DISCRIMINANT CORRELATION ANALYSIS FOR FEATURE LEVEL FUSION WITH APPLICATION TO MULTIMODAL BIOMETRICS

M. Haghighat [a], *M. Abdel-Mottaleb* [ab], *W. Alhalabi* [b]

[a] Department of Electrical and Computer Engineering, University of Miami
[b] Department of Computer Science, Effat University

ABSTRACT

In this paper, we present Discriminant Correlation Analysis (DCA), a feature level fusion technique that incorporates the class associations in correlation analysis of the feature sets. DCA performs an effective feature fusion by maximizing the pair-wise correlations across the two feature sets, and at the same time, eliminating the between-class correlations and restricting the correlations to be within classes. Our proposed method can be used in pattern recognition applications for fusing features extracted from multiple modalities or combining different feature vectors extracted from a single modality. It is noteworthy that DCA is the first technique that considers class structure in feature fusion. Moreover, it has a very low computational complexity and it can be employed in real-time applications. Multiple sets of experiments performed on various biometric databases show the effectiveness of our proposed method, which outperforms other state-of-the-art approaches.

Index Terms— multimodal biometric identification, feature level fusion, class structure, correlation analysis.

1. INTRODUCTION

Fusion of information can occur at different levels of a recognition system, *i.e.*, at the feature level, matching-score level, or decision level. However, feature level fusion is believed to be more effective owing to the fact that a feature set contains richer information about the input biometric data than the matching score or the output decision of a classifier [1].

Three well-known and typical feature fusion methods are: serial feature fusion [2], parallel feature fusion [3], and feature fusion based on Canonical Correlation Analysis (CCA) [4]. Serial feature fusion works by simply concatenating two sets of feature vectors into a single feature vector. If the first source feature vector, x, is p-dimensional and the second source feature vector, y, is q-dimensional, the fused feature vector, z, will be $(p+q)$-dimensional. Parallel feature fusion, on the other hand, combines the two source feature vectors into a complex vector $z{=}x{+}\mathrm{i}y$ (i being an imaginary unit). Note that if the dimensions of the two input vectors are not equal, the one with the lower dimension is padded with zeros. CCA-based feature fusion uses the correlation between two sets of features to find two sets of transformations such that the transformed features have maximum correlation across the two feature sets, while being uncorrelated within each feature set. Recently, CCA-based methods have become popular and other related and improved methods have also been proposed [5–8].

In this paper, we propose a feature fusion method that considers the class associations in feature sets. Our method, called *Discriminant Correlation Analysis (DCA)*, eliminates the between-class correlations and restricts the correlations to be within classes. DCA provides the characteristics of the CCA-based methods in maximizing the correlation of corresponding features across the two feature sets and in addition decorrelates features that belong to different classes within each feature set. Moreover, our method is computationally efficient and it does not have the small sample size (SSS) problem faced by the CCA-based algorithms. Although our method is general and can be used for any recognition system, in our experiments we focus on multimodal biometrics.

This paper is organized as follows: Section 2 describes the CCA-based feature level fusion method and its properties. Section 3 presents our proposed discriminant correlation analysis. The implementation details and experimental results are presented in Section 4. Finally, Section 5 concludes the paper.

2. FEATURE-LEVEL FUSION USING CANONICAL CORRELATION ANALYSIS

Suppose that $X \in \mathbb{R}^{p\times n}$ and $Y \in \mathbb{R}^{q\times n}$ denote two matrices, each contains n training feature vectors from two different modalities. Let $S_{xx} \in \mathbb{R}^{p\times p}$ and $S_{yy} \in \mathbb{R}^{q\times q}$ denote the within-sets covariance matrices of X and Y and $S_{xy} \in \mathbb{R}^{p\times q}$ denote the between-set covariance matrix (note that $S_{yx} = S_{xy}^T$). CCA aims to find the linear combinations, $X^* = W_x^T X$ and $Y^* = W_y^T Y$, that maximize the pair-wise correlations across the two feature sets. The transformation matrices, W_x and W_y, are found by solving the eigenvalue equations [9]:

