1212: DEEP LEARNING TECHNIQUES FOR INFRARED IMAGE/VIDEO UNDERSTANDING

Check for
updates

Enhanced near-infrared periocular recognition through collaborative rendering of hand crafted and deep features

Ritesh Vyas[1] (ORCID)

Abstract

Periocular recognition leverage from larger feature region and lesser user cooperation, when compared against the traditional iris recognition. Moreover, in the current scenario of Covid-19, where majority of people cover their faces with masks, potential of recognizing faces gets reduced by a large extent, calling for wide applicability of periocular recognition. In view of these facts, this paper targets towards enhanced representation of near-infrared periocular images, by combined use of hand-crafted and deep features. The hand-crafted features are extracted through partitioning of periocular image followed by obtaining the local statistical properties pertaining to each partition. Whereas, deep features are extracted through the popular convolutional neural network (CNN) ResNet-101 model. The extensive set of experiments performed with a benchmark periocular database validates the promising performance of the proposed method. Additionally, investigation of cross-spectral matching framework and comparison with state-of-the-art, reveal that combination of both types of features employed could prove to be extremely effective.

Keywords Near-infrared · Periocular · CNN · Biometrics

1 Introduction

Biometrics has been solving the personal authentication and identification tasks for long now. Among all the popular physiological biometric traits, periocular region has emerged as a potential trait [5]. The primary reason behind its increasing popularity is its ability to facilitate recognition at-a-distance and its viability to work with partially covered faces [33]. Both these attributes make periocular recognition a potential biometric contender, especially for public applications like surveillance. Moreover, in the recent times of pandemic, when

---

✉ Ritesh Vyas
   ritesh.vyas157@gmail.com

1   Lancaster University, Lancaster, UK

roaming around with uncovered faces could lead to high health risks, the need of a better periocular recognition system becomes even greater.

Periocular region being in the vicinity of the eye, comprises of several unique structures, like eyelid, eye shape and skin texture [13, 19]. These unique inclusions make the periocular region potential biometrics, which offers a trade-off between using the entire facial region or the iris [20, 21]. Owing to various advantages like less user-cooperation, no need of controlled lighting setup, periocular recognition has been researched heavily by the research community.

The periocular region is set to find profound utility in view of the recent distension of the Covid-19 pandemic. The natural reason behind this is the prevalent mandate of wearing face coverings like masks. The large-scaled use of face-masks has certainly raised questions on the utility of facial recognition, as majority of the discerning regions of face gets covered behind the masks. In view of this, periocular recognition can provide a subtle solution to identify the human beings. Therefore, the real-world applicability of the proposed periocular recognition approach stands high. Furthermore, the conjunction of hand-crafted and deep features paves a new way for enhanced and distinctive representation of the periocular images. Harnessing the detailed (through deep learning model) and textural (through hand-crafted descriptor) information of the periocular images facilitates the outstanding behaviour of the proposed method.

2 Related work

Numerous researchers have worked in the emerging field of periocular recognition in the last decade. Hollingsworth et al. [14] carried out an intensive study to identify the important features of the periocular region from near infrared (NIR) and visible wavelength (VW) spectrum, based on the interpretations of human and machine. Bhardwaj et al. [8] studied the use of periocular recognition in a scenario when iris recognition fails, by discussing a global descriptor along with the effects of capturing distances. Mahalingam and Ricanek [18] used local binary pattern (LBP) for periocular recognition from facial images. Smereka and Kumar [32] explored the good periocular regions by extracting features through probabilistic deformation models and m-SIFT approaches. Raja et al. [23] adopted three popular descriptors, namely SIFT, SURF and BSIF, for accomplishing periocular recognition on smartphone devices.

In another work, Raja et al. [25] investigated binarized statistical features (BSIF) for combined iris and periocular recognition. Proenca and Briceno [21] proposed a modified elastic graph matching (EGM) approach which was made more globally coherent by avoiding sudden angular changes and modelling non-linear distortions more faithfully. Santos et al. [27] studied about the cross-sensor recognition for combined iris and periocular recognition, where periocular features were extracted through LBP and histogram of gradient (HoG). Gangwar and Joshi [10] implemented a robust periocular recognition system via score-level fusion of local phase quantization (LPQ) and Gabor wavelet descriptors.

Sharma et al. [31] conducted investigations on cross-spectral periocular recognition by jointly training a neural network. Uzair et al. [33] investigated the potential of periocular recognition from RGB and NIR videos, along with hyperspectral image cubes, where PCA (principal component analysis) and LBP assisted features were processed through two-stage fusion to achieve improved results. Behera et al. [6] proposed an illumination normalization based cross-spectral periocular matching scheme, where features were extracted through

HoG. Additional details about some other existing periocular recognition approaches and related databases may be found in [4, 5].

Aginako et al. [1] completed an exhaustive study on extracting iris and periocular features through popular local descriptors like LBP, variants of LBP, LPQ, and Weber local descriptor (WLD). Ahmed et al. [2] explored the fusion of iris and periocular scores for mobile database, where periocular features were extracted via LBP. Kumar et al. [15] proposed non-overlapped interpolated LBP (iLBP) for periocular recognition, where histogram from non-overlapping sub-regions of the iLBP image were utilized as potential features.

The excessive use of LBP in the state-of-the-art for periocular recognition limits the full potential of periocular features. It remains sensitive to the noise due to its binary nature. Besides, it does not consider the effect of center pixel of a local patch because it only comprises of the difference of the neighboring pixels with the center pixel. Concurrently, other variants of LBP do have some or the other limitations in terms of their neighborhood or calculation of the individual bins. Other common descriptors like HoG and LPQ have high sensitivity to factors like rotation and blur, respectively. In order to overcome these limitations, the proposed approach utilizes a different hand-crafted descriptor which analyse the input images in a multi-resolution and multi-orientation manner. This procedure of the proposed descriptor enriches the extracted feature set and enhances the overall differentiability.

