CNN Based Periocular Recognition Using Multispectral Images

Vineetha Mary Ipe(✉) and Tony Thomas

Indian Institute of Information Technology and Management, Kerala, India
{vineetha.mary,tony.thomas}@iiitmk.ac.in

Abstract. Over the recent years, the periocular region has emerged as a potential unconstrained biometric trait for person authentication. For a biometric identification scenario to operate reliably round the clock, it should be capable of subject recognition in multiple spectra. However, there is limited research associated with the non-ideal multispectral imaging of the periocular trait. This is critical for real life applications such as surveillance and watch list identification. The existing techniques for multispectral periocular recognition rely on fusion at the feature level. However, these handcrafted features are not primarily data driven and there even exists possibilities for more novel features that could better describe the same. One possible solution to address such issues is to resort to the data driven deep learning strategies. Accordingly, we propose to apply the attributes extracted from pretrained CNN for subject authentication. To the best of our knowledge, this is the first study of multispectral periocular recognition employing deep learning. For our work, the IIITD Multispectral Periocular (IMP) database is used. The best classification accuracy reported for this dataset is 91.8%. This value is not precise enough for biometric identification tasks. The off-the-shelf CNN features employed in our work gives an improved accuracy of 97.14% for the multispectral periocular images.

Keywords: Multispectral periocular recognition · Deep learning · Biometrics · Convolutional neural network (CNN)

1 Introduction

Periocular recognition refers to the automated process of recognizing individuals using periocular images. Periocular is the facial region around the eye including eyelids, eyelashes, and eyebrows [2]. The typical elements of the periocular region are as shown in Fig. 1. Periocular (periphery of ocular) region is unique for each individual because of the distinctive markings of the iris, sclerotic blood vessels and skin texture [3].

The significance of periocular biometrics lies in the fact that it represents a trade-off between using face and iris [13] which can be elucidated as follows:

S. M. Thampi et al. (Eds.): SIRS 2019, CCIS 1209, pp. 94–105, 2020.
https://doi.org/10.1007/978-981-15-4828-4_9

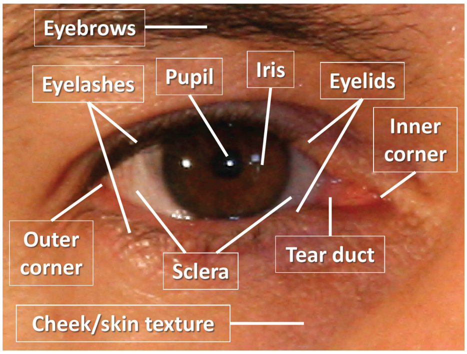

Fig. 1. Elements of the periocular region [2]

- While trying to capture the entire face from a distance, iris information will be of poor quality or low resolution. On the contrary, when the iris is captured from close distance, the whole face may not be visible thereby forcing the recognition system to rely only on the iris. In both these cases, periocular images may be captured with sufficient clarity over a wide range of distances.
- When portions of the face such as the nose and mouth are occluded, the periocular region may be used to determine the identity.
- In cases where the iris cannot be reliably obtained, such as blinking or off-angle poses, the skin in periocular region may be used to recognize the identity.

This implies that the periocular biometric requires less subject cooperation compared to other ocular biometrics which in turn makes the periocular region a potential trait for unconstrained biometrics. However, there is limited research focusing on periocular recognition in multispectral non-ideal imaging scenario. This is critical for real life applications such as tracking, watch list identification, surveillance and security, where it is required to track or authenticate individuals based on the biometric traits in varied illumination conditions. Thus, there is a need for such biometric applications to operate in multiple spectra round the clock. Also, these multispectral images provide rich visual information for authentication and is tougher to be spoofed. This has paved the way for multispectral periocular recognition.

Existing implementations for multispectral periocular recognition rely on fusion of periocular features in different spectra (feature-level fusion) [1]. But, these handcrafted features are not primarily data driven and are based on human assumptions to be the best descriptors for the multispectral periocular trait. There even exists possibilities for more novel or unrevealed features that could better describe the same. Moreover, different handcrafted features exhibit differently with respect to different spectra and this can give rise to compatibility issues while fusing such features for different spectrums. To address such issues, the data driven deep learning strategies can be used.