$$\begin{cases} S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx} \hat{W}_x = R^2 \hat{W}_x \\ S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy} \hat{W}_y = R^2 \hat{W}_y \end{cases}, \tag{1}$$

where $\hat{W}_x$ and $\hat{W}_y$ are the eigenvectors and R^2 is the diagonal matrix of eigenvalues or squares of the *canonical correla-*

tions. The number of non-zero eigenvalues in each equation is $d = rank(S_{xy}) \leq min(n,p,q)$, which will be sorted in decreasing order, $r_1 \geq r_1 \geq \ldots \geq r_d$. The transformation matrices, W_x and W_y, consist of the sorted eigenvectors corresponding to the non-zero eigenvalues. $X^*, Y^* \in \mathbb{R}^{d \times n}$ are known as canonical variates.

As defined in [4], feature-level fusion is performed either by concatenation or summation of the transformed feature vectors:

$$Z_1 = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} W_x^T X \\ W_y^T Y \end{pmatrix} = \begin{pmatrix} W_x & 0 \\ 0 & W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{2}$$

or

$$Z_2 = X^* + Y^* = W_x^T X + W_y^T Y = \begin{pmatrix} W_x \\ W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \tag{3}$$

where Z_1 and Z_2 are called the *Canonical Correlation Discriminant Features (CCDFs)*.

3. FEATURE-LEVEL FUSION USING DISCRIMINANT CORRELATION ANALYSIS

The feature fusion method described above has two disputable issues. The first issue is encountered in case of a small sample size problem. In many real world applications, the number of samples is usually less than the number of features ($n < p$ or $n < q$). This makes the covariance matrices S_{xx} and S_{yy} singular and non-invertible. A solution to overcome this issue is to consider a two stage PCA + CCA approach [6].

The second issue is that CCA-based approaches neglect the class structure among samples. CCA decorrelates the features, but in pattern recognition problems, we are also interested in separating the classes. The dimensionality reduction approaches based on Linear Discriminant Analysis (LDA) [10] consider this matter by finding projections that best separate the classes. However, a *two stage* LDA + CCA will not be an effective solution due to the fact that the transformation applied by the second stage, *i.e.*, CCA, will not preserve the properties achieved by the first stage, *i.e.*, LDA. Therefore, we need transformations to not only maximize the pair-wise correlations across the two feature sets, but also to *simultaneously* separate the classes within each feature set. In this section, we present a solution to achieve this goal. Our proposed approach, called Discriminant Correlation Analysis (DCA), is described below.

Let's assume that the samples in the data matrix are collected from c separate classes. Accordingly, the n columns of the data matrix are divided into c separate groups, n_i columns comprising the i^{th} class ($n = \sum_{i=1}^{c} n_i$). Let $x_{ij} \in X$ denote the feature vector corresponding to the j^{th} sample in the i^{th} class. $\bar{x}_i$ and $\bar{x}$ denote the means of the x_{ij} vectors in the i^{th} class and in the whole feature set, respectively. That is, $\bar{x}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} x_{ij}$ and $\bar{x} = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{n_i} x_{ij} = \frac{1}{n}\sum_{i=1}^{c} n_i \bar{x}_i$. The between-class scatter matrix is defined as

$$S_{bx_{(p\times p)}} = \sum_{i=1}^{c} n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T = \Phi_{bx}\Phi_{bx}^T, \tag{4}$$

where

$$\Phi_{bx_{(p\times c)}} = \left[\sqrt{n_1}(\bar{x}_1 - \bar{x}), \sqrt{n_2}(\bar{x}_2 - \bar{x}), \ldots, \sqrt{n_c}(\bar{x}_c - \bar{x})\right]. \tag{5}$$

If the number of features is higher than the number of classes ($p \gg c$), it is computationally easier to calculate the covariance matrix as $(\Phi_{bx}^T \Phi_{bx})_{c\times c}$ rather than $(\Phi_{bx}\Phi_{bx}^T)_{p\times p}$. As presented in [11], the most significant eigenvectors of $\Phi_{bx}\Phi_{bx}^T$ can be efficiently obtained by mapping the eigenvectors of $\Phi_{bx}^T\Phi_{bx}$. Therefore, we only need to find the eigenvectors of the $c \times c$ covariance matrix $\Phi_{bx}^T\Phi_{bx}$.