There have been various attempts from the researchers to examine the performance of deep features in the domain of periocular recognition. Raja et al. [24] proposed deep sparse filtering for smartphone-based periocular recognition, attributing to easy learning of the sparse filters via only one parameter which is the number of features. Zhang et al. [35] utilized a convolutional neural network (CNN) with 'maxout' layer and concatenation of iris and periocular features through the deep learning model itself. Luz et al. [17] explored the deep representation of periocular regions for the application of video surveillance, where VGG deep CNN was employed with transfer learning. Alahmadi et al. [3] effectively utilized the sparsity present in the activations of convolutional layers of CNN models and employs a sparsity augmented collaborative representation and classification.

Hernandez-Diaz et al. [12] evaluated multiple pre-trained CNNs in the problem of periocular recognition and claimed that ResNet101 yields superior results. Zhao and Kumar [36] proposed a improved periocular recognition approach where explicit attention was used to emphasize the important regions of the periocular image. Raja et al. [22] adopted a collaborative representation of smartphone-based periocular images using deep sparse features and deep sparse time frequency features.

In view of the above literature survey, it can be strongly inferred that there has not been any attempt from the research fraternity to investigate the fusion of traditional hand-crafted features and contemporary deep features for the problem of periocular recognition. Despite the outstanding performances of deep features for biometrics applications, there has always been a pressing need of improving their recognition rates and incorporating the interpretability measures. We believe that inclusion of hand-crafted features could cater to both these needs in an effective manner. The same has been reflected through the outcomes of current work.

This paper investigates the combination of hand-crafted and deep features for the problem of periocular recognition, especially for the near-infrared acquisition framework. For extracting the hand-crafted features, an effective approach based on partitioning of Gabor-

filtered images and calculation of local statistical measures, is adopted. Whereas, for the extraction of deep features, a deep convolutional neural network, named ResNet, is employed. Furthermore, individual and collaborative performances for both the NIR and VW acquisition frameworks are reported in the form of popular metrics and curves.

The rest of the paper is organized in the following manner. Section 2 details about the related work and Section 3 explains the preliminaries for both the hand-crafted and deep features. The details pertaining to the experiments conducted are presented in Section 4 and the paper is concluded in Section 5.

3 Methodology

The proposed approach leverage from the benefits of both the hand-crafted and deep feature descriptors. On one hand, hand-crafted features (HCF) enable us to understand the behavior of the multiscale and multi-orientation features of the periocular image, avoiding hyper-parameter tuning at the same time. On the other hand, deep features (DF) aid to the recognition accuracy by virtue of their distinctive representation. Figure 1 demonstrates the overall block diagram of the proposed collaborative approach. Each of the individual blocks is explained in subsequent subsections.

3.1 Hand-crafted features

There has been a great thrust in the research community for devising novel hand-crafted feature descriptors for biometric recognition. The advantage with hand-crafted descriptors is that one does not need to finetune a massive set of hyper parameters for them, as is usually desired for deep learning-based descriptors. Moreover, the hand-crafted descriptors are developed by using traditional, yet effective image representation means which have got the long-lasting support of established image processing techniques.

In this paper, the hand-crafted features (HCF) of the periocular images are captured through extracting the local statistical properties of the image at different levels of partitions. This work is inspired by our earlier work on iris recognition [34], where the said descriptor performed well for iris images from NIR as well as VW illuminations. In current work, the periocular images are first resized to one fifth of their original sizes and then partitioned at two different levels by having equal number of partitions at both the levels. The resizing is done to speed up the feature extraction process. Whereas, the equal number of partitions are chosen in order to attain uniform vertical and horizontal resolutions in all the sub-images
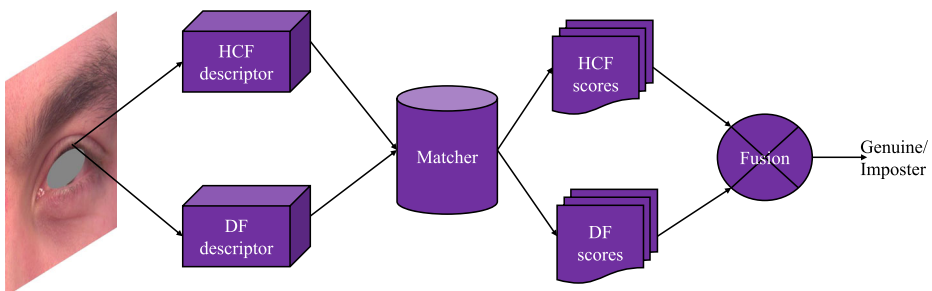

Fig. 1 Overall block diagram of the proposed approach

occurring at the same level of partitioning. It is because of these partitions, that the statistical measures (like mean and variance) exhibit the amount of texture variations occurring at the local level.