For our work, we empoly the multispectral images from the IIITD Multispectral Periocular (IMP) database [17], which provides the relevant multispectral ocular images belonging to three spectrums namely near infrared, visible and night vision. The best classification accuracy reported with this database in literature is 91.8% and such works rely on a number of feature descriptors. However, this value is not precise enough to settle with the claim that the features or methodology involved in such tasks are the best approaches for biometric recognition using multispectral periocular images. One possible technique to improve the recognition performance is to resort to the recent trends in deep learning.

The advantage of deep learning strategies lies in the fact that they are primarily data driven i.e. they are able to learn features automatically based on the input data rather than by human hypotheses as in the case of conventional features. Even though deep learning techniques call for huge amount of training data and time for designing and training new CNN architectures for specific applications, deploying off-the-shelf CNN features from pretrained existing architectures help resolve these demands. Accordingly, we propose to apply the extracted attributes from pretrained convolutional neural network (CNN), Alexnet for subject authentication.

Alexnet is a CNN that is trained on 1.2 million high resolution images from the ImageNet database [6]. Hence, the network has learned rich feature depictions for a wide variety of images. This representational power of pretrained deep networks can be effectively utilized by means of feature extraction. In our work, we extract the off-the-shelf CNN features from Alexnet so as to train an SVM based image classifier. We present the efficiency of our method by means of classification accuracy, which is the ratio of the number of exact classifications to the total number of prediction outputs. The off-the-shelf CNN features used in our work provide an improved 97.14% accuracy. Our research work is novel in the prospect that this is the first study for multispectral periocular recognition employing deep learning.

2 Related Works

The pioneering work in the field of periocular biometrics was reported in the literature by Park et al. in 2009 [14]. They made a feasibility study of whether periocular can be used as a biometric trait for recognition and confirmed the same [13]. Bharadwaj et al. also supported the concept of using periocular biometrics when iris recognition fails [4]. Park et al. used conventional methods for attribute extraction, such as scale invariant feature transform (SIFT) [7], local binary patterns (LBP) [12] and histograms of oriented gradients (HOG) [21] in order to assess performance in the presence of various performance degradation factors such as occlusions, scale, pose, and gaze.

Deep learning approaches especially CNNs have gained boundless popularity for computer pattern analysis and computer vision tasks. However, surveys on periocular biometrics [2] suggest that few studies have taken advantage of deep learning methods for improving periocular recognition rates. Zhao et al. [22] presented a semantics assisted convolutional neural network (SCNN) method which uses semantic information apart from learning identities. This technique involves learning from scratch which is computationally expensive and does not have a significant boost in the recognition performance. Proenca et al. [15] proposed a deep learning framework for periocular recognition without iris and sclera (Deep-PRWIS). Though this is an attempt to highlight the importance of periocular region alone, research suggests that use of entire periocular region without masking provides better recognition performance [13], which is the crux of the biometrics. Luz et al. [8] proposed a periocular region recognition (PRR) using VGG (visual geometry group) transfer learning. Kevin et al. [5] performed periocular recognition using off-the-shelf CNN features. Compared against the conventional periocular features such as LBP, HOG, and SIFT, they show an EER (equal error rate) reduction up to 0.4. Considering the pace at which research in deep learning is progressing,we can anticipate much more precise and fast algorithms.

Research in periocular biometrics has paved its way to emerge as a potential trait for unconstrained biometrics. For such a biometric system to operate 24×7, it should be capable of person identification/authentication in multiple spectra. This has led to the development of multispectral biometrics. The multispectral recognition has been performed on the biometric traits such as face, fingerprint, iris, sclera and palmprint [20] and these images have shown to provide integral information owing to the fact that these traits exhibit different physical characteristics in different spectra. Combining the images belonging to different spectrums improves the strength of the biometric systems against spoofing or counterfeiting thus increasing the robustness of the system.

