Discriminant Correlation Analysis: Real-Time Feature Level Fusion for Multimodal Biometric Recognition

Mohammad Haghighat*, *Member, IEEE,* Mohamed Abdel-Mottaleb, *Fellow, IEEE,*
and Wadee Alhalabi, *Member, IEEE*

***Abstract*—Information fusion is a key step in multimodal biometric systems. 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. The goal of feature fusion for recognition is to combine relevant information from two or more feature vectors into a single one with more discriminative power than any of the input feature vectors. In pattern recognition problems, we are also interested in separating the classes. In this paper, we present Discriminant Correlation Analysis (DCA), a feature level fusion technique that incorporates the class associations into the correlation analysis of the feature sets. DCA performs an effective feature fusion by maximizing the pairwise correlations across the two feature sets, and at the same time, eliminating the between-class correlations and restricting the correlations to be within the 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, and using different feature extraction techniques, 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, discriminant correlation analysis.**

I. Introduction

BIOMETRIC identifiers are distinctive and measurable characteristics used to label and describe individuals. Some of the well-known biometrics used for human identification are fingerprints, face, ear, iris, voice and DNA. Most of the real-world biometric systems, so-called *unimodal*, rely on the evidence of a single source of biometric information. *Multimodal* biometric systems, on the other hand, fuse multiple sources of biometrics information to make a more reliable recognition. Fusion of the biometrics information can occur at different stages of a recognition system. In case of *feature level fusion*, the data itself or the features extracted from multiple biometrics are fused. *Matching-score level fusion* consolidates the scores generated by multiple classifiers pertaining to different modalities. Finally, in case of *decision level fusion* the final results of multiple classifiers are combined via techniques such as majority voting [1]–[3].

Feature level fusion is believed to be more effective than the other levels of fusion because the feature set contains richer information about the input biometric data than the matching score or the output decision of a classifier. Therefore, fusion at the feature level is expected to provide better recognition results [3]–[5]. However, matching-score level fusion and decision level fusion are more popular in the literature and there is not much research on feature level fusion. The reason is the difficulty of feature level fusion in cases where the features are not compatible, *e.g.*, eigen-coefficients of faces and minutiae set of fingerprints, or when commercial biometric systems do not provide access to the feature sets (nor the raw data), which they use in their products [3]. The goal of the feature fusion for recognition is to combine relevant information from two or more feature vectors into a single one, which is expected to be more discriminative than any of the input feature vectors.

Two well-known and typical feature fusion methods are: serial feature fusion [6] and parallel feature fusion [7], [8]. Serial feature fusion works by simply concatenating two sets of feature vectors into a single feature vector. Obviously, 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+iy (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.

Recently, feature fusion based on Canonical Correlation Analysis (CCA) [9] has attracted the attention in the area of multimodal recognition. 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. This method is described in details in Section II. Recently, CCA-based methods have become popular and other related and improved methods have also been proposed [10]–[14].

Manuscript received Month xx, 20xx; revised Month xx, 20xx.

M. Haghighat and M. Abdel-Mottaleb are with the Department of Electrical and Computer Engineering, College of Engineering, University of Miami, Coral Gables, FL 33146 USA (emails: *haghighat@umiami.edu, mottaleb@miami.edu*).

W. Alhalabi is with the Department of Computer Science, Effat University, and the Department of Computer Science, King Abdulaziz University, Jeddah, Saudi Arabia (email: *wsalhalabi@kau.edu.sa*).

Kettenring [15] proposed a generalized extension of CCA for several sets of variables. Nielsen [16] improved Kettenrings method to present a multiset canonical correlation analysis (MCCA), which can be used to analyze relationships between more than two sets of variables. Although Kettenrings and Nielsens methods [15], [16] are able to analyze multi-group variables, they do not demonstrate the integral relation among the multi-set variables, and the constraints do not guarantee that the transformed variables are statistically uncorrelated [12]. Recently, Yuan *et al.* [17] proposed a multi-set integrated canonical correlation analysis (MICCA) framework for the multi-set problems. MICCA can distinctly express the integral correlation among multi-set features. However, it follows an iterative approach, which reduces its efficiency.