If the classes were well-separated, $\Phi_{bx}^T\Phi_{bx}$ would be a diagonal matrix. Since $\Phi_{bx}^T\Phi_{bx}$ is symmetric positive semi-definite, we can find transformations that diagonalize it:

$$P^T (\Phi_{bx}^T\Phi_{bx}) P = \hat{\Lambda}, \tag{6}$$

where P is the matrix of orthogonal eigenvectors and $\hat{\Lambda}$ is the diagonal matrix of real and non-negative eigenvalues sorted in decreasing order.

Let $Q_{(c\times r)}$ consist of the first r eigenvectors, which correspond to the r largest non-zero eigenvalues, from matrix P. We have:

$$Q^T (\Phi_{bx}^T\Phi_{bx}) Q = \Lambda_{(r\times r)}. \tag{7}$$

The r most significant eigenvectors of S_{bx} can be obtained with the mapping: $Q \rightarrow \Phi_{bx}Q$ [11]:

$$(\Phi_{bx}Q)^T S_{bx} (\Phi_{bx}Q) = \Lambda_{(r\times r)}. \tag{8}$$

$W_{bx} = \Phi_{bx}Q\Lambda^{-1/2}$ is the transformation that unitizes S_{bx} and reduces the dimensionality of the data matrix, X, from p to r. That is:

$$W_{bx}^T S_{bx} W_{bx} = I, \tag{9}$$

$$X'_{(r\times n)} = W_{bx_{(r\times p)}}^T X_{(p\times n)}. \tag{10}$$

X' is the projection of X in a space, where the between-class scatter matrix is I and the classes are separated. Note that there are at most $c-1$ nonzero generalized eigenvalues; therefore, an upper bound on r is $c-1$ [12]. Other upper bounds for r are the ranks of the data matrices, *i.e.*, $r \leq min(c-1, rank(X), rank(Y))$.

Similar to the above approach we solve for the second feature set, Y, and find a transformation matrix W_{by}, which unitizes the between-class scatter matrix for the second modality, S_{by} and reduces the dimensionality of Y from q to r:

$$W_{by}^T S_{by} W_{by} = I, \tag{11}$$

$$Y'_{(r\times n)} = W_{by_{(r\times q)}}^T Y_{(q\times n)}. \tag{12}$$

The updated Φ'_{bx} and Φ'_{by} are non-square $r \times c$ orthonormal matrices. Although $S'_{bx} = S'_{by} = I$, the matrices ${\Phi'_{bx}}^T \Phi'_{bx}$ and

${\Phi'_{by}}^T \Phi'_{by}$ are strict diagonally dominant matrices, where the diagonal elements are close to one and the non-diagonal elements are close to zero. This makes the centroids of the classes have minimal correlation with each other, and thus, the classes are separated.

Now that we have transformed X and Y to X' and Y', where the between-class scatter matrices are unitized, we need to make the features in one set have nonzero correlation only with their corresponding features in the other set. To achieve this, we need to diagonalize the between-set covariance matrix of the transformed feature sets, $S'_{xy} = X'{Y'}^T$. We use singular value decomposition (SVD) to diagonalize S'_{xy}:

$$S'_{xy_{(r\times r)}} = U\Sigma V^T \quad \Rightarrow \quad U^T S'_{xy} V = \Sigma. \tag{13}$$

Note that X' and Y' are of rank r and $S'_{xy_{(r\times r)}}$ is nondegenerate. Therefore, Σ is a diagonal matrix whose diagonal elements are non-zero. Let $W_{cx} = U\Sigma^{-1/2}$ and $W_{cy} = V\Sigma^{-1/2}$, we have:

$$(U\Sigma^{-1/2})^T S'_{xy} (V\Sigma^{-1/2}) = I, \tag{14}$$

which unitizes the between-set covariance matrix, S'_{xy}. Now, we transform the feature sets as follows:

$$X^* = W_{cx}^T X' = \underbrace{W_{cx}^T W_{bx}^T}\, X = W_x X, \tag{15}$$

$$Y^* = W_{cy}^T Y' = \underbrace{W_{cy}^T W_{by}^T}\, Y = W_y Y. \tag{16}$$

where $W_x = W_{cx}^T W_{bx}^T$ and $W_y = W_{cy}^T W_{by}^T$ are the final transformation matrices for X and Y, respectively.