Very few works are reported in literature for multispectral periocular recognition. Tapia et al. [18] presented a gender classification method from multispectral periocular images using features extracted from shape, intensity and texture of periocular images. Algashaam et al. [1] presented a multispectral periocular classification with multimodal compact multi-linear pooling in which complementary data from multispecral images are fused. This work provides the best classification accuracy with the IIITD Multispectral Periocular (IMP) database, the only publically available database for multispectral periocular images. The highest accuracy reported is 91.8% which is not lucrative enough to accept the above approach as the perfect methodology for multispectral periocular recognition. One possible remedy for performance improvement is to incorporate deep learning techniques. Based on these facts, we extract the off-the-shelf CNN features from Alexnet so as to train an SVM based image classifier.

3 Proposed SR Based Periocular Recognition

The framework for periocular recognition is as shown in Fig. 2. This recognition framework is one of its kind whereby off-the-shelf CNN features are extracted from multispectral periocular images of the IMP database.

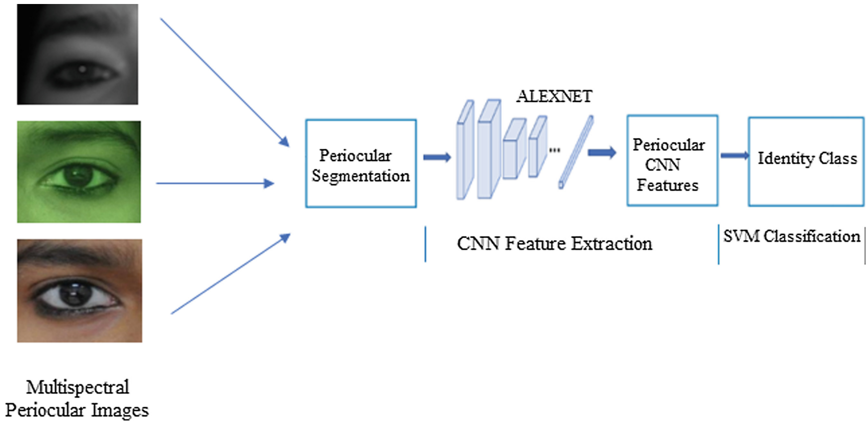

Fig. 2. Experimental framework

3.1 Periocular Segmentation

The periocular region is segmented from these images by modifying the Viola-Jones algorithm [19]. Such a segmentation routine is precise ensuring the incorporation of vital periocular information such as eye, eyebrow, and surrounding skin texture. After segmentation, the periocular images belonging to distinct individuals are grouped together.

3.2 CNN Feature Extraction

The segmented periocular images are fed into the CNN feature extraction module. The CNN used for feature extraction in our work is Alexnet. It consists of 5 convolutional (conv.) layers and 3 fully connected (f_c) layers. In addition, there are ReLU and max-pooling layers. Multiple convolutional kernels (filters) extract features in an image. The first convolutional layer of Alexnet contains 96 kernels of size $11 \times 11 \times 3$. The architecture of Alexnet is shown in Fig. 3.

Convolutional layer is the integral element of CNN. The convolutional layers comprise a series of kernels or learnable filters which derives the local attributes from the periocular images [9]. Each kernel generates a kernel map or feature map which is produced by convolution operation ($*$). The first convolutional layer extracts low-level features and the succeding convolutional layer extracts higher-level features.