Another important aspect of the employed descriptor is the execution of image filtering via a 2D Gabor filter bank, which tends to highlight the multifaceted textural and spatial information present in the periocular image. The filter bank is designed to have varying scales and frequencies, which would aid the comprehensive representation of input image. The mathematical form of 2D Gabor filter bank can be expressed as:

$$G(p, q) = g\left(u, v, \varepsilon_p, \rho_p, \alpha_q\right) \tag{1}$$

where, p and q are known as the filter identifiers and stand responsible for selecting specified values of (ϵ, ρ) and α, respectively. Whereas, detailed expression for individual filters would be as follows:

$$g\left(u, v, \varepsilon_p, \rho_p, \alpha_q\right) = \frac{1}{\sqrt{2\pi \varepsilon_p^2}} \exp\left(-\frac{u^2 + v^2}{\varepsilon_p^2}\right) \exp\left(2\pi i \rho_p \left(u \cos \alpha_q + v \sin \alpha_q\right)\right) \tag{2}$$

with $\varepsilon_p = 1.9863 \times \sqrt{2^p}$, $\rho_p = 0.2592/\sqrt{2^p}$ and $\alpha_q = q\pi/4$ being the scale, frequency and orientation of the Gabor filter, respectively. With these huge variations in the Gabor parameters, the filter bank becomes able to reveal the prominent texture of the periocular image, which can further be apprehended through mean and standard deviation of each sub-region of the image. The filtered images of one sample periocular image are shown in Fig. 2.

Thereafter, the absolute difference of standard deviations pertaining to a first-level sub-block and the second-level sub-blocks underneath it, form the feature vector. Subsequently, the hereof formed feature vector for each individual filter from the filter bank is concatenated side by side to obtain the overall feature template pertaining to one image. Finally, the feature templates of query and gallery images are matched through city-block distance metric, to yield matching scores. The detailed information about the matching scores are presented in Section 4.

3.2 Deep features

Unlike the hand-crafted features, deep features (DF) are obtained through convolutional neural networks (CNNs), which demand high computational power for training and subsequent feature extraction. However, the CNNs are capable of obtaining more distinctive features of the images, attributing to their in-depth layered structures and convolutions occurring at intermittent layers. Besides, the increased computational powers of personal computers (PCs) nowadays has provided a boost to implementation of CNNs with reasonable training times.

Hence, this paper considers a popular CNN, called residual neural network (ResNet) [11], for extracting deep features of the periocular images. This model is selected for its outstanding performance in complex problems like image classification, object detection and face recognition. The core idea of ResNet was to introduce skipping identity connections for one or more layers, thus making the training of deeper networks free from accuracy degradation problem. Moreover, ResNet has fast convergence and astonishing classification accuracy, which motivated us to adopt this model for the problem of periocular recognition.
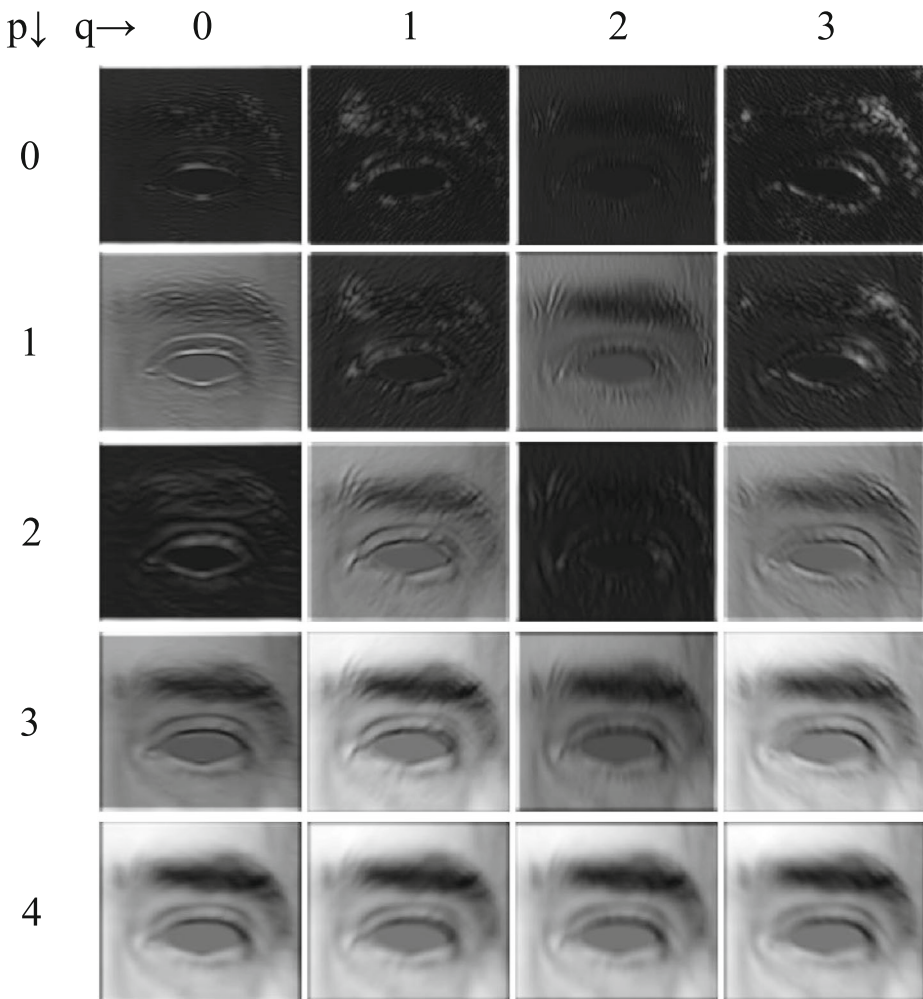
Fig. 2 Filter responses for the 2D Gabor filter bank

In the current work, all the experiments for deep features are conducted through ResNet-101 model. This model is chosen as a trade-off between depth and training time of the model. The ResNet-101 model, pretrained on millions of images from the ImageNet database [9], is utilized for the current task. The pretrained model is modified to have the classification layer as per the requirement of number of classes. Thereafter, weights of the initial layers of the network are frozen by setting their learning rates to zero. This is usually done to avoid overfitting in case of retraining of the CNN models on small datasets. The model is then retrained with 60% of the periocular samples to achieve distinctive representation of their features. Notably, the deep features of the periocular samples are obtained from the "global average pooling" layer of ResNet-101 model. The matching between query and gallery feature vectors is performed through city-block distance metric, in order to have uniform score ranges and distributions. Other details about the training options and used hardware resources are provided in Section 4.

