

Received June 22, 2017, accepted July 18, 2017, date of publication July 24, 2017, date of current version August 14, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2731118

# Multispectral Periocular Classification With Multimodal Compact Multi-Linear Pooling

**FAISAL M. ALGASHAAM<sup>1</sup>, (Graduate Student Member, IEEE), KIEN NGUYEN<sup>1</sup>, (Member, IEEE), MOHAMED ALKANHAL<sup>2</sup>, VINOD CHANDRAN<sup>1</sup>, (Senior Member, IEEE), WAGEEH BOLES<sup>1</sup>, (Member, IEEE), AND JASMINE BANKS<sup>1</sup>, (Member, IEEE)**

<sup>1</sup>Queensland University of Technology, Brisbane, QLD 4001, Australia

<sup>2</sup>King Abdulaziz City for Science and Technology, Riyadh, 12371, Saudi Arabia

Corresponding author: Faisal M. Algashaam (faisal.algashaam@hdr.qut.edu.au)

The work of F. M. Algashaam was supported by the Scholarship of the Ministry of Education, Saudi Arabian Government 2014–2018.

**ABSTRACT** Feature-level fusion approaches for multispectral biometrics are mainly grouped into two categories: 1) concatenation and 2) elementwise multiplication. While concatenation of feature vectors has benefits in allowing all elements to interact, it is difficult to learn output classification. Differently, elementwise multiplication has the benefits in enabling multiplicative interaction, but it is difficult to learn input embedding. In this paper, we propose a novel approach to combine the benefits of both categories based on a compact representation of two feature vectors' outer product, which is called the multimodal compact multi-linear pooling technique. We first propose to expand the bilinear pooling technique for two inputs to a multi-linear technique to accommodate for multiple inputs (multiple inputs from multiple spectra are frequent in the multispectral biometric context). This fusion approach not only allows all elements to interact and enables multiplicative interaction, but also uses a small number of parameters and low computation complexity. Based on this fusion proposal, we subsequently propose a complete multispectral periocular recognition system. Employing higher order spectra features with an elliptical sampling approach proposed by Algashaam *et al.*, our proposed system achieves the state-of-the-art performance in both our own and the IIT multispectral periocular data sets. The proposed approach can also be extended to other biometric modalities.

**INDEX TERMS** Periocular recognition, multispectral biometrics, feature-level fusion, higher order spectra, bilinear pooling.

## I. INTRODUCTION

Biometrics have been shown to be critical to deal with the increasing incidents of fraud challenges in highly secure identity authentication systems. Unlike traditional token-based (e.g. cards, keys) and knowledge-based (e.g. PINs, passwords) approaches, biometrics cannot be lost, forgotten or shared. A number of human physiological and behavioral characteristics such as face, iris, fingerprint, keystroke, palm vein, retina and ear have been successfully used as biometrics [1]. Among these, the human periocular region, often referred to as skin textures and anatomical features of the face region in the vicinity of the eye, possibly including the eye, eyelids, eyelashes and eyebrows [2], [3], has emerged as a promising biometric trait for human identification, promising an attractive trade-off between the iris alone and the entire face, especially for cases where neither the iris nor a full facial image can be acquired. However, similar to face and iris,

recognizing periorculars 24/7 in both day- and night-times in less constrained conditions is challenging due to the variations highlighted by different spectra.

Multispectral biometric technologies are emerging to address this challenge owing to two major advantages: (1) they offer richer information details for extracting features and (2) they are more robust to spoof attacks since they are more difficult to be duplicated or counterfeited [4]. These approaches capture the same biometric trait using different sensors at different wavelengths of the electromagnetic spectrum, e.g. Visible, Near-Infrared (NIR), Short-wave Infrared (SWIR) and Long-wave Infrared (LWIR) [5]. With their complementary information, fusing images from different spectra has been shown to be very effective in improving both recognition performance and robustness (i.e. versatility, usability and security) of the biometric systems for face [6], iris [7], [8], sclera [9], [10], palmprint [11], and

fingerprint [12] recognition. Despite sharing many similar properties with the face, the irises and the sclera, periocular has seen limited research in the multispectral fusion [10]. In this paper, we aim at effectively fusing complementary information from multiple spectral images to improve the classification accuracy of the periocular biometric system.