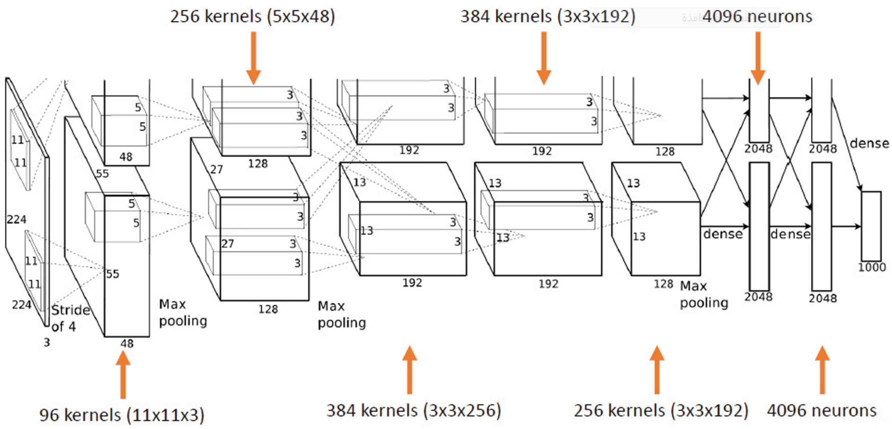

Fig. 3. Alexnet architecture [6]

The non-linear layers introduce non-linearity into the CNN thus enabling the learning of complex models. In Alexnet the non-linear activation function used is ReLU. It applies the following function

$$f(x) = max(0, x) \tag{1}$$

The pooling or subsampling layer reduces the dimensionality of the feature maps. In Alexnet, max pooling is used. The max pooling returns the maximum output inside a rectangular neighbourhood thus making the features robust to minute variations for formerly learned features.

The feature maps of the preceeding layer are convolved with learnable kernels at a convolution layer and passed to the activation function to form the output feature map. Each output combines convolutions with many input maps. This can be represented by Eq. (2).

$${x_j}^L = f\left(\sum_{i \epsilon M_j} {x_i}^{L-1} * k_{ij}^L + b_j^L\right) \tag{2}$$

where

$$\begin{aligned}
{x_j}^L &- j^{th}\ output\ of\ L^{th}\ layer \\
k_{ij}^L &- kernel\ for\ the\ L^{th}\ layer \\
b_j^L &- additive\ biases\ for\ the\ L^{th}\ layer \\
M_j &- set\ of\ input\ maps
\end{aligned}$$

The CNN feature extraction of periocular images consists of multiple similar steps and each step is made of cascading three aforementioned layers.

3.3 SVM Classification

The features extracted from Alexnet are then fed into the classification module. The periocular recognition task being a multiclass identification problem, we choose to perform classification using simple multi-class SVM [16], which is quite popular in image classification tasks because of its robustness, accuracy, and effectiveness while using small training set. Since SVM is mostly used for the binary classification problem, ECOC (error-correcting output codes) based SVM is used for implementing the multiclass problem, whereby the direct multiclass task is split into several binary classification tasks and the results from these indirect classifiers are then combined to give correct predictions of identity class.

4 Experimental Results

4.1 Database

For our experiments IIITD multispectral periocular (IMP) database [17] is used. It is the only publically available database for multispectral periocular images. It contains 1240 multispectral images of 62 persons in three spectrums namely visible, night vision and near infrared. The visible wavelength images are captured with a digital camera, the near infrared dataset with a iris scanner and the night vision dataset with a Sony camcorder in night vision mode.

4.2 Performance Metric and Baseline Method

To report the performance of our work, we rely on classification accuracy. It is a popular metric for assessing the classification models. It is defined as the ratio of the number of correct predictions to the total number of predictions.

The baseline methodology we used for comparison is the multimodal compact multi-linear pooling by Algashaam et al. [1]. This baseline work achieved periocular recognition accuracy of 91.8%, the best reported in the literature with IMP dataset.

4.3 Experimental Setup

The periocular images obtained from the IMP database are grouped into three categories corresponding to the visible, night vision and near infrared spectra. Each category comprises 62 distinct classes with each class containing the periocular region of a particular individual. Thus these 62 classes correspond to 62 distinct individuals. Additionally, we also group the multispectral images into another 62 classes such that each class contains the images of all the three spectra of an individual. For each of these four categories, we choose 70% of the data corresponding to each class on a random basis for training and the remaining 30% for testing.

In this work, we use Alexnet to extract features from the periocular images. This deep CNN is the ILSVRC 2012 winner with a top-5 error rate of 15.3%, compared to 26.2% achieved by the runner up [6]. This network is trained on more than one million images of the ImageNet database and can classify images into 1000 object categories. This deep network is thus capable of capturing and encoding complex attributes of the periocular images.