3.3 Matching scores

The features obtained from hand-crafted and deep feature extractors are compared through the city-block or Manhattan distance [7]. This distance function is also known as 1-norm distance between two feature vectors. It yields smaller values for similar vectors and larger values for dissimilar vectors. The mathematical definition of the city-block distance for N-dimensional query (FV_Q) and gallery (FV_G) feature vectors is given in (3).

$$D_{Q,G} = \sum_N \left| FV_Q - FV_G \right| \tag{3}$$

3.4 Score-level fusion

This is a fusion scheme where scores from different classifiers are combined to furnish the final decision [16, 26]. The score-level fusion is viable to achieve without having knowledge of underlying feature extraction procedures. This is the reason score-level fusion is generally preferred over other types of fusion, like feature-level or decision-level fusions. The score-level fusion can take place in two different ways. In the first way, the scores from different approaches are combined through the popular fusion schemes like maximum, minimum, average, and geometric mean. Whereas in the second way, a separate second-level classifier is trained on the concatenations of individual scores, leading to a more effective set of scores. Among both the ways, the former is said to be more versed as it does not need the additional training which is otherwise needed in the latter way. Hence, this paper utilizes the combination approach of score-level fusion, where "average" rule of fusion is adopted for all the experimentation. This essentially means that a new set of genuine and imposter scores are obtained by taking the average (or mean) of the individual genuine and imposter scores from the corresponding frameworks of hand-crafted and deep features. Since, scores in both the employed features are calculated through city-block distance, they have similar ranges and distributions. This excludes the need of executing any score normalization procedures.

4 Experimental results

In order to evaluate the performances of HCF and DF, a capacious set of experiments were conducted on the benchmark CrossEyed Periocular database [29, 30], which constitutes of registered periocular images acquired from the near infrared (NIR) and visible wavelength (VW) illuminations, respectively. This database is highly suitable for periocular experiments as the images within this database have the iris regions masked, so that any evaluation would solely reflect the potential of periocular features. Regarding the size of the database, this database consists of periocular images from 120 subjects at the sample rate of 8 samples per subject. In addition to that, the database provides samples from both the left and right periocular regions for all 120 subjects. Hence, there are 960 images in each of the four acquisition frameworks: right periocular in NIR, left periocular in NIR, right periocular in VW and left periocular in VW, leading to a total of 3840 samples from 480 different classes (as left and right periocular of the same subject form different classes).

For the training task of ResNet-101, the "MiniBatchSize" and "InitialLearnRate" are set to 10 and 0.0003, respectively. Further, training is accomplished by "sgdm" optimizer. Since the training data is small in size, several data augmentation strategies (like random translations and scaling in both horizontal and vertical directions, and random rotations) are

employed. The training is completed for 10 epochs with 20% periocular samples to be used for validation. All the training tasks are completed in MATLAB R2019b on a Windows system with Intel Core i5-10300H CPU @ 2.50 GHz, 8 GB RAM and NVIDIA GeForce GTX 1650 (4 GB) GPU. Besides, the performance evaluation is analysed in terms of popular metrics like equal error rate (EER), decidability index (DI), genuine acceptance rate (GAR), receiver operator characteristics (ROC) curves and area under the ROC curves (AUC). Following subsections discuss about the experimental results for both HCF and DF and their combination.

4.1 HCF

The performance of proposed hand-crafted feature descriptor is presented in the ROC curves illustrated in Fig. 3, where EER, DI and GAR (@ false acceptance rate (FAR) of 0.1% and 1%) are specified for both the left and right periocular regions. For the sake of completeness, results of the proposed approach for both NIR and VW based images are mentioned in the curves. As observed from Fig. 3, EER and GAR of right periocular images (12.28% and 70.59%, respectively) are better than that of other matching frameworks. However, the EERs and AUC values for all four frameworks fall in the interval of 12-14.5% and 0.9235-0.9361 respectively, which do not seem reasonable for practical biometric applications hence limiting the use of standalone HCF. Nevertheless, the proposed HCF possess significant interpretability, as this descriptor turns the highlighted textural information into corresponding numerical features.