Most recently, sparse representation has attracted the interest of many researchers, both for reconstructive and discriminative tasks [18]–[20]. The assumption is that a query sample belonging to a specific class can be represented with a linear combination of the training samples from that class. Therefore, it aims to find a sparse vector having non-zero elements only in the indices corresponding to that class. As indicated in the definition of feature level fusion, "the feature sets originating from multiple biometric algorithms are consolidated into a single feature set" [21]. Although not following this definition in building a single feature set that can be used by any classifier, Joint Sparse Representation Classification (JSRC) [22] is considered as a feature level fusion technique. JSRC builds multiple corresponding dictionaries each using training samples of a modality. Having a query consisting of multiple modalities, it aims to find joint sparse vectors that share the same sparsity pattern and have non-zero values only in the indices corresponding to a mutual class in multiple modalities. That is, training samples of the same class from the different modalities are used to reconstruct the query data. Bahrampour *et al.* [23] improved the performance of this method by using a multimodal task-driven dictionary learning algorithm.

In this paper, we propose a feature fusion method that considers the class associations in feature sets[1]. Our method, called *Discriminant Correlation Analysis (DCA)*, eliminates the between-class correlations and restricts the correlations to be within classes. DCA has 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. To the best of our knowledge, no other feature fusion method in the literature considered the class structure, and our method is the first to incorporate the class structure into the feature level fusion. It is worth mentioning that our method does not have the small sample size (SSS) problem faced by the CCA-based algorithms. Moreover, we propose a multiset method to generalize DCA to be applicable to more than two sets of variables. *Multiset Discriminant Correlation Analysis (MDCA)* follows a cascade approach and applies DCA on two sets of variables at a time. Extensive experiments performed on several multimodal biometric databases verify the effectiveness of our proposed method, which outperforms the state-of-the-art feature level fusion techniques[2].

This paper is organized as follows: Section II describes the CCA-based feature level fusion method and its properties. Section III presents our proposed discriminant correlation analysis method. The implementation details and experimental results on several databases are presented in Section IV. Finally, Section V concludes the paper.

II. Feature-Level Fusion Using Canonical Correlation Analysis

Canonical correlation analysis (CCA) is one of the valuable multi-data processing methods, which has been widely used to analyze the mutual relationships between two sets of variables. 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. That is, for each sample, two feature vectors with p and q dimensions are extracted from the first and second modalities, respectively.

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 $S_{yx} = S_{xy}^T$). The overall $(p+q)\times(p+q)$ covariance matrix, S, contains all the information on associations between pairs of features:

$$S = \begin{pmatrix} cov(x) & cov(x,y) \\ cov(y,x) & cov(y) \end{pmatrix} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}. \qquad (1)$$

However, the correlation between these two sets of feature vectors may not follow a consistent pattern, and thus, understanding the relationships between these two sets of feature vectors from this matrix is difficult [25]. CCA aims to find the linear combinations, $\overset{*}{X}= W_x^T X$ and $\overset{*}{Y}= W_y^T Y$, that maximize the pair-wise correlations across the two feature sets:

$$corr(\overset{*}{X}, \overset{*}{Y}) = \frac{cov(\overset{*}{X}, \overset{*}{Y})}{var(\overset{*}{X}).var(\overset{*}{Y})}, \qquad (2)$$

where $cov(\overset{*}{X}, \overset{*}{Y}) = W_x^T S_{xy} W_y$, $var(\overset{*}{X}) = W_x^T S_{xx} W_x$ and $var(\overset{*}{Y}) = W_y^T S_{yy} W_y$. Maximization is performed using Lagrange multipliers by maximizing the covariance between $\overset{*}{X}$ and $\overset{*}{Y}$ subject to the constraints $var(\overset{*}{X}) = var(\overset{*}{Y}) = 1$. The transformation matrices, W_x and W_y, are then found by solving the eigenvalue equations [25]:

$$\begin{cases} S_{xx}^{-1} S_{xy} S_{yy}^{-1} S_{yx} \hat{W}_x = R^2 \hat{W}_x \\ S_{yy}^{-1} S_{yx} S_{xx}^{-1} S_{xy} \hat{W}_y = R^2 \hat{W}_y \end{cases}, \qquad (3)$$