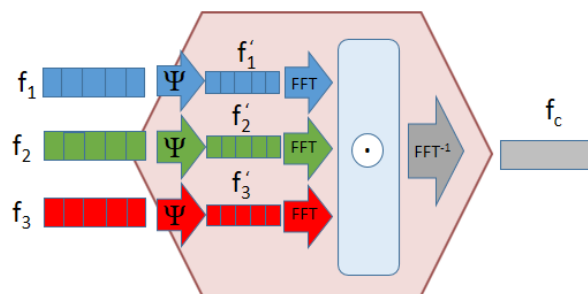
There are four levels of fusion depending on the stage in which the fusion is performed: sensor-level, feature-level, score-level and decision-level [13]. While early fusion retains more information, it is also vulnerable to noise and unnecessary details which are not used for classification. In contrast, late fusion will work with more compact features which are more directly used by the classification task, but lose details due to the feature extraction step. Feature-level fusion provides a balance between the amount of information and the robustness to noise. However, feature-level fusion has seen limited research in the literature since it normally requires access to the feature representation, which may not be available, especially in commercial systems. In addition, the feature representations extracted from different modalities may not be compatible. Hence, even though feature-level fusion is believed to be more effective than score-level and decision-level, it has been employed less frequently. However, in the context of multi-spectral fusion, feature-level fusion is a perfect candidate since the feature representations used in different spectra can be similar and compatible.

There are two typical categories of feature-level fusion including [14]:

- Concatenation: the benefit of this approach is that it allows all elements to interact. However, it is difficult to learn output classification.
- Element-wise multiplication: the benefit of this approach is that it enables multiplicative interaction.

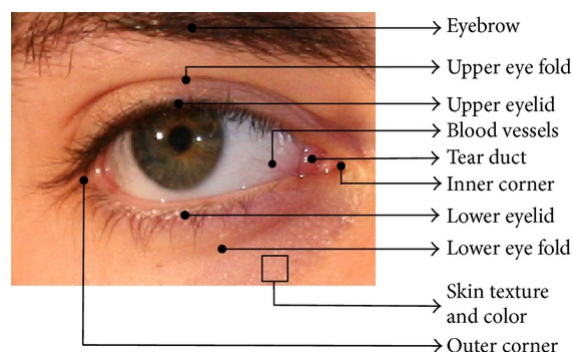
However, it is difficult to learn input embedding.

Bilinear pooling has been shown to perform extremely well for feature-level fusion because it can combine the benefits of both categories [15]. The major challenge for bilinear pooling is that it requires a huge number of parameters and expensive computations. Recently, Gao *et al.* [16] proposed a new approach to deal with these requirements, effectively combining two feature vectors. This has brought in a new approach for multimodal combining, which has been shown to outperform other feature-fusion approaches for combining visual and textual feature vectors in the visual question answering task [17] and neural machine translation [18]. In this paper, we propose a new approach for fusing multispectral periocular images based on compact bilinear pooling. We first extend the bilinear pooling to multi-linear pooling to accommodate for multiple inputs as shown in Figure 1. The proposed extended fusion approach allows all elements in each feature to interact, enabling multiplicative interaction with a small number of parameters and reduced computational cost. Then based on this extension, we propose a complete framework for multispectral periocular system recognition based on the feature-level fusion.



**FIGURE 1.** Our proposed Multimodal Compact Multi-linear Pooling (MCMP) with three input feature vectors. Three input feature vectors are combined in an efficient way to generate a fused feature vector.

The remainder of this paper is structured as follows: Section II reviews related work in periocular recognition and multispectral fusion; Section III introduces multimodal compact bilinear pooling; Section IV describes the proposed approach; Section V presents our experiments; and Section VI concludes the paper.



**FIGURE 2.** The areas surrounding the eyes that can be used for the periocular recognition task [19], [20].

## II. RELATED WORK

### A. PERIOULAR RECOGNITION

In 2009, Park *et al.* [19] first investigated and proved that the periocular region is sufficiently unique and reliable to serve as a standalone biometric. It can also be used complementary to iris and face biometrics, especially in unconstrained conditions when the full face cannot be obtained due to occlusion and the iris image is of poor quality due to its small size and short imaging distance constraint [20], [21]. Since then, periocular recognition has emerged as one of the most active research areas in biometrics [21]–[27]. Figure 2 depicts one example of a periocular region. A typical periocular recognition approach first captured the periocular of the subject using a visible or near-infrared camera. If the quality is acceptable with high resolution, sufficient illumination and blur free [28], the periocular region is then segmented. The periocular region is normally a rectangular region localized by the eye center or the inner and outer corners of the eye.

Choosing the best periocular region for recognition is still an open question for the research community [29]. The next step is extracting features. Choosing features representing reliable and discriminative properties of the periocular is one of the most critical tasks in a periocular recognition process. There are two feature groups: whole-image features and key-point features. While the whole-image features such as Histogram of Gradient (HOG) and Local Binary Pattern (LBP) are extracted from the whole image or region of interest (ROI), the key-point features are extracted from a set of discrete points using such approaches as Scale Invariant Feature Transform (SIFT) [30] and Speeded Up Robust Features (SURF) [31]. The whole-image features are usually in the form of color, texture of the Region Of Interest (ROI) and shape of eyelids and eyebrows. The final features are subsequently compared with the features stored in the database to find the match. The matching is performed through different classification techniques such as k Nearest Neighbor (k-NN), Support Vector Machines (SVMs) and Gaussian Mixture Models (GMMs). Recently, Algashaam *et al.* have shown the advantages of Higher-Order-Spectra features for the periocular classification task. They proposed to use the advantages of 2D Higher-Order-Spectra to encode the periocular region [32], [33]. The authors show that the proposed Higher-Order-Spectra feature extraction technique coupled with a novel elliptical sampling approach has improved the classification performance remarkably.