It can be easily shown that the between-class scatter matrices of the transformed feature sets are still diagonal; hence, the classes are separated. The between-class scatter matrix for X^* is calculated as:

$$S^*_{bx} = W_{cx}^T \underbrace{W_{bx}^T S_{bx} W_{bx}}\, W_{cx}. \tag{17}$$

From Eq. (9), $W_{bx}^T S_{bx} W_{bx} = I$ and since U is an orthogonal matrix, we have:

$$S^*_{bx} = (U\Sigma^{-\frac{1}{2}})^T (U\Sigma^{-\frac{1}{2}}) = \Sigma^{-1}. \tag{18}$$

Similarly, we can show that $S^*_{by} = \Sigma^{-1}$, which is diagonal.

Similar to the CCA method, feature level fusion can be performed either by concatenation or summation of the transformed feature vectors, as shown in Eqs. (2) and (3). In our experiments, we use the summation method, shown in Eq. (3), for both CCA and DCA approaches. In case of having more than two sets of features, we follow a cascade approach and apply DCA on two sets of features at a time.

4. EXPERIMENTS AND ANALYSIS

In this paper, we present two sets of experiments to demonstrate the performance of our proposed technique in combining feature vectors extracted from different biometric modalities. Section 4.1 presents experiments on the fusion of fingerprint and iris modalities from Multimodal Biometric Dataset Collection, BIOMDATA [13]; and Section 4.2 presents experiments on fusing information from weak biometric modalities, *i.e.*, periocular, mouth, and nose regions, extracted from face images in AR face database [14].

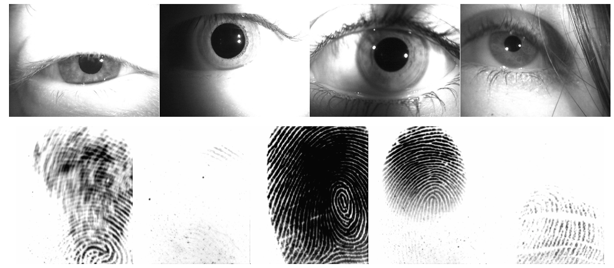

Fig. 1. Examples of challenging samples in BIOMDATA database. The images are corrupted with blur, occlusion, shadows, and sensor noise.

4.1. Multimodal Fusion: : BIOMDATA Multimodal Biometric Dataset

In this set of experiments, we use the multimodal biometric dataset (BIOMDATA) collected in West Virginia University [13]. This dataset is challenging, as many of the samples suffer from various artifacts such as blur, shadows, and sensor noise, as shown in Fig. 1. Following the experimental setting in [15], we chose iris and fingerprint modalities for our experiments. All the evaluations are performed on a subset of 219 subjects that have samples in both modalities. In total, there are two iris and four fingerprint modalities.

We segmented the iris images using the method proposed in [16]. Iris regions are then normalized and 25×240 bit-wise iris templates are generated by extracting log-Gabor features using the publicly available source code of Masek and Kovesi [17]. On the other hand, we enhanced the fingerprint images using the filtering methods described in [18]. Then, the core points of the fingerprints are detected [19] and Gabor features in eight orientations are extracted around the core points [20–22].

Four samples randomly chosen from each modality are used for training and the remaining samples are used for testing. The recognition results are averaged over five runs. Table 1 shows the rank-1 recognition rate for the individual iris and fingerprint modalities, and Table 2 shows the multimodal fusion results. We compare the proposed feature level fusion technique with the serial feature fusion, the parallel feature fusion, the CCA-based feature fusion, and the most recently published Joint Sparse Representation Classification (JSRC) [15] method. In order to prevent the SSS problem in the CCA-based approach, dimensionality reductions based on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are applied [6,11]. PCA and LDA are also used for dimensionality reduction of the results of the serial and parallel methods. Except for the JSRC method, which is restricted to work with a sparse representation classifier, all other experiments in this paper use a simple KNN classifier with $K = 1$

Table 1. Rank-1 recognition rates obtained by a KNN classifier using individual modalities in BIOMDATA database.