4.2 DF

As opposed to HCF, DF have high capacity to achieve improved EER and GAR values, reason being the employed end-to-end CNN model and extraction of meaningful features from the deepest layer. The performance of deep features in all four matching frameworks is depicted in the ROC curves of Fig. 4. It is apparent from Fig. 4 that performance of deep features is certainly at par as compared against that of HCF. The best EER (1.44%) achieved is for right periocular images from VW illumination. Concurrently, both the GAR (at 0.1%
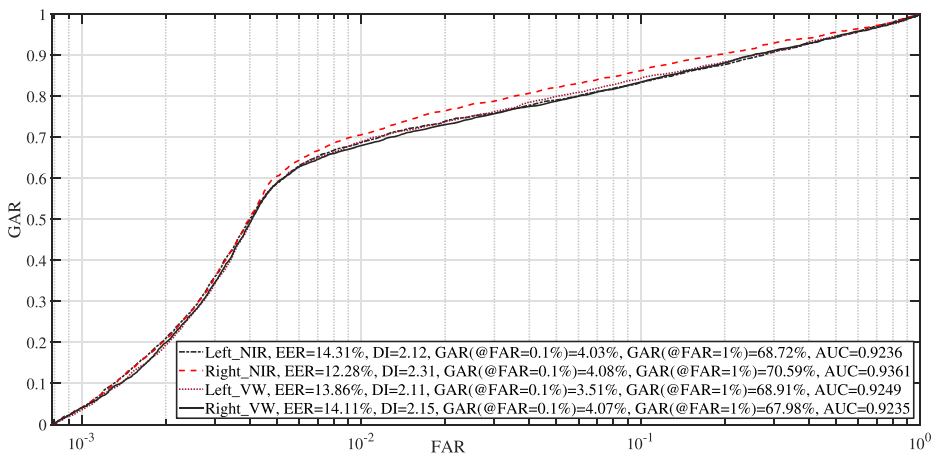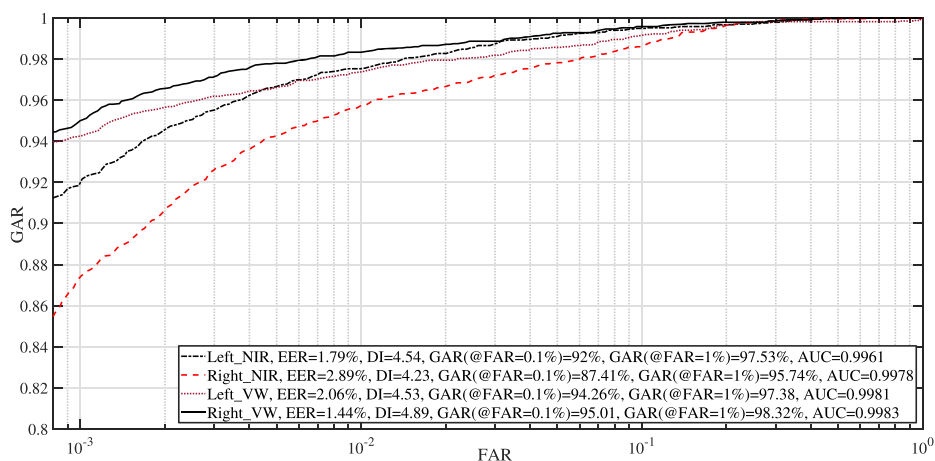

Fig. 3 ROC curves for HCF

Fig. 4 ROC curves for DF

and 1% FAR) and AUC values are extremely high, when compared against proposed non-deep features. This certainly facilitates the application of deep features with highly accurate system and minimum possible false acceptance rate. In addition to the performance metrics, Fig. 5 shows the gradient-weighted class activation mappings (Grad-CAM) [28], which provides a coarse localization map to highlight the important regions in the image, which are more responsible for the predictions made by the network. These maps could be useful in identifying the discerning regions of the periocular images. Figure 5 demonstrates the explanations generated for the sample periocular images from all four frameworks, where it is revealed that the regions just above the eye, especially comprising the eyelids, eyelashes, eyebrows and surrounding skin, emerge as the discerning regions.
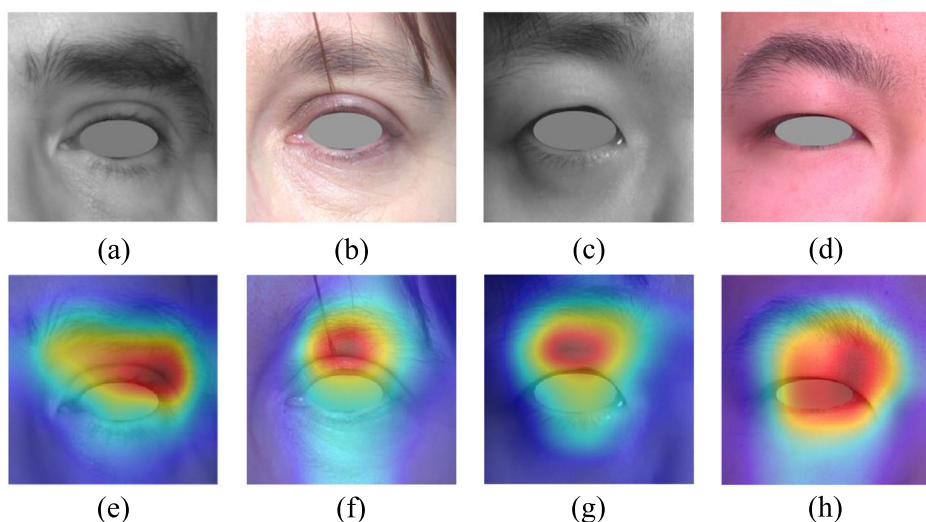