For feature extraction, we deploy the pretrained Alexnet model in MATLAB R2018b. With the output from each layer used as an attribute vector to characterise the multispectral periocular images, the performance of each layer is assessed using classification accuracy.

The pretrained CNN model Alexnet is not altered at all using the training data. Only features are extracted from CNN. These activations corresponding to the training images, in turn, are used to train the multiclass SVMs. The multiclass SVM used in our work is equivalent to combining multiple binary SVM classifiers and the number of class depends on the count of subjects in the periocular database.

4.4 Performance Analysis

Different levels of visual content are encoded by each layer of CNN. With the output from each layer used as an attribute vector to characterise the multispectral periocular trait, the performance of each layer is assessed using classification accuracy. The accuracies obtained for different layers of the CNN model are shown in Fig. 4.

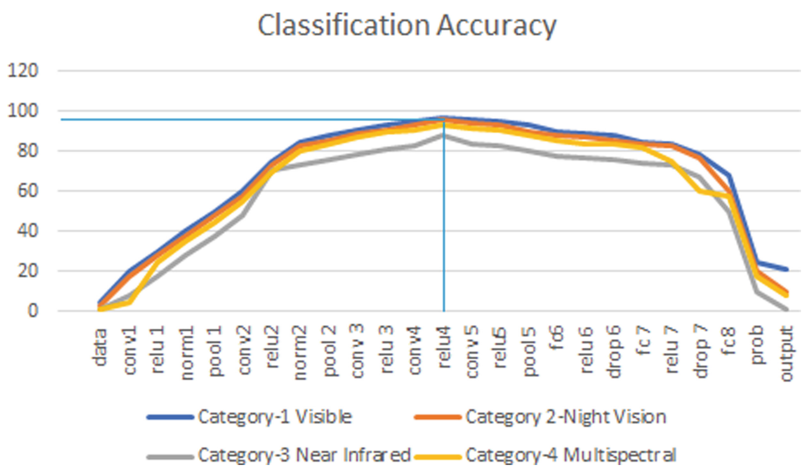

Fig. 4. Performance comparison

For each category, we get a maximum classification accuracy with 'relu4' (fourth rectified linear unit layer) in Alexnet model in Matlab R2018b). The highest classification accuracies obtained are 97.14% (category 1-visible), 95.86% (category 2-night vision), 88% (category 3-near infrared) and 93.55% (category

