN models are given in Table 4.

Table 3 Performance of proposed CNN architecture in gender classification

Database name	Male vs female image ratio in dataset	Accuracy
UBIPr	Male: 54.4%; Female: 45.6%	95%
Ethnic Ocular	Male: 62.06% Female: 37.94%	92.5%
Color FERET	Male: 59.61% Female: 40.39%	94.4%

Table 4 Parameter description (for pretrained and proposed CNN models)

Parameter	Value (pretrained CNN models)	Value (proposed CNN model)
Learning algorithm	Stochastic Gradient Descent Momentum (SGDM)	Stochastic Gradient Descent Momentum (SGDM)
Momentum	0.9000	0.9000
Learning Rate	0.0001 (AlexNet and VGG models) 0.0003 (ResNet and inception models)	0.0008
Weight factor	10	10
Bias factor	10	10
L2 regularization	1.0000e-04	1.0000e-04
Gradient threshold method	'l2norm'	'l2norm'
No of Epochs	6	8
Mini Batch size	10	16

For the proposed study, SGDM is used as learning algorithm as it is computationally very fast and since it updates parameters very frequently, it converges fast. The value of momentum lies between 0 and 1 and it help to jump over from local-minima. Large value of momentum provides speed and quick convergence but if momentum is large then learning rate should be small. Hence, in this study good results are obtained with small learning rate such as 0.0001, 0.0003 and 0.0008. If value of momentum and learning rate both is large then algorithm can skip local minima and degrades its performance. L2 regularization also known as 'weight decay' reduces the possibility of overfitting by taking care of the values of bias and weights.

Then, all the three feature vectors (handcrafted, non handcrafted and gender) are concatenated to create the final feature vector. Final size of feature vector after concatenation is provided in Table 5.

In order to evaluate the efficiency of the proposed ensemble feature vector, the multiclass SVM classifier is implemented. The best classification results were obtained with the RBF kernel.

4 Results and discussion

Since the proposed biometric system is a closed set identification system, the performance of the proposed system is evaluated using two appropriate performance metrics: Rank 1 recognition accuracy and Cumulative Match Characteristic curve.

Rank 1 Recognition Accuracy

Rank 1 accuracy tells that how many times the label predicted by the system is same as ground truth label and it is defined as:

$$\text{Rank 1 recognition accuracy} = \frac{\text{sum}(\text{predicted class label} = \text{true class label})}{\text{Total class label}} \times 100 \quad (7)$$

Table 5 Size of final feature vector

Ensemble feature vector	Size
HOG + AlexNet + Gender info	8100 + 4096 + 1024 = 13,220
HOG + VGG16 + Gender info	
HOG + VGG19 + Gender info	
HOG + GoogleNet + Gender info	8100 + 1024 + 1024 = 10,148
HOG + ResNet 50 + Gender info	8100 + 2048 + 1024 = 11,172
HOG + ResNet 50 + Gender info	
HOG + InceptionV3 + Gender info	

Cumulative Match Characteristics (CMC) Curve

It is used to assess the probability of identification by the system at various ranks and it plots the number of correct identifications occurred in top k matches or ranks.

4.1 Experiment 1: Subjects are wearing glasses

Now a days, a large part of population is wearing eye glasses which may hide a part of periocular region. Considering this fact, this experiment assess the accuracy of the proposed approach when subjects are wearing eye glasses. Rank1 recognition accuracy obtained by the proposed approach on UBIPr, Ethnic Ocular and Color FERET dataset is shown in Table 6 with best result highlighted in bold and italic. Comparison of Rank 1 to Rank 10 recognition accuracy using CMC curve is illustrated in Fig. 9.

Table 6 UBIPRVs Ethnic Ocular Vs Color FERET (subjects are wearing glasses)

Feature vector		Rank 1 accuracy (UBIPr) (%)	Rank 1 accuracy (Ethnic Ocular) (%)	Rank 1 accuracy (Color FERET) (%)
Hand crafted feature	HOG	77	54	54.5
Convolutional neural networks	Alex Net	91.22	64.10	61.75
	Google Net	91.1	70.55	68.62
	Resnet50	91.89	73.56	75.5
	Resnet101	95	73.88	77.87
	VGG16	88.33	70	72.62
	VGG 19	98.8	71.88	73.75
	Inception V3	89.33	69.05	69.87
CNN + HOG + Gender information (GI)	Alex Net + HOG + GI	92	70	78
	Google Net + HOG + GI	92	74.64	76.25
	Resnet50 + HOG + GI	92.33	78.22	82
	Resnet101 + HOG + GI	96.78	79.8	83.62
	VGG16 + HOG + GI	93.33	75.36	76.5
	VGG 19 + HOG + GI	98.9	76	78.62
	Inception V3 + HOG + GI	90.33	74	78