Fig. 5 Grad-CAM [28] explanations for the deep features; (a-d) Sample periocular images, (e-f) Grad-CAM explanations corresponding to samples in (a-d)
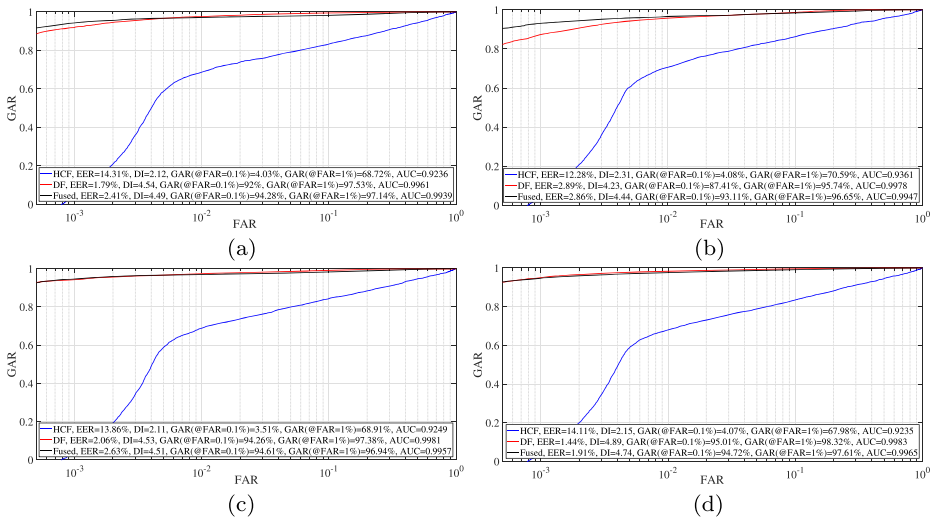
Fig. 6 Roc curves depicting effects of fusion (a) Left_NIR, (b) Right_NIR, (c) Left_VW, (d) Right_VW

4.3 Fusion

The next part of experimentation deals with the score-level fusion of HCF and DF individual performances. The results of score-level fusion are presented via ROC curves for each of the four matching frameworks. As a result of fusion, the performance metrics (like EER/GAR/DI) either improve or remain comparable to the individual performance of deep features. For instance, the EER becomes 2.41% after fusion as compared against 1.79% of the individual deep features, in case of left periocular regions captured in NIR (please see Fig. 6a). However, the GAR (@FAR=0.1%) value for the same scenario improves as a result of fusion. Explicitly, it improves from 92% to 94.28% which counts to 2.48% improvement. The corresponding values of AUC does not differ significantly, proving the fact that the proposed fusion yields comparable AUC (if not improved) with respect to that of DF, simultaneously exhibiting significant improvement with respect to HCF. Hence, it can be inferred that fusion of HCF and DF facilitates the use of the system at lower FAR values like 0.1%, which is a significant gain. Similar pattern of improvement could be observed from Fig. 6b, for the case of right periocular images from NIR illumination. In this case, the improvement in GAR (@FAR=0.1%) is even greater, counting to be 6.52% improved value (i.e. 93.11% as compared against 87.41%). In addition to that, EER and GAR(@FAR=1%) are also improved. On the contrary, fusion of HCF and DF does not always lead to an improvement for the images captured in VW illumination. As can be observed from Fig. 5c, and d that the fusion results in enhancement of performance metrics for left periocular images, but it does not improve the parameters for right periocular images. This could be attributed to the several challenges posed by the visible wavelength, like specular reflections and non-uniform illumination. Therefore, in view of the significant improvements with NIR images, the overall idea of collaborative representation, to improve the recognition performance, proves to be effective. Additionally, constant high values of AUC for all four matching frameworks prove the effectiveness of the proposed approach.

Table 1 Performance metrics for cross-spectral matching

Features	EER (%)	DI	GAR (%) (@FAR=0.1%)	GAR (%) (@FAR=1%)	AUC
Left_HCF	27.55	1.14	9.10%	23.55%	0.7955
Right_HCF	25.20	1.35	8.49%	27.37%	0.8284
Left_DF	19.86	1.68	14.35%	36.48%	0.8845
Right_DF	20.17	1.65	15.18%	35.18%	0.8776
Left_Fused	**16.83**	**1.94**	**26.09%**	**47.88%**	**0.9128**
Right_fused	**15.73**	**2.05**	**27.31%**	**51.55%**	**0.9212**

4.4 Cross-spectral matching

Another important investigation conducted in this paper is about the performance of both the employed descriptors in the context of cross-spectral matching. This is essentially an important evaluation framework, where features of images from one wavelength (say VW) are matched against those of another (say NIR). This sort of matching is challenging in terms of uncorrelated VW and NIR features, therefore leading to reasonable drop in the performance metrics. However, the current investigation targets on evaluating the fusion of hand-crafted and deep features in this challenging matching scenario. Table 1 reports the values of various important metrics for cross-spectral matching with individual and combined features, respectively.

It is evident from Table 1 that the hand-crafted features perform poorly when used in cross-spectral matching scenario, with EER as low as 27.55% and 25.20% for left and right periocular regions, respectively. Whereas, the deep features show noteworthy improvement over hand-crafted feature, with EERs close to 20%. Besides, last two rows of the table report the metrics for evaluation conducted with combined HCF and DF (the combination is simply performed through concatenation). The combined features show significant results (as highlighted by bold entries in the table), with huge improvements in the performance metrics. For instance, the relative improvements in EER with respect to the standalone HCF and DF performances are (38.91%, 37.58%) and (15.26%, 22.01%), respectively for left
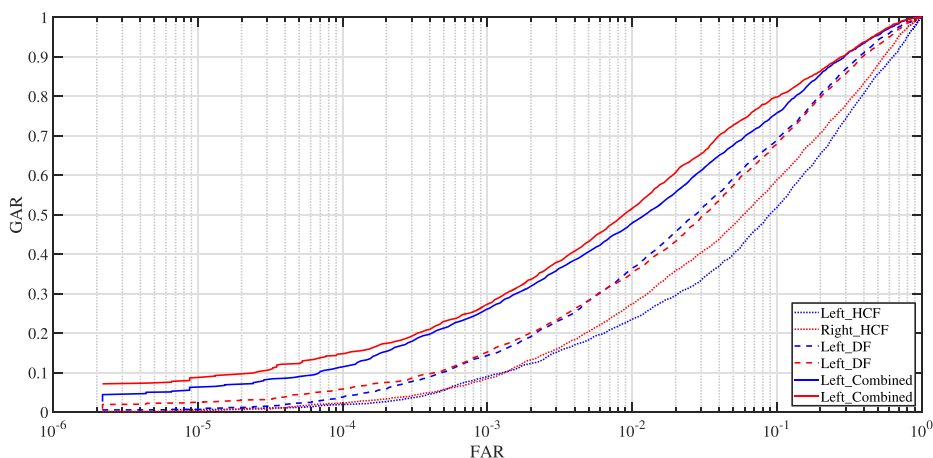