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 correlations*. 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. $\overset{*}{X}$, $\overset{*}{Y} \in \mathbb{R}^{d\times n}$ are known as canonical variates. For the transformed data, the sample covariance matrix defined in Eq. (1) will be of the form:

[1] A preliminary version of this work appeared in ICASSP 2016 [24].

[2] The MATLAB® source code for the proposed DCA method is provided at http://www.mathworks.com/matlabcentral/fileexchange/55405.

$$\overset{*}{S}=\left(\begin{array}{cccc:cccc} 1 & 0 & \dots & 0 & r_1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & r_2 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & r_d \\ \hdashline r_1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & r_2 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & & \ddots & & \vdots & & \ddots & \\ 0 & 0 & \dots & r_d & 0 & 0 & \dots & 1 \end{array}\right).$$

The above matrix shows that the canonical variates have nonzero correlation only on their corresponding indices. The identity matrices in the upper left and lower right corners show that the canonical variates are uncorrelated within each feature set.

As defined in [9], feature-level fusion is performed either by concatenation or summation of the transformed feature vectors:

$$Z_1=\begin{pmatrix} \overset{*}{X} \\ \overset{*}{Y} \end{pmatrix}=\begin{pmatrix} W_x^T X \\ W_y^T Y \end{pmatrix}=\begin{pmatrix} W_x & 0 \\ 0 & W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (4)$$

or

$$Z_2=\overset{*}{X}+\overset{*}{Y}=W_x^T X+W_y^T Y=\begin{pmatrix} W_x \\ W_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}, \quad (5)$$

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

III. Incorporating Class Structure in Multivariate Correlation Analysis

The feature fusion method described in the previous section 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 singular and non-invertible. Therefore, we will face a major problem in inverting the S_{xx} and S_{yy} matrices used in Eq. (3). A solution to overcome this issue is to reduce the dimensionality of the feature vectors before applying CCA. Therefore, a two stage PCA + CCA approach can be considered [10].

The second issue in CCA-based approaches is their negligence of 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) [26] 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 that not only maximize the pair-wise correlations across the two feature sets, but also *simultaneously* separate the classes within each set of features. In this section, we present a solution to achieve this goal.

Correlation analysis and discriminant analysis have been previously used in a combined way in [27] and [28]. However, the problem definition and the presented methods are totally different from our problem setting and proposed technique. These methods do not consider the problem of multimodal recognition or feature level fusion, which is the problem discussed in our paper. In [27] and [28], the correlation analysis is used for the cross-domain matching problem in unimodal recognition systems. For example, [27] proposes a cross-view face recognition system, where the query face image is in a different view angle than the one given for enrollment. In the cross-domain matching problem, the correlation analysis aims to extract the correlated features from feature vectors of the different domains.

In our method, we incorporate the class structure, *i.e.*, memberships of the samples in classes, into the correlation analysis, which helps in highlighting the differences between classes and at the same time maximizing the pair-wise correlations between features across the two feature sets. This helps fusing the relevant information captured by different modalities in multimodal recognition systems. Our proposed approach, called Discriminant Correlation Analysis (DCA), is described below.