Fig. 9 CMC curve (subjects are wearing glasses): handcrafted vs non-handcrafted vs feature fusion approach **a** UBIPr **b** Ethnic ocular **c** color FERET

Table 7 UBIPr Vs Ethnic Ocular Vs Color FERET (eye portion is masked)

Feature vector		Rank 1 Accuracy (UBIPr) (%)	Rank 1 Accuracy (Ethnic Ocular) (%)	Rank 1 Accuracy (Color FERET) (%)
Hand crafted feature	HOG	68	59	62
Convolutional Neural Networks	Alex Net	62	59.68	59.16
	Google Net	67	63.71	64.86
	Resnet50	69	64.93	66.33
	Resnet101	71.2	65.75	68.73
	VGG16	72.6	63.21	67.23
	VGG 19	78	61	64.16
	Inception V3	65	63.94	65.48
CNN + HOG + Gender Information (GI)	Alex Net + HOG + GI	69	63.45	63.71
	Google Net + HOG + GI	73.5	69.34	70.53
	Resnet50 + HOG + GI	76.32	72.45	73
	Resnet101 + HOG + GI	80	75.42	77
	VGG16 + HOG + GI	78.54	64.48	70.45
	VGG 19 + HOG + GI	82	69.46	74.55
	Inception V3 + HOG + GI	74	68	70

4.2 Experiment 2: Occlusion because of masked eye region

This experiment is carried out to analyze the accuracy of a true periocular biometric system that considered only pre eye orbital region and disregard all the eye components. This may be interesting when subjects' eyes are closed or hidden intentionally. To evaluate this fact, this experiment considered images with eye portion such as sclera, iris and eye shape are manually masked or hide. Rank1 recognition accuracy obtained by the proposed approach on UBIPr, ethnic ocular and color FERET dataset is shown in Table 7 with best result highlighted in bold and italic. Comparison of Rank 1 to Rank 10 recognition accuracy using CMC curve is illustrated in Fig. 10.

4.3 Experiment 3: Pose variation

This experiment is implemented to assess the efficiency of the proposed approach when images are suffered with pose variations. Here, for the training features are extracted only from frontal (0 degree pose variation) images, whereas for testing, images with different pose variation were used. Recognition accuracy obtained by proposed approach on UBIPr and color FERET dataset is shown in Table 8 with best result highlighted in bold and italic. Comparison of Rank 1 to Rank 10 recognition accuracy using CMC curve is illustrated in Fig. 11.

4.4 Experiment 4: Complete UBIPr, Ethnic Ocular and color FERET database

For fair comparison with pre-existing approaches, the recognition accuracy of the proposed feature fusion-based approach is evaluated on complete UBIPr, Ethnic Ocular and color FERET databases and is shown in Table 9 with best result highlighted in bold and italic. Comparison of Rank 1 to Rank 10 recognition accuracy using CMC curve is illustrated in Fig. 12.

4.5 Comparison with pre-existing work in literature

Table 10 shows a comparison between the proposed model and state of the art models. For this, the best result obtained by the proposed feature fusion approach on complete UBIPr, Ethnic Ocular and Color FERET database is used.

From the results shown in Table 10, it is observed that the proposed approach outperforms the state-of-the-art models in periocular authentication. This observation suggests that the concept of combining handcrafted and non-handcrafted features with semantic information is successfully contributing useful information and thus improving the overall discriminative capability of the periocular system.

Some important findings from all of the experiments are enumerated below:

1. The recognition accuracy obtained for first two experiments reveals that the proposed approach is performing low for ethnic ocular and Color FERET dataset as compare to UBIPr dataset. Reason for low performance

Fig. 10 CMC curve (eye portion is masked): handcrafted vs non-handcrafted vs feature fusion approach **a** UBIPr **b** Ethnic ocular **c** color FERET

Table 8 UBIPR vs color FERET (pose variation)