### B. MULTISPECTRAL BIOMETRIC FUSION

The physiological characteristics of the same biometric trait can be revealed differently using sensors of different wavelengths of the electromagnetic spectrum. Other types of sensors beyond RGB cameras such as NIR cameras and thermal cameras (long-wavelength infrared) have been experimented with for biometric modalities such as face, iris, sclera, palm-print and fingerprint [4]. These multispectral images have been shown to contain complementary information to each other. Fusing multispectral images is expected to improve both the recognition accuracy and the robustness of the biometric system. Most of the fusion approaches for multispectral in the literature are based on score-level [7], [8], [10] owing to the fact that the scores are readily available at the end of the matching process. Feature-level fusion approaches are believed to be more effective than the score-level counterparts owing to the fact that the feature vectors contain richer details than the score values [34]. Many feature-level fusion approaches have been proposed. Two of the most popular approaches of feature-level fusion are concatenation and element-wise multiplication [4]. There are a number of other approaches such as Wild *et al.* [35] who proposed to select and keep the most common bit values for each location in the iris feature vectors. Although it has not been used for multispectral fusion yet, it is worth noting a body of work in multimodal biometric fusion using Canonical Correlation Analysis (CCA) [36] and its extension, Discriminant Correlation Analysis (DCA) [14], for feature-level fusion.

One interesting approach, which can be considered as feature-level in the multispectral task, is fusing multiple neural networks into one as shown in [37] and [38]. However, using deep neural networks is extremely computationally demanding (e.g. with tens of Graphical Processing Units (GPUs)) and normally requiring a huge amount of training data. Our proposed feature-level fusion approach explicitly models the relationship between multiple input feature vectors with affordable computational complexity.

### III. COMPACT BILINEAR POOLING

Bilinear pooling calculates the outer product of two feature vectors,  $f_1 \in \mathfrak{N}^{n_1}$  and  $f_2 \in \mathfrak{N}^{n_2}$ , and learns a model  $W$  (here linear), i.e.  $z = W[f_1 \otimes f_2]$ , where  $\otimes$  denotes the outer product ( $f_1 f_2^T$ ) and  $[\cdot]$  denotes linearizing the matrix in a vector [39]. Bilinear pooling is attractive for fusion since it allows all elements of both vectors to interact with each other in a multiplicative manner. However, calculating the outer product of high dimensional feature vectors is very computationally demanding. To deal with this computation problem, Gao *et al.* [16] proposed to project the high dimensional feature vectors into a low dimensional feature space based on Count Sketch projection function  $\Psi$ . The benefit of this projection is that the Count Sketch of the outer product is equivalent to convolution of two component Count Sketches as,  $\Psi(f_1 \otimes f_2) = \Psi(f_1) * \Psi(f_2)$  [16]. The convolution theory states that the convolution in the time domain is equivalent to the element-wise product in the frequency domain as,

$$\Psi(f_1) * \Psi(f_2) \approx FFT^{-1}(FFT(\Psi(f_1)) \odot FFT(\Psi(f_2))),$$

where  $FFT$  and  $FFT^{-1}$  denote Fast Fourier Transform and its inverse. Hence the outer product can be efficiently calculated, allowing multiplicative interaction between all elements of component vectors with a small number of parameters, activations and computation [17].

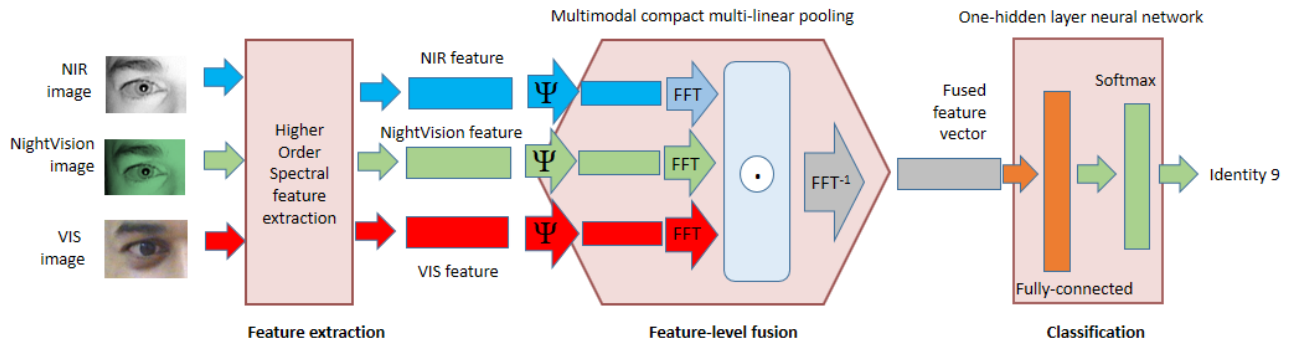
Compact bilinear pooling theory has been effectively applied to combine visual and textual features and is the state-of-the-art in the visual question answering task [17] and neural machine translation task [18]. Considering its advantages when combining the benefits from both concatenation and element-wise multiplication [17], [18], it would be an ideal candidate to learn and model the correlation between multispectral modalities of the same biometric trait in the feature level. To the best of our knowledge, this technique has never been explored for the multispectral fusion task.

### IV. THE PROPOSED APPROACH

In this section, we first discuss the proposal extending bilinear pooling to multi-linear pooling to accommodate multiple inputs. Then the proposed complete framework for fusing and classifying multispectral periocular images is presented.

#### A. MULTI-LINEAR POOLING

Consider the context of fusing multispectral periocular images, there may be more than two spectra such as the

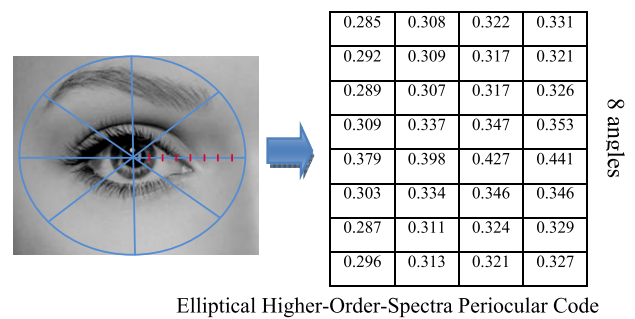


**FIGURE 3.** The proposed fusion approach. First, the features are extracted from the NIR periocular image, NightVision periocular image and the VIS periocular image using Higher-Order-Spectral features. These features are then fused based on the proposed multimodal compact multi-linear pooling. The fused feature vector is optionally passed through a fully connected layer, followed by a softmax layer for the classification task.

Visible, Near-InfraRed, and NightVision spectra. While bilinear pooling only accepts two inputs, we propose to extend it to multi-linear pooling theory to accommodate multiple inputs. The extension is inspired by the element-wise product in the frequency domain, which can be easily extended to multiple inputs. The proposed multimodal compact multi-linear pooling is shown in Figure 1. Three input feature vectors,  $f_1, f_2,$  and  $f_3$  are first projected to a low dimensional space by a Count Sketch projection function  $\Psi$ . Subsequently these three low dimensional representations,  $f'_1, f'_2,$  and  $f'_3$  are transformed into the frequency domain using Fast Fourier Transform. Three corresponding frequency-domain feature vectors are then element-wise multiplied, followed by an inverse Fourier transform to convert the signal back to the time domain. It is worth noting that the number of input feature vectors can be trivially increased to more than 3 using the same process.

**B. PERIOCLAR RECOGNITION**

There are many different approaches for periocular recognition as shown in Section II. The proposed fusion approach works with any features. In this work, we choose to demonstrate the performance of the proposed approach on the Elliptical Higher-Order-Spectra Periocular Code features which were introduced by Algashaam et al. in [33] due to the robustness of the proposed feature extraction technique. The proposed technique combines the invariance properties of 1D Higher Order Spectral features with an elliptical coordinate sampling technique to generate a feature vector called eHPC, which achieves robustness in scale, translation and head rotation. Figure 4 depicts an example of the Elliptical Higher-Order-Spectra Periocular Code coupled with the proposed sampling technique. This approach achieves state-of-the-art recognition accuracy in two periocular datasets, 99.52% for the Face Recognition Grand Challenge (FRGC) [40] and 97.71% for the Japanese Female Facial Expression (JAFFE) [41] datasets. Interested readers are referred to [32] and [33] for in-depth details.



**FIGURE 4.** An example of the Elliptical Higher-Order-Spectra Periocular Code. Reprinted with permission from [33].

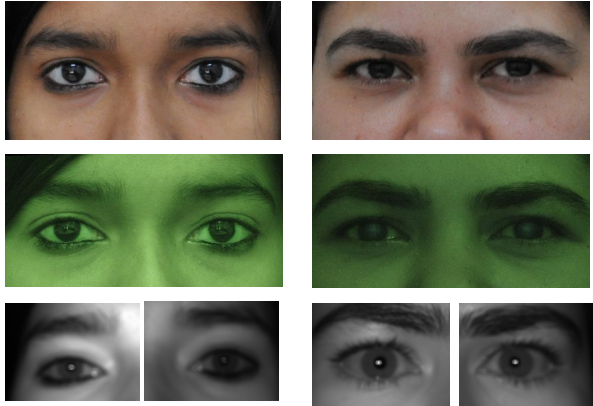
**C. MULTIMODAL COMPACT MULTI-LINEAR POOLING (MCM) FOR THE MULTISPECTRAL PERIOCLAR CLASSIFICATION TASK**

Based on the proposed MCM approach, we propose a complete system for combining and classifying multispectral periocular images as illustrated in Figure 3. There are three main phases in the classification process:

- *Feature extraction:* three periocular images of the same subject using three sensors of different wavelengths of the electromagnetic spectrum are passed through the feature extraction module to extract three eHPC [33] feature vectors correspondingly for three spectrum.
- *Feature fusion:* three feature vectors from three different spectra are then passed through the proposed multimodal compact multi-linear pooling fusion module to perform the feature-level fusion, generating one fused feature vector.
- *Classification:* the fused feature vector is passed through a one-hidden layer neural network for classification. A fully connected neural network layer of the same dimension to further increase the interaction between elements of the feature vector, before fed into a softmax layer for the classification task. A softmax classifier is a generalization of the binary form of Logistic Regression to multiple classes [43], which makes the outputs



**FIGURE 5.** A sample of the periocular region of one subject in our dataset with visible images in the first column, NightVision images in the second column, and Near-infrared in the third column.



**FIGURE 6.** Sample multispectral images of the periocular regions of two subject (1 and 12) in the IIIT multispectral periocular dataset. The first row shows visible spectrum images, the second row shows night vision images, and the third row shows NIR images.

interpretable as posterior probabilities for a categorical target variable. The probability assigned to the correct label,  $y_i$ , given the image,  $x_i$ , and parameterized by  $W$  is calculated as,

$$P(y_i|x_i; W) = \frac{e^{f_{y_i}}}{\sum_j e^{f_{y_j}}}.$$

This probabilistic interpretation makes the outputs intuitive.

## V. EXPERIMENTAL RESULTS

In this research, we choose two datasets for experiments:

- Our own dataset: we collected 212 images pertaining to 53 individuals captured using visible spectrum and NIR spectrum using a Sony DCR-DVD653E camera and NightVision spectrum using a IP2M-842B Wireless Network IP Surveillance camera. The size of each image is 800 x 600 pixels. A sample of the periocular region of one subject in our dataset is illustrated in Figure 5.
- IIIT Multispectral Periocular dataset [42]: this is the only public dataset containing multispectral images of perioculars. It contains 1240 images pertaining to 62 individuals captured using the visible spectrum, night vision, and NIR iris cameras. Similar to [42], the periocular region is resized to 260 x 270 pixel. Sample multispectral images of the periocular regions of two subject in the IIIT multispectral periocular dataset are illustrated in Figure 6.

### A. PERFORMANCE OF THE eHPC FEATURES COUPLED WITH ONE HIDDEN-LAYER NEURAL NETWORK CLASSIFIER

First, performance of the eHPC features for each modality in both datasets are investigated. The eHPC features coupled

with one-hidden layer neural network classifier lead to promising results on both datasets. We achieved the classification accuracies of 89.1%, 84.5% and 79.6% for visible spectrum, NIR spectrum and NightVision spectrum respectively for our own dataset. For the IIIT dataset, we achieved 83.8%, 74.2% and 78.7% for those spectra respectively. It is noteworthy that these results are promising considering the perioculars in both datasets are very challenging with various facial expressions, illuminations, resolutions and particularly head rotations. In the IIIT dataset, the NIR images are usually blurred, consequentially leading to lower classification results in comparison with other spectra. The accuracies are summarized in Table 1.

**TABLE 1.** Performance of the eHPC features coupled with one hidden-layer neural network classifier for each spectrum on the two datasets.

	Visible	NIR	Night Vision
Own	89.1	84.5	79.6
IIIT	83.8	74.2	78.7

### B. COMPARISON WITH OTHER FEATURE-LEVEL FUSION APPROACHES

To investigate the performance of the proposed feature-level fusion, we compare with two popular feature-level fusion categories: Concatenation (Concat) and Element-wise Multiplication (EWM). The proposed Multi-modal Compact Multi-linear pooling fusion approach outperforms both of them in both datasets. These superior accuracies attribute to the advantages of the outer product in the low dimensional feature space generated by the Count Sketch projection. We achieved 2.0% and 2.7% improvements in comparison with Concat and EWM respectively on our own dataset. We recorded higher gains for the IIIT dataset with 5.5% and 4.7%.

We also compare with two most recent proposals on the feature-level fusion: Canonical Correlation Analysis (CCA) and Discriminant Correlation Analysis (DCA). Our proposal outperforms both of them in terms of the accuracy but with less gains than two traditional categories. On our own dataset, the gains are 1.1% and 0.8% in comparison with CCA and DCA respectively. On the IIIT dataset, the gains are higher with 2.3% and 1.7% respectively. The comparison results are presented in Table 2.

### C. COMPARISON WITH OTHER-LEVEL FUSION APPROACHES

As discussed in the Introduction, four levels of fusion can be applied for combining multiple modalities. For score-level fusion, we implemented weighted sum of the scores. The optimal weights are searched for by grid search. For sensor-level fusion, we took the mean of component images in different spectra. For decision-level, we ranked the component decisions and took a majority vote for the final decision.

**TABLE 2. Comparison of the proposed feature-level fusion with other fusion approaches on the two datasets: two popular categories: concatenation (Concat) and element-wise multiplication (EWM); and two state-of-the-art approaches: canonical correlation analysis (CCA) and discriminant correlation analysis (DCA).**

	Concat	EWM	CCA [31]	DCA [14]	Our MCM
Own	91.3	90.8	92.2	92.7	93.3
IIIT	86.3	87.1	89.5	90.1	91.8

**TABLE 3. Comparison of the proposed feature-level fusion with other-level fusion approaches on the two datasets.**

	Feature-level (Ours)	Score-level (Weighted sum)	Sensor-level (Mean)	Decision-level (Rank)
Own	93.3	91.2	88.8	90.1
IIIT	90.8	88.2	86.9	87.9

The proposed feature-level fusion consistently outperforms other-level fusions on both our own datasets and the IIIT dataset, acknowledging the effectiveness of our proposed feature-level fusion approach. The comparison results are presented in Table 3.

**VI. CONCLUSIONS**

In this paper, we have proposed a novel feature-level fusion approach to effectively fuse multispectral feature vectors of the periocular biometric trait. The proposed approach, which we named Multi-model Compact Multi-linear pooling, based on an efficient outer product calculation of feature vectors in a low dimensional space generated by the Counter Sketch projection, allowing all elements to interact, enabling multiplicative interaction with a small number of parameters and low computation complexity. Based on the proposed feature-level fusion, a complete multispectral periocular recognition system has also been demonstrated using the elliptical Higher-order-spectra features. The proposals have been validated in two datasets: our own periocular dataset and the IIIT periocular dataset. The proposed approach can also be applicable for other biometric modalities such as multispectral face recognition and multispectral iris recognition.

**REFERENCES**

[1] A. K. Jain, A. Ross, and S. Prabhakar, "An introduction to biometric recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 4–20, Jan. 2004.

[2] R. Jillela et al., "Iris segmentation for challenging periocular images," in *Handbook of Iris Recognition*, J. M. Burge and W. K. Bowyer, Eds. London, U.K.: Springer, 2013, pp. 281–308.

[3] R. Ambika, R. Radhika, and D. Sehsachalam, "Periocular recognition using reduced features," *Int. J. Comput. Sci. Inf. Technol.*, vol. 5, no. 4 pp. 151–158, 2013.

[4] G. Lu, Y. Gong, D. Zhang, Z. Guo, and L. Zhang, *Multispectral Biometrics*. New York, NY, USA: Springer-Verlag, 2016.

[5] T. Bourlai and B. Cukic, "Multi-spectral face recognition: Identification of people in difficult environments," in *Proc. IEEE Int. Conf. Intell. Secur. Inform. (ISI)*, Jun. 2012, pp. 196–201.

[6] T. Bourlai, Ed., *Face Recognition Across the Imaging Spectrum*, vol. 1. Cham, Switzerland: Springer-Verlag, 2016.

[7] C. Boyce, A. Ross, M. Monaco, L. Hornak, and L. Xin, "Multispectral iris analysis: A preliminary study," in *Proc. Comput. Vis. Pattern Recognit. Workshop*, Jun. 2006, p. 51.

[8] H. T. Ngo, R. W. Ives, J. R. Mately, J. Dormo, M. Rhoads, and D. Choi, "Design and implementation of a multispectral iris capture system," in *Proc. Conf. Rec. 43rd Asilomar Conf. Signals, Syst. Comput.*, 2009, pp. 380–384.

[9] S. Crihalmeanu and A. Ross, "Multispectral scleral patterns for ocular biometric recognition," *Pattern Recogn. Lett.*, vol. 33, no. 14, pp. 1860–1869, 2012.

[10] S. Crihalmeanu and A. Ross, "Multispectral ocular biometrics," in *Face Recognition Across the Imaging Spectrum*. Cham, Switzerland: Springer, 2016, pp. 355–380.

[11] Z. Guo, D. Zhang, L. Zhang, and W. Liu, "Feature band selection for online multispectral palmprint recognition," *IEEE Trans. Inf. Forensics Security*, vol. 7, no. 3, pp. 1094–1099, Jun. 2012.

[12] R. K. Rowe, K. A. Nixon, and P. W. Butler, "Multispectral fingerprint image acquisition," in *Advances in Biometrics*. London, U.K.: Springer, 2008, pp. 3–23.

[13] A. Ross, A. Jain, and K. Nandakumar, "Levels of fusion in biometrics," in *Handbook of Multibiometrics*. New York, NY, USA: Springer, 2006, pp. 59–90.

[14] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 9, pp. 1984–1996, Sep. 2016.

[15] T.-Y. Lin, A. RoyChowdhury, and S. Maji, "Bilinear CNN models for fine-grained visual recognition," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1449–1457.

[16] Y. Gao, O. Beijbom, N. Zhang, and T. Darrell, "Compact bilinear pooling," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 317–326.

[17] A. Fukui, D. H. Park, D. Yang, A. Rohrbach, T. Darrell, and M. Rohrbach. (2016). "Multimodal compact bilinear pooling for visual question answering and visual grounding." [Online]. Available: <https://arxiv.org/abs/1606.01847>

[18] J.-B. Delbrouck and S. Dupont, "Multimodal compact bilinear pooling for multimodal neural machine translation," presented at the Int. Conf. Learn. Represent. (ICLR) Workshop, 2017.

[19] U. Park, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum: A feasibility study," in *Proc. IEEE Int. Conf. BTAS*, Sep. 2009, pp. 1–6.

[20] U. Park, R. Jillela, A. Ross, and A. K. Jain, "Periocular biometrics in the visible spectrum," *IEEE Trans. Inf. Forensics Security*, vol. 6, no. 1, pp. 96–106, Mar. 2011.

[21] I. Nigam, M. Vatsa, and R. Singh, "Ocular biometrics: A survey of modalities and fusion approaches," *Inf. Fusion*, vol. 26, pp. 1–35, Nov. 2015.

[22] F. Alonso-Fernandez and J. Bigun, "A survey on periocular biometrics research," *Pattern Recognit. Lett.*, vol. 82, pp. 92–105, Oct. 2016.

[23] S. Bakshi, P. K. Sa, and B. Majhi, "A novel phase-intensive local pattern for periocular recognition under visible spectrum," *Biocybern. Biomed. Eng.*, vol. 35, no. 1, pp. 30–44, 2015.

[24] J. R. Lyle, P. E. Miller, S. J. Pundlik, and D. L. Woodard, "Soft biometric classification using local appearance periocular region features," *Pattern Recognit.*, vol. 45, no. 11, pp. 3877–3885, 2012.

[25] S. Bakshi, P. K. Sa, and B. Majhi, "Phase intensive global pattern for periocular recognition," in *Proc. Annu. IEEE India Conf. (INDICON)*, Dec. 2014, pp. 1–5.

[26] J. R. Lyle, P. E. Miller, S. J. Pundlik, and D. L. Woodard, "Soft biometric classification using periocular region features," in *Proc. 4th IEEE Int. Conf. Biometrics, Theory, Appl. Syst. (BTAS)*, Sep. 2010, pp. 1–7.

[27] S. Kumari, S. Bakshi, and B. Majhi, "Periocular gender classification using global ICA features for poor quality images," *Procedia Eng.*, vol. 38, pp. 945–951, Jun. 2012.

[28] C.-W. Tan and A. Kumar, "Towards online iris and periocular recognition under relaxed imaging constraints," *IEEE Trans. Image Process.*, vol. 22, no. 10, pp. 3751–3765, Oct. 2013.