Fig. 7 ROC curves for cross-spectral matching scenario

Table 2 Comparative analysis

Periocular region	Illum.	Descriptor	EER (%)	DI	GAR (%) (@FAR=0.1%)	GAR (%) (@FAR=1%)	AUC
Left	NIR	LBP	4.07	2.31	4.19	92.78	0.9883
		LPQ	3.13	1.92	3.95	94.65	0.9878
		Proposed	**2.41**	**4.49**	**94.28**	**97.14**	**0.9939**
	VW	LBP	4.30	1.36	3.56	88.33	0.9849
		LPQ	4.15	1.46	4.19	92.77	0.9871
		Proposed	**2.63**	**4.51**	**94.61**	**96.94**	**0.9957**
Right	NIR	LBP	3.47	2.41	4.23	92.22	0.9894
		LPQ	2.87	2.06	3.85	95.49	0.9902
		Proposed	**2.86**	**4.44**	**93.11**	**96.65**	**0.9947**
	VW	LBP	3.98	1.47	4.13	90.24	0.9888
		LPQ	4.53	1.54	3.97	91.08	0.9880
		Proposed	**1.91**	**4.74**	**94.72**	**97.61**	**0.9965**

and right periocular images. Whereas the relative improvements in AUC values with respect to their individual counterparts are (14.75%, 11.20%) and (3.20%, 4.97%), respectively for the left and right periocular images. The ROC curves depicted in Fig. 7 also follow the similar patterns of improvement. These improvements clearly validate the effectiveness of the collaborative representations proposed in the current work.

4.5 Comparison with state-of-the-art

In order to furnish qualitative comparison of the proposed approach with state-of-the-art periocular recognition approaches, this paper provides comparative results from two benchmark descriptors, namely LBP and LPQ. The comparative performance metrics are presented in Table 2, with bold entries corresponding to the best results produced by the proposed approach. It is not difficult to observe from Table 2 that the proposed collaborative representation of periocular images outperforms the LBP and LPQ descriptors significantly. The major point of difference between the proposed scheme and state-of-the-art is the high values of GAR at lower FAR of 0.1%. It is worth mentioning that the specified GAR position is improved considerably by the fusion of HCF and DF, as detailed in the Section 4.3. In addition to this, an overall improvement of the proposed approach can also be observed from the increased AUC values for each of the matching frameworks. This is the reason proposed scheme of periocular recognition is at par with the highly effective benchmark schemes for all four acquisition frameworks (which are left/right in NIR/VW), as highlighted by the bold-faced values in Table 2.

5 Conclusion

This paper deals with collaborative representation of near infra-red periocular images through traditional hand-crafted and end-to-end deep features. On one hand, hand-crafted feature descriptor is free from any sort of learning and/or hyperparameter tuning. While on the other hand, deep features are advantageous in terms of superior performance and facility of transfer learning. Hence, the current work focuses on extraction of both hand-crafted and deep features to study the potential of periocular recognition. The hand-crafted features are extracted through multiresolution and multi-scale analysis of the periocular image through a 2D Gabor filter bank, followed by calculation of local statistical measures from the image partitions. Whereas, deep features are extracted through fine-tuning of a popular CNN, namely ResNet-101, and extracting features from its deepest pooling layer. It has been observed from the experimental results that combined knowledge of hand-crafted and deep features can certainly lead to improvement in performance metrics. Notably, the system becomes applicable to highly demanding security venues, where lower FARs remain preferable. This statement is specially supported by the rise in GAR values at lower FARs of 0.1%, which happens as a result of the proposed collaborative representation. In addition to the above, the proposed approach exhibits promising results for challenging matching scenarios like cross-spectral matching. The comparison with state-of-the-art further proves the utility of the proposed approach.