Feature Vector		UBIPr		Color FERET		
		30 degree	-30 degree	15 degree	22.5 degree	45 degree
Hand crafted Feature	HOG	74	76	72.72	67.73	67
Convolutional Neural Networks	Alex Net	85	85	80.89	72.65	73.3
	Google Net	81	83	79	71.77	69.2
	Resnet50	93	90	81.83	76.11	75.89
	Resnet101	83	90	81.83	77.22	77
	VGG16	88	90	79.56	74.78	74.45
	VGG 19	94	96	79	76.10	75
	Inception V3	82	82	79.56	70	68.73
	Inception V3 + HOG + GI	89.83	88.46	82	75	70
CNN + HOG + Gender Information (GI)	Alex Net + HOG + GI	86.83	86.73	82	76	75
	Google Net + HOG + GI	88.54	87.50	83.4	77	72.89
	Resnet50 + HOG + GI	94.12	92.50	85	79.8	78
	Resnet101 + HOG + GI	94.73	93.63	86.4	82.56	80
	VGG16 + HOG + GI	91.50	92.12	83.89	76.75	74.82
	VGG 19 + HOG + GI	95.76	96.50	83.89	76.55	76.75

Fig. 11 CMC curve (pose variation): handcrafted vs non-handcrafted vs feature fusion approach **a** UBIPr **b** Color FERET

Table 9 UBIPr Vs Ethnic Ocular Vs Color FERET (complete database)

Feature vector		Rank 1 Accuracy (UBIPr) (%)	Rank 1 Accuracy (Ethnic Ocular) (%)	Rank 1 Accuracy (Color FERET) (%)
Hand crafted feature	HOG	79	70	75
Convolutional Neural Networks	Alex Net	79.80	72.71	77.50
	Google Net	87.20	73.19	72.22
	Resnet50	88	74.66	75.87
	Resnet101	86	81.71	79.04
	VGG16	89.20	79.72	78.84
	VGG 19	89.50	80.72	78.94
	Inception V3	82.83	74.32	76.22
CNN + HOG + Gender Information (GI)	Alex Net + HOG + GI	86.83	80.32	77.56
	Google Net + HOG + GI	90.67	86.73	78.77
	Resnet50 + HOG + GI	91.83	85.77	85.52
	Resnet101 + HOG + GI	92.83	87.83	86.56
	VGG16 + HOG + GI	90.50	85.63	82.13
	VGG 19 + HOG + GI	93.83	86.46	85.27
	Inception V3 + HOG + GI	87.67	85.12	78.52

Fig. 12 CMC curve (complete database): handcrafted vs non-handcrafted vs feature fusion approach **a** UBIPr **b** Ethnic ocular **c** color FERET

may be the high degree of randomness in images used in ethnic ocular dataset and images with high pose variation and varying age in Color FERET dataset.

- Experimental results illustrate that for periocular matching, the combination of handcrafted, non-handcrafted features and semantic information obtained remarkable recognition accuracy for all the non-ideal scenarios and with all pretrained CNNs.
- It is found that the Eye shape, corner points etc. are critical components for periocular matching since masking of eye region significantly degrade the recognition accuracy whereas presence of eyeglasses does not make much effect on the performance of the system.
- It is also found that the combination of handcrafted and non-handcrafted features neutralizes the effect of pose variation up to a greater extent and obtained good results as compared to other methods in literature.
- From Figs. 9, 10, 11 to 12, it can be easily observed that proposed fusion approach outperform handcrafted and non-handcrafted features in every scenario from Rank 1 to Rank 10 recognition accuracy.

Table 10 Comparison with pre-existing approaches

Methodology		Rank1 Accuracy(%)
UBIPr database		
Proposed approach		93.83
Tiong et al. (2019)	OCLBP+ dual stream CNN	91.28
Raffei et al. (2019)	Rotation Invariant Uniform LBP	85.7
Ethnic Ocular Database		
Proposed approach		87.83
Tiong et al. (2019)	OCLBP+ dual stream CNN	85.03
Color FERET Database		
Proposed approach		86.56
Kumar et al. (2019)	LBP+ Chi Square distance	80.73
Bakshi et al. (2018)	Reduced-PILP	85.64

*OCLBP orthogonal combination of Local binary pattern, R-PILP reduced phase intensive local pattern

5 Conclusion and future work

In order to cope up with the COVID-19 pandemic, this research proposes a robust periocular region based biometric authentication system. The proposed system uses handcrafted and non-handcrafted features assisted with semantic information for image matching. The proposed model gained remarkable improvement in recognition accuracy for the non-ideal images, captured in three nonideal scenarios: subjects are wearing eyeglasses (which can partially occlude the periocular region), masking of eye region (effect of partial/full closure of eyes) and pose variation (images with tilted head). The experimental results provide very strong support to the proposed feature level fusion approach which is unique of its type and to the best of our knowledge nothing like this is ever proposed by anyone in the area of periocular biometrics. The future work aims to improve the performance of the proposed approach by assisting it with some more semantic information and binding it with attention mechanism on some critical components in a single learning model.