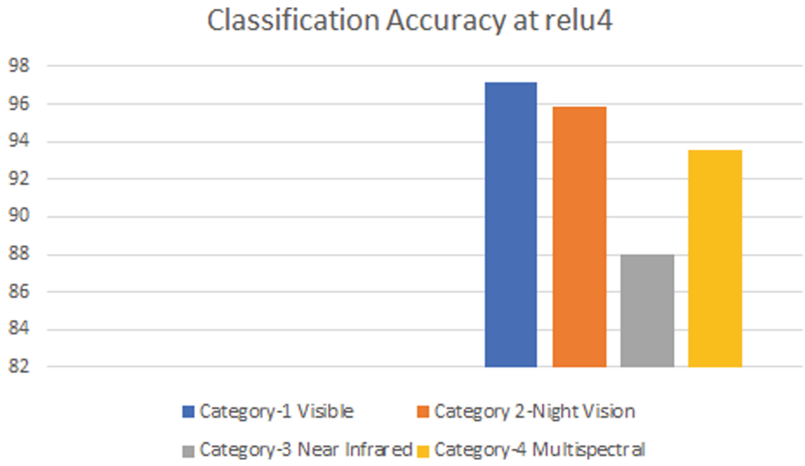

Fig. 5. Performance at relu4

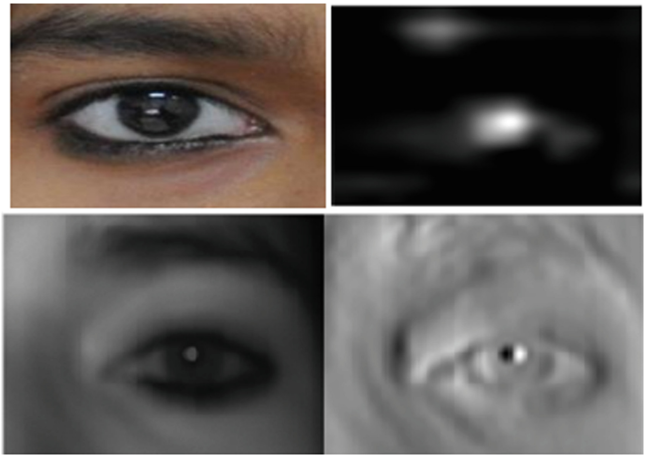

Fig. 6. Activations in Alexnet model

4-multispectral). This is illustrated in Fig. 5. A comparatively low accuracy for near infrared images owes to the fact that these images in the IMP database are blurred and rich visual details are hence not available for recognition.

The difference in accuracy in the distinct layers of Alexnet owes to the distinct levels of features or activations learnt by them. The channels in earlier layers learn simple features like color, edges etc., while those in deeper layers learn much more complex features like that of eyes. The low accuracy in the initial layers is due to the fact that they retain the majority of the input as such or only learn simple features which are not quite efficient for distinguishing individuals. The final layers may not activate at all since there is nothing more to learn beyond a particular layer. Consequently, the recognition accuracy falls towards the last layers.

For visualisation, we normalise the activations to gray-scale such that minimum activation is 0 and the maximum is 1. Such a mapping leads to white pixels representing strong positive activations and gray representing weak activations. Figure 6 shows activations for periocular image in the visible spectrum and in the near infrared spectrum from the layer that gives the maximum accuracy i.e. 'relu4'. From these activations it is evident that the one corresponding to visible spectrum has strong white pixels around the center of the periocular region which is absent in the other case and this strong activation leads to high accuracy.

Comparison of the proposed method with other approaches for multispectral periocular recognition on IMP dataset is shown in Table 1. From this, it is evident that the proposed approach has much better recognition accuracy when compared to the existing methods.

Table 1. Comparison with existing approaches

Methods	Recognition accuracy (%)
Concatenation [20]	86.3
Element-wise multiplication [20]	87.1
Canonical correlation analysis [23]	89.5
Discriminant correlation analysis [24]	90.1
Multimodal compact multi-linear pooling [1]	91.8
Proposed approach	97.14

5 Conclusion

In this work, we have addressed the approach of multispectral periocular recognition from a deep learning perspective. Our experiment shows that there is significant improvement in recognition performance by the application of off-the-shelf CNN features. There are various open questions and challenges regarding the deployment of deep learning to the problem of periocular recognition, even though CNN such as Alexnet is effective in encoding discriminative features for periocular recognition. The computational intricacy of CNNs used for recognition task can be addressed by using model reduction techniques such as pruning and compression [5]. The low accuracy for near infrared images can be addressed using super-resolution [11], a technique for improving the resolution of images. Also, this deep learning approach can also be extended to cross-spectral periocular recognition. Recent developments in Deep Reinforcement Learning and Evolution Theory allow the networks to adapt themselves [10] and achieve better results for the task of periocular recognition. There are several other frameworks in the field of deep learning such as Recurrent Neural Network (RNN), Stacked Auto-Encoder (SAE) and so on. To improve the representation capability of periocular region, these architectures can be used independently or combined with conventional CNNs.

Acknowledgement. This work is done as a part of the project CEPIA (Centre of Excellence in Pattern and Image Analysis) 2019–20, which is funded by the Kerala state planning board.