Declarations

References

1. Aginako N, Castrillón-Santana M, Lorenzo-Navarro J, Martínez-Otzeta JM, Sierra B (2017) Periocular and iris local descriptors for identity verification in mobile applications. Pattern Recogn Lett 91:52–59
2. Ahmed NU, Cvetkovic S, Siddiqi EH, Nikiforov A, Nikiforov I (2017) Combining iris and periocular biometric for matching visible spectrum eye images. Pattern Recogn Lett 91:11–16
3. Alahmadi A, Hussain M, Aboalsamh H, Azmi A (2020) Convsrc: Smartphone-based periocular recognition using deep convolutional neural network and sparsity augmented collaborative representation. Journal of Intelligent & Fuzzy Systems 38(3):3041–3057
4. Alonso-Fernandez F, Bigun J (2016) Periocular biometrics: databases, algorithms and directions. In: 2016 4Th international conference on biometrics and forensics (IWBF), IEEE
5. Alonso-Fernandez F, Bigun J (2016) A survey on periocular biometrics research. Pattern Recogn Lett 82:92–105
6. Behera SS, Gour M, Kanhangad V, Puhan N (2017) Periocular recognition in cross-spectral scenario. In: 2017 IEEE International joint conference on biometrics (IJCB), IEEE
7. Bekhet S, Ahmed A (2020) Evaluation of similarity measures for video retrieval. Multimedia Tools and Applications 79(9):6265–6278
8. Bharadwaj S, Bhatt HS, Vatsa M, Singh R (2010) Periocular biometrics: When iris recognition fails. In: 2010 Fourth IEEE international conference on biometrics: theory, applications and systems (BTAS), IEEE
9. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) Imagenet: A large-scale hierarchical image database. In: CVPR09
10. Gangwar A, Joshi A (2014) Robust periocular biometrics based on local phase quantisation and gabor transform. In: 2014 7Th international congress on image and signal processing, IEEE
11. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE conference on computer vision and pattern recognition (CVPR), IEEE
12. Hernandez-Diaz K, Alonso-Fernandez F, Bigun J (2018) Periocular recognition using cnn features off-the-shelf. In: 2018 international conference of the biometrics special interest group (BIOSIG), IEEE, pp 1–5
13. Hollingsworth K, Bowyer KW, Flynn PJ (2010) Identifying useful features for recognition in near-infrared periocular images. In: 2010 Fourth IEEE international conference on biometrics: theory, applications and systems (BTAS), IEEE
14. Hollingsworth KP, Darnell SS, Miller PE, Woodard DL, Bowyer KW, Flynn PJ (2012) Human and machine performance on periocular biometrics under near-infrared light and visible light. IEEE Transactions on Information Forensics and Security 7(2):588–601
15. Kumar G, Bakshi S, Sa PK, Majhi B (2020) Non-overlapped blockwise interpolated local binary pattern as periocular feature. Multimedia Tools and Applications
16. Lumini A, Nanni L (2017) Overview of the combination of biometric matchers. Inform Fusion 33:71–85
17. Luz E, Moreira G, Junior LAZ, Menotti D (2018) Deep periocular representation aiming video surveillance. Pattern Recogn Lett 114:2–12
18. Mahalingam G, Ricanek K (2013) LBP-Based periocular recognition on challenging face datasets EURASIP. Journal on Image and Video Processing 2013(1)
19. Park U, Jillela RR, Ross A, Jain AK (2011) Periocular biometrics in the visible spectrum. IEEE Transactions on Information Forensics and Security 6(1):96–106
20. Park U, Ross A, Jain AK (2009) Periocular biometrics in the visible spectrum: a feasibility study. In: 2009 IEEE 3Rd international conference on biometrics: theory, applications, and systems, IEEE
21. Proença H, Briceño JC (2014) Periocular biometrics: constraining the elastic graph matching algorithm to biologically plausible distortions. IET Biometrics 3(4):167–175
22. Raja K, Ramachandra R, Busch C (2020) Collaborative representation of blur invariant deep sparse features for periocular recognition from smartphones. Image Vis Comput 101:103979
23. 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, IEEE
24. Raja KB, Raghavendra R, Busch C (2016) Collaborative representation of deep sparse filtered features for robust verification of smartphone periocular images. In: 2016 IEEE international conference on image processing (ICIP), IEEE
25. Raja KB, Raghavendra R, Stokkenes M, Busch C (2015) Multi-modal authentication system for smartphones using face, iris and periocular. In: 2015 international conference on biometrics (ICB), IEEE
26. Ross A, Jain A (2003) Information fusion in biometrics. Pattern Recognition Letters 24(13):2115–2125

27. Santos G, Grancho E, Bernardo MV, Fiadeiro PT (2015) Fusing iris and periocular information for cross-sensor recognition. Pattern Recogn Lett 57:52–59
28. Selvaraju RR, Cogswell M, Das A, Vedantam R, Parikh D, Batra D (2017) Grad-CAM: Visual explanations from deep networks via gradient-based localization. In: 2017 IEEE international conference on computer vision (ICCV), IEEE
29. Sequeira A, Chen L, Wild P, Ferryman J, Alonso-Fernandez F, Raja KB, Raghavendra R, Busch C, Bigun J (2016) Cross-eyed - cross-spectral iris/periocular recognition database and competition. In: 2016 international conference of the biometrics special interest group (BIOSIG), IEEE
30. Sequeira AF, Chen L, Ferryman J, Wild P, Alonso-Fernandez F, Bigun J, Raja KB, Raghavendra R, Busch C, de Freitas Pereira T, Marcel S, Behera SS, Gour M, Kanhangad V (2017) Cross-eyed 2017: Cross-spectral iris/periocular recognition competition. In: 2017 IEEE international joint conference on biometrics (IJCB), IEEE
31. Sharma A, Verma S, Vatsa M, Singh R (2014) On cross spectral periocular recognition. In: 2014 IEEE international conference on image processing (ICIP), IEEE
32. Smereka JM, Kumar BV (2013) What is a 'good' periocular region for recognition? In: 2013 IEEE conference on computer vision and pattern recognition workshops, IEEE
33. Uzair M, Mahmood A, Mian A, McDonald C (2015) Periocular region-based person identification in the visible, infrared and hyperspectral imagery. Neurocomputing 149:854–867
34. Vyas R, Kanumuri T, Sheoran G (2018) Cross spectral iris recognition for surveillance based applications. Multimedia Tools and Applications 78(5):5681–5699
35. Zhang Q, Li H, Sun Z, Tan T (2018) Deep feature fusion for iris and periocular biometrics on mobile devices. IEEE Transactions on Information Forensics and Security 13(11):2897–2912
36. Zhao Z, Kumar A (2018) Improving periocular recognition by explicit attention to critical regions in deep neural network. IEEE Transactions on Information Forensics and Security 13(12):2937–2952