References

- Alonso-Fernandez F, Bigun J (2014) Eye detection by complex filtering for periocular recognition. In: International workshop on biometrics and forensics, Valletta, Malta. <https://doi.org/10.1109/IWBF.2014.6914250>
- Alonso-Fernandez F, Bigun J (2016) A survey on periocular biometrics research. *Pattern Recognit Lett* 82:92–105. <https://doi.org/10.1016/j.patrec.2015.08.026>
- Bakshi S, Sa PK, Majhi B (2015) A novel phase-intensive local pattern for periocular recognition under visible spectrum. *Bio-cybern Biomed Eng* 35(1):30–44. <https://doi.org/10.1016/j.bbe.2014.05.003>
- Bakshi S, Sa PK, Wang H, Barpanda SS, Majhi B (2018) Fast periocular authentication in handheld devices with reduced phase intensive local pattern. *Multimed Tools Appl* 77(14):17595–17623. <https://doi.org/10.1007/s11042-017-4965-6>
- Castrillón-Santana M, Lorenzo-Navarro J, Ramón-Balmaseda E (2016) On using periocular biometric for gender classification in the wild. *Pattern Recogn Lett* 82:81–189. <https://doi.org/10.1016/j.patrec.2015.09.014>
- Chen H, Gao M, Ricanek K, Xu W, Fang B (2017) A novel race classification method based on periocular features fusion. *Int J Pattern Recognit Artif Intell* 31(08):1750026. <https://doi.org/10.1142/S0218001417500264>
- He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778. <https://doi.org/10.1109/cvpr.2016.90>
- Hollingsworth KP, Darnell SS, Miller PE, Woodard DL, Bowyer KW, Flynn PJ (2011) Human and machine performance on periocular biometrics under near-infrared light and visible light. *IEEE Trans Inf Forensics Secur* 7(2):588–601. <https://doi.org/10.1109/TIFS.2011.2173932>
- Kandaswamy C, Monteiro JC, Silva LM, Cardoso JS (2017) Multi-source deep transfer learning for cross-sensor biometrics. *Neural Comput Appl* 28(9):2461–2475. <https://doi.org/10.1007/s00521-016-2325-5>
- Karahan S, Karaöz A, Özdemir OF, Gul AG, Uludag U (2014) On identification from periocular region utilizing sift and surf. In: 22nd European Signal Processing Conference, Lisbon, Portugal, pp 1392–1396
- Karahan S, Yildirim MK, Kirtac K, Rende FS, Butun G, Ekenel HK (2016) How image degradations affect deep cnn based face recognition. In: International Conference of the Biometrics Special Interest Group (BIOSIG). <https://doi.org/10.1109/BIOSIG.2016.7736924>
- Krizhevsky A, Sutskever I, Hinton GE (2012) Image net classification with deep convolutional neural networks. *Adv Neural Inf Process Syst*. <https://doi.org/10.1145/3065386>
- Kumar KK, Pavani M (2019) Periocular region-based age-invariant face recognition using local binary pattern. In: *Microelectronics, Electromagnetics and Telecommunications*. pp. 713–720. Springer, Singapore. https://doi.org/10.1007/978-981-13-1906-8_72
- Kumari P, Seeja KR (2019) Periocular Biometrics: A survey. *Journal of King Saud University Computer and Information Sciences*. <https://doi.org/10.1016/j.jksuci.2019.06.003>
- LeTHN, Prabhu U, Savvides M (2014) A novel eyebrow segmentation and eyebrow shape-based identification. In: IEEE International Joint Conference on Biometrics. pp. 1–8. <https://doi.org/10.1109/BTAS.2014.6996262>
- Liu P, Guo J, Tseng S, Wong K, Lee JD, Yao C, Zhu D (2017) Ocular recognition for blinking eyes. *IEEE Trans Image Process* 26(10):5070–5081. <https://doi.org/10.1109/TIP.2017.2713041>
- Mahalingam G, Ricanek K Jr (2013) LBP-based periocular recognition on challenging face datasets. *EURASIP J Image Video Process* 2013, 36. <https://doi.org/10.1186/1687-5281-2013-36>
- Nigam I, Vatsa M, Singh R (2015) Ocular biometrics: a survey of modalities and fusion approaches. *Inf Fusion* 26:1–35. <https://doi.org/10.1016/j.inffus.2015.03.005>
- Padole CN, Proenca H (2012) Periocular recognition: Analysis of performance degradation factors. In: 5th IAPR International Conference on Biometrics, New Delhi, India, pp 439–445. <https://doi.org/10.1109/ICB.2012.6199790>
- Park U, Ross A, Jain AK (2009) Periocular Biometrics in the visible spectrum: A Feasibility study. In: 3rd IEEE International Conference on Biometrics: Theory Applications and Systems Washington DC, USA, pp 153–158. <https://doi.org/10.1109/BTAS.2009.5339068>
- Park U, Jillela R, Ross A (2011) Periocular biometrics in the visible spectrum. *IEEE Trans Inf Forensics Secur* 6(01):96–106. <https://doi.org/10.1109/TIFS.2010.2096810>
- Phillips PJ, Moon H, Rizvi SA, Rauss PJ (2000) The FERET evaluation methodology for face recognition algorithms. *IEEE Trans Pattern Anal Mach Intell* 22(10):1090–1104. <https://doi.org/10.1109/34.879790>
- Pizer SM, Amburn EP, Austin JD, Cromartie R, Geselowitz A, Greer T, Zuiderveld K (1987) Adaptive histogram equalization and its variations. *Comput Vis Graph Image Process* 39(3):355–368. [https://doi.org/10.1016/S0734-189X\(87\)80186-X](https://doi.org/10.1016/S0734-189X(87)80186-X)
- Proenca H, Neves JC, Santos G (2014) Segmenting the periocular region using a hierarchical graphical model fed by texture/shape information and geometrical constraints. In: IEEE international joint conference on biometrics, pp 1–7. <https://doi.org/10.1109/BTAS.2014.6996228>
- Raffei AFM, Sutikno T, Asmuni H, Hassan R, Othman RM, Kasim S, Riyadi MA (2019) Fusion iris and periocular recognitions in non-cooperative environment. *Indonesian J Electr Eng Inf* 7(3):543–554. <https://doi.org/10.11591/ijeei.v7i3.114740>
- Raja KB, Raghavendra R, Busch C (2014) Binarized statistical features for improved iris and periocular recognition in visible spectrum. In: 2nd International Workshop on Biometrics and Forensics,