References

1. Algashaam, F.M., Nguyen, K., Alkanhal, M., Chandran, V., Boles, W., Banks, J.: Multispectral periocular classification with multimodal compact multi-linear pooling. IEEE Access **5**, 14572–14578 (2017)
2. Alonso-Fernandez, F., Bigun, J.: A survey on periocular biometrics research. Pattern Recogn. Lett. **82**, 92–105 (2016)
3. Bakshi, S., Sa, P.K., Majhi, B.: A novel phase-intensive local pattern for periocular recognition under visible spectrum. Biocybern. Biomed. Eng. **35**(1), 30–44 (2015)
4. Bharadwaj, S., Bhatt, H.S., Vatsa, M., Singh, R.: Periocular biometrics: when iris recognition fails. In: 2010 Fourth IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS) (2010)
5. Hernandez-Diaz, K., Alonso-Fernandez, F., Bigun, J.: Periocular recognition using CNN features off-the-shelf. In: 2018 International Conference of the Biometrics Special Interest Group (BIOSIG) (2018)
6. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. Commun. ACM **60**(6), 84–90 (2017)
7. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004). https://doi.org/10.1023/B:VISI.0000029664.99615.94
8. Luz, E., Moreira, G., Junior, L.A.Z., Menotti, D.: Deep periocular representation aiming video surveillance. Pattern Recogn. Lett. **114**, 2–12 (2018)
9. Namatēvs, I.: Deep convolutional neural networks: structure, feature extraction and training. Inf. Technol. Manag. Sci. **20**(1), 40–47 (2017)
10. Nguyen, K., Fookes, C., Ross, A., Sridharan, S.: Iris recognition with off-the-shelf CNN features: a deep learning perspective. IEEE Access **6**, 18848–18855 (2018)
11. Nguyen, K., Fookes, C., Sridharan, S., Tistarelli, M., Nixon, M.: Super-resolution for biometrics: a comprehensive survey. Pattern Recognit. **78**, 23–42 (2018)
12. Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Trans. Pattern Anal. Mach. Intell. **24**(7), 971–987 (2002)
13. Park, U., Jillela, R., Ross, A., Jain, A.: Periocular biometrics in the visible spectrum. IEEE Trans. Inf. Forensics Secur. **6**(1), 96–106 (2011)
14. Park, U., Ross, A., Jain, A.K.: Periocular biometrics in the visible spectrum: a feasibility study. In: 2009 IEEE 3rd International Conference on Biometrics: Theory, Applications, and Systems (2009)
15. Proenca, H., Neves, J.C.: Deep-PRWIS: periocular recognition without the iris and sclera using deep learning frameworks. IEEE Trans. Inf. Forensics Secur. **13**(4), 888–896 (2018)

16. Scholkopf, B., Smola, A.J.: Learning with Kernels: Support Vector Machines, Regularization, Optimization, and Beyond. MIT Press, Cambridge (2001)
17. Sharma, A., Verma, S., Vatsa, M., Singh, R.: On cross spectral periocular recognition. In: 2014 IEEE International Conference on Image Processing (ICIP), pp. 5007–5011. IEEE (2014)
18. Tapia, J., Viedma, I.: Gender classification from multispectral periocular images. In: 2017 IEEE International Joint Conference on Biometrics (IJCB) (2017)
19. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR (2001)
20. Zhang, D., Guo, Z., Gong, Y.: Multispectral biometrics systems. In: Zhang, D., Guo, Z., Gong, Y. (eds.) Multispectral Biometrics, pp. 23–35. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-22485-5_2
21. Zhang, S., Wang, X.: Human detection and object tracking based on histograms of oriented gradients. In: 2013 Ninth International Conference on Natural Computation (ICNC) (2013)
22. Zhao, Z., Kumar, A.: Accurate periocular recognition under less constrained environment using semantics-assisted convolutional neural network. IEEE Trans. Inf. Forensics Secur. **12**(5), 1017–1030 (2017)
23. Haghighat, M., Abdel-Mottaleb, M., Alhalabi, W.: Fully automatic face normalization and single sample face recognition in unconstrained environments. Expert Syst. Appl. **47**, 23–34 (2016)
24. Haghighat, M., Abdel-Mottaleb, M., Alhalabi, W.: Discriminant correlation analysis for feature level fusion with application to multimodal biometrics. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) (2016)