Modality	Recognition Rate
Iris (Left)	51.29
Iris (Right)	57.33
Fingerprint (Left thumb)	78.22
Fingerprint (Left index)	90.10
Fingerprint (Right thumb)	79.60
Fingerprint (Right index)	91.29

Table 2. Rank-1 recognition rates for multimodal fusion of iris and fingerprint biometrics in BIOMDATA database.

Method \ Modality	2 Irises	4 Fingerprints	All 6 Modalities
Serial + PCA	62.48	94.46	94.85
Serial + LDA	70.31	96.22	96.22
Parallel + PCA	68.22	-	-
Parallel + LDA	72.25	-	-
PCA + CCA	78.51	96.32	97.20
LDA + CCA	78.90	96.40	97.51
JSRC	78.20	97.60	98.60
DCA	84.16	98.71	99.60

for classification. Note that in case of more than two modalities (four fingerprints or all six modalities) parallel feature fusion method cannot be applied and Multiset-CCA [5] is used. Table 2 shows that the proposed DCA technique outperforms the other fusion methods.

4.2. Multimodal Fusion: AR Face Database

In this set of experiments, we show the applicability of the proposed DCA algorithm in fusing information from weak biometric modalities extracted from face images. These modalities include left and right periocular, mouth, and nose regions, as shown in Fig. 2. It was shown that the periocular regions, nose and mouth can be considered as useful biometrics [23, 24]; however, they are not as discriminative as the whole face [15].

We evaluated our algorithms on a set of 100 subjects from AR face database [14, 25]. The AR face database consists of frontal face images with varying facial expressions and illumination. The face images are captured in two sessions. Similar to the setup in [15], seven images of each subject from the first session are used for training and seven images from the second session are used for testing. Gabor features in five scales and eight orientations are extracted from all modalities.

Table 3 shows the rank-1 recognition rates for the individual modalities. The major challenge here is to be able to fuse weak modalities with a strong modality based on the whole face without deteriorating the accuracy of the strong modality [26]. Table 4 shows the recognition rates for different feature fusion methods using combinations of different modalities. It is obvious that the proposed method has a higher recognition rate than the other feature level fusion techniques. Moreover, the results show that adding more modalities increases the accuracy of the multimodal system over the performance of all the individual modalities.

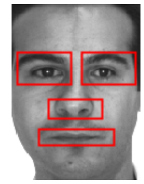

Fig. 2. Face mask used to crop out different modalities.

Table 3. Rank-1 recognition rates obtained by a KNN classifier using individual modalities in AR database. Modalities include 1. left periocular, 2. right periocular, 3. nose, 4. mouth, and 5. face.

Modality	1	2	3	4	5
Recognition Rate	84.14	84.29	73.57	74.29	90.57

Table 4. Rank-1 recognition rates for multimodal fusion in AR database.

Method \ Modality	{1,2}	{1,2,3}	{1,2,3,4}	{1,2,3,4,5}
Serial + PCA	85.57	88.71	90.42	90.71
Serial + LDA	89.43	92.14	92.86	93.57
PCA + CCA	90.57	92.86	94.43	96.57
LDA + CCA	91.28	92.57	93.71	97.00
JSRC	92.14	92.86	94.43	98.57
DCA	92.71	93.28	97.43	99.14

5. CONCLUSIONS

In this paper, we presented a feature fusion technique based on correlation analysis of the feature sets. Our proposed method, called Discriminant Correlation Analysis, contemplates the class associations of the samples in its analysis. It aims to find transformations that maximize the pair-wise correlations across the two feature sets and at the same time separate the classes within each set. These characteristics make DCA an effective feature fusion tool for pattern recognition applications. Experimental results demonstrated the efficacy of our proposed approach in fusion of multimodal feature sets or different feature sets extracted from a single modality. In addition, the DCA method is computationally efficient and can be employed in real-time applications.