- Valletta, Malta, pp 1–6. <https://doi.org/10.1109/IWBF.2014.6914249>
- Rattani A, Reddy N, Derakhshani R (2017) Gender prediction from mobile ocular images: a feasibility study. In: IEEE international symposium on technologies for homeland security (HST), pp 1–6. <https://doi.org/10.1109/THS.2017.7943489>
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
- Szegedy C, liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A (2015) Going deeper with convolutions. In: IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, pp 1–9. <https://doi.org/10.1109/CVPR.2015.7298594>
- Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna ZB (2016) Rethinking the inception architecture for computer vision. In: IEEE Conference on Computer Vision and Pattern Recognition, pp 2818–2826. <https://doi.org/10.1109/cvpr.2016.308>
- Tiong L, Teoh ABJ, Lee Y (2019) Periocular recognition in the wild with orthogonal combination of local binary coded pattern in dual-stream convolutional neural network. *Appl Sci* 9(13):2709. <https://doi.org/10.3390/app9132709>
- Woodard DL, Pundlik SJ, Lyle JR, Miller PE (2010) Periocular region appearance cues for biometric identification. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition-Workshops, pp 162–169. <https://doi.org/10.1109/CVPRW.2010.5544621>
- Yadav AR, Anand RS, Dewal ML, Gupta S (2015) Multiresolution local binary pattern variants-based texture feature extraction techniques for efficient classification of microscopic images of hardwood species. *Appl Soft Comput* 32:101–112. <https://doi.org/10.1016/j.asoc.2015.03.039>
- Zhang Y (2012) Support vector machine classification algorithm and its application. In: Liu C, Wang L, Yang A (eds) Information computing and applications. Communications in Computer and Information Science. https://doi.org/10.1007/978-3-642-34041-3_27
- Zhao Z, Kumar A (2016) Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. *IEEE Trans Inf Forensics Secur* 12(5):1017–1030. <https://doi.org/10.1109/TIFS.2016.2636093>
- Zhao Z, Kumar A (2018) Improving periocular recognition by explicit attention to critical regions in deep neural network. *IEEE Trans Inf Forensics Secur* 13(12):2937–2952. <https://doi.org/10.1109/TIFS.2018.2833018>
- Zuiderveld K (1994) Contrast limited adaptive histogram equalization. *Graphics Gems IV*, Academic Press Professional, Inc. pp 474–485
- Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.