A. Feature-Level Fusion Using Discriminant Correlation Analysis

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, where n_i columns belong to 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 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

$$S_{bx_{(p\times p)}}=\sum_{i=1}^{c} n_i(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})^T=\Phi_{bx}\Phi_{bx}^T, \quad (6)$$

where

$$\Phi_{bx_{(p\times c)}}=\left[\sqrt{n_1}(\bar{x}_1-\bar{x}),\sqrt{n_2}(\bar{x}_2-\bar{x}),\dots,\sqrt{n_c}(\bar{x}_c-\bar{x})\right]. \quad (7)$$

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 [29], 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:

$$P^T\left(\Phi_{bx}^T\Phi_{bx}\right)P=\hat{\Lambda}, \quad (8)$$

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:

$$Q^T\left(\Phi_{bx}^T\Phi_{bx}\right)Q = \Lambda_{(r\times r)}\,. \tag{9}$$

The r most significant eigenvectors of S_{bx} can be obtained with the mapping: $Q \rightarrow \Phi_{bx}Q$ [29]:

$$\left(\Phi_{bx}Q\right)^T S_{bx}\left(\Phi_{bx}Q\right) = \Lambda_{(r\times r)}\,. \tag{10}$$

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

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

$$X'_{(r\times n)} = W_{bx_{(r\times p)}}^T X_{(p\times n)}\,. \tag{12}$$

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 $c-1$ [30]. Other upper bounds for r are the ranks of the data matrices, *i.e.*, $r \leq min\left(c-1, rank\left(X\right), rank\left(Y\right)\right)$.

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:

$$W_{by}^T S_{by} W_{by} = I\,, \tag{13}$$

$$Y'_{(r\times n)} = W_{by_{(r\times q)}}^T Y_{(q\times n)}\,. \tag{14}$$

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 $\left(\forall i: |a_{ii}| > \sum_{j\neq i}|a_{ij}|\right)$, 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}:

$$S'_{xy_{(r\times r)}} = U\Sigma V^T \quad \Rightarrow \quad U^T S'_{xy} V = \Sigma\,. \tag{15}$$

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

$$(U\Sigma^{-1/2})^T S'_{xy}(V\Sigma^{-1/2}) = I\,, \tag{16}$$

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

$$\overset{*}{X} = W_{cx}^T X' = \underbrace{W_{cx}^T W_{bx}^T}_{} X = W_x X\,, \tag{17}$$

$$\overset{*}{Y} = W_{cy}^T Y' = \underbrace{W_{cy}^T W_{by}^T}_{} Y = W_y Y\,. \tag{18}$$

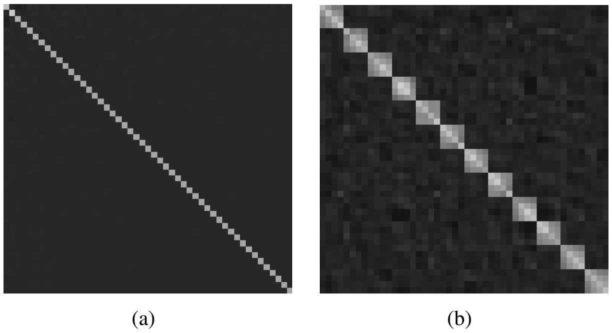

(a) (b)

Fig. 1. Visualization of covariance matrices (black color represents zero values and the elements with higher values are illustrated brighter). (a) Covariance between features ($\overset{*}{X}\overset{*}{X}^T$). (b) Covariance between samples ($\overset{*}{X}^T\overset{*}{X}$).

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 $\overset{*}{X}$ is calculated as:

$$\overset{*}{S}_{bx} = W_{cx}^T \underbrace{W_{bx}^T S_{bx} W_{bx}}_{} W_{cx}\,. \tag{19}$$

From Eq. (11), $W_{bx}^T S_{bx} W_{bx} = I$ and since U is an orthogonal matrix, we have:

$$\overset{*}{S}_{bx} = (U\Sigma^{-\frac{1}{2}})^T(U\Sigma^{-\frac{1}{2}}) = \Sigma^{-1}\,. \tag{20}$$

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

Fig. 1(a) shows the covariance between features in a transformed feature set ($\overset{*}{X}\overset{*}{X}^T$), which is a strict diagonally dominant matrix. Black color represents zero values and the elements with higher values are brighter. The results show that the correlation between different features in an individual feature set is minimal. On the other hand