- [29] J. Smereka and V. Kumar, "What is a 'good' periocular region for recognition?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recogn. Workshops*, Jun. 2013, pp. 117–124.
- [30] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. IEEE Int. Conf. Comput. Vis.*, vol. 2, Sep. 1999, pp. 1150–1157.
- [31] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Computer Vision—ECCV*, vol. 3951. A. Leonardis, H. Bischof, and A. Pinz, Eds. Berlin, Germany: Springer, 2006, pp. 404–417.
- [32] F. AlGashaam, K. Nguyen, W. Boles, and V. Chandran, "Periocular recognition under expression variation using higher order spectral features," in *Proc. Int. Conf. Image Vis. Comput. New Zealand (IVCNZ)*, 2015, pp. 1–6.
- [33] F. M. Algashaam, K. Nguyen, V. Chandran, and J. Banks, "Elliptical higher-order-spectra periocular code," *IEEE Access*, vol. 5, pp. 6978–6988, 2017.
- [34] A. Ross, K. Nandakumar, and A. K. Jain, "Introduction to multibiometrics," in *Handbook Biometrics*, A. K. Jain, P. Flynn, and A. A. Ross, Eds. New York, NY, USA: Springer, 2008, pp. 271–292.
- [35] P. Wild, P. Radu, and J. Ferryman, "On fusion for multispectral iris recognition," in *Proc. Int. Conf. Biometrics (ICB)*, 2015, pp. 31–37.
- [36] Q.-S. Sun, S.-G. Zeng, L. Yan, P.-A. Heng, and D.-S. Xia, "A new method of feature fusion and its application in image recognition," *Pattern Recogn.*, vol. 38, no. 12, pp. 2437–2448, 2005.
- [37] J. Liu, S. Zhang, S. Wang, and D. M. Metaxas, "Multispectral deep neural networks for pedestrian detection," in *Proc. Brit. Mach. Vis. Conf.*, York, U.K., 2016, p. 1.
- [38] J. Wagner, V. Fischer, M. Herman, and S. Behnke, "Multispectral pedestrian detection using deep fusion convolutional neural networks," in *Proc. 24th Eur. Sympo. Artif. Neural Netw., Comput. Intell. Mach. Learn. (ESANN)*, Bruges, Belgium, Apr. 2016, pp. 509–514.
- [39] J. B. Tenenbaum and W. T. Freeman, "Separating style and content with bilinear models," *Neural Comput.*, vol. 12, no. 6, pp. 1247–1283, Jun. 2000.
- [40] NIST. *Face Recognition Grand Challenge Database*. Accessed on Jun. 20, 2017. [Online]. Available: <http://www.nist.gov/itl/iad/ig/frgc.cfm>
- [41] M. N. Dailey et al., "Evidence and a computational explanation of cultural differences in facial expression recognition," *Emotion*, vol. 10, no. 6, pp. 874–893, 2010.
- [42] A. Sharma, S. Verma, M. Vatsa, and R. Singh, "On cross spectral periocular recognition," in *Proc. IEEE Int. Conf. Image Process.*, Oct. 2014, pp. 5007–5011.
- [43] Y. Tang. (2013). "Deep learning using linear support vector machines." [Online]. Available: <https://arxiv.org/abs/1306.0239v4>



**FAISAL M. ALGASHAAM** (GS'17) received the B.Eng. degree from the Queensland University of Technology (QUT) in 2012 and the M.Eng. degree from The University of Queensland in 2013. He is currently pursuing the Ph.D. degree at the School of Electrical Engineering and Computer Science, QUT. His research interests include wireless networks, renewable energy, spectrum sensing, biometrics, image processing, and computer vision.



**KIEN NGUYEN** (S'09–M'16) received the B.Eng. and M.Sc. degrees in computer engineering from the Hanoi University of Technology, Vietnam, in 2005 and 2008, respectively, and the Ph.D. degree from the Queensland University of Technology in 2013. His Ph.D. research was on less-constrained human identification at a distance using iris. He is currently a Post-Doctoral Research Fellow with the Speech, Audio, Image, and Video Technology Laboratory, with a

research focus on computer vision, biometrics, deep learning, and scene understanding.



**MOHAMED ALKANHAL** received the B.Sc. degree (Hons.) in computer engineering from King Saud University, Riyadh, Saudi Arabia, the M.Sc. degree from the University of Southern California, Los Angeles, CA, USA, and the Ph.D. degree from Carnegie Mellon University, Pittsburgh, PA, USA. He was with the Computer Research Institute, King Abdulaziz City for Science and Technology (KACST), Riyadh, as a Research Faculty Member from 2001 to 2006 and the Director from 2006 to 2014. He is currently the Director of the Communications and Information Technology Research Institute, KACST. He has authored or co-authored many technical papers and has three patents pending. He has done extensive research in several areas, including pattern recognition, digital signal processing, and natural language processing.



**VINOD CHANDRAN** (M'90–SM'01) received the B.S. degree in electrical engineering from IIT Madras, Chennai, the M.S. degree in electrical engineering from Texas Tech University, the M.S. degree in computer science from Washington State University, and the Ph.D. degree in electrical and computer engineering from Washington State University in 1990. He is currently an Adjunct Professor with the Queensland University of Technology, Australia. He has supervised 14 Ph.D. students as the Principal Supervisor to completion and has authored or co-authored over 170 journal and conference papers. His research contributions span signal processing, image processing, and pattern recognition with applications to biometrics and biomedical systems.



**WAGEEH BOLES** (M'13) is currently a Professor with the School Electrical Engineering, Computer Science, Computational Intelligence and Signal Processing, Science and Engineering Faculty, Queensland University of Technology, Brisbane. His research interests include biometric identification using iris and palm features, object recognition, robot vision, technology and education, work integrated learning, curriculum design, and the study and utilization of learners' cognitive styles.



**JASMINE BANKS** (M'14) received the B.Eng. degree in electronics and the B.I.T. and Ph.D. degrees from the Queensland University of Technology in 1993 and 2000, respectively. She is currently a Lecturer with the School of Engineering Systems, Queensland University of Technology. Her research interests include image processing, computer vision, sensor fusion, and electronics and reconfigurable hardware.

...