6. REFERENCES

[1] A. Ross and A.K. Jain, "Multimodal biometrics: An overview," in *12th European Signal Processing Conference (EUSIPCO)*, 2004, pp. 1221–1224.

[2] C. Liu and H. Wechsler, "A shape-and texture-based enhanced fisher classifier for face recognition," *IEEE*

Transactions on Image Processing, vol. 10, no. 4, pp. 598–608, 2001.

[3] J. Yang, J. Yang, D. Zhang, and J. Lu, "Feature fusion: parallel strategy vs. serial strategy," *Pattern Recognition*, vol. 36, no. 6, pp. 1369–1381, 2003.

[4] Q.S. Sun, S.G. Zeng, Y. Liu, P.A. Heng, and D.S. Xia, "A new method of feature fusion and its application in image recognition," *Pattern Recognition*, vol. 38, no. 12, pp. 2437–2448, 2005.

[5] Y.O. Li, T. Adali, W. Wang, and V.D. Calhoun, "Joint blind source separation by multiset canonical correlation analysis," *IEEE Transactions on Signal Processing*, vol. 57, no. 10, pp. 3918–3929, 2009.

[6] N.M. Correa, T. Adali, Y. Li, and V.D. Calhoun, "Canonical correlation analysis for data fusion and group inferences," *IEEE Signal Processing Magazine*, vol. 27, no. 4, pp. 39–50, 2010.

[7] K.H. Pong and K.M. Lam, "Multi-resolution feature fusion for face recognition," *Pattern Recognition*, vol. 47, no. 2, pp. 556–567, 2014.

[8] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Fully automatic face normalization and single sample face recognition in unconstrained environments," *Expert Systems with Applications*, vol. 47, pp. 23–34, 2016.

[9] W.J. Krzanowski, *Principles of multivariate analysis: a user's perspective*, Oxford University Press, Inc., 1988.

[10] P.N. Belhumeur, J.P. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[11] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[12] R.O. Duda and P.E. Hart, *Pattern classification and scene analysis*, Wiley New York, 1973.

[13] S. Crihalmeanu, A. Ross, S. Schuckers, and L. Hornak, "A protocol for multibiometric data acquisition, storage and dissemination," *Technical Report, WVU, Lane Department of Computer Science and Electrical Engineering*, 2007.

[14] A. Martinez and R. Benavente, "The AR face database," *CVC Technical Report*, vol. 24, 1998.

[15] S. Shekhar, V.M. Patel, N.M. Nasrabadi, and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 113–126, 2014.

[16] S.J. Pundlik, D.L. Woodard, and S.T. Birchfield, "Non-ideal iris segmentation using graph cuts," in *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2008, pp. 1–6.

[17] L. Masek and P. Kovesi, "Matlab source code for a biometric identification system based on iris patterns," *The School of Computer Science and Software Engineering, The University of Western Australia*, vol. 26, 2003.

[18] S. Chikkerur, C. Wu, and V. Govindaraju, "A systematic approach for feature extraction in fingerprint images," in *Biometric Authentication*, pp. 344–350. Springer, 2004.

[19] A.K. Jain, S. Prabhakar, L. Hong, and S. Pankanti, "Filterbank-based fingerprint matching," *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 846–859, 2000.

[20] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002.

[21] M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb, "Identification using encrypted biometrics," in *Computer Analysis of Images and Patterns (CAIP)*. Springer, 2013, pp. 440–448.

[22] M. Haghighat, S. Zonouz, and M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," *Expert Systems with Applications*, vol. 42, no. 21, pp. 7905–7916, 2015.

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

[24] M. Balasubramanian, S. Palanivel, and V. Ramalingam, "Real time face and mouth recognition using radial basis function neural networks," *Expert Systems with Applications*, vol. 36, no. 3, pp. 6879–6888, 2009.

[25] A.M. Martinez and A.C. Kak, "PCA versus LDA," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 228–233, 2001.

[26] H. Li, K.A. Toh, and L. Li, *Advanced topics in biometrics*, World Scientific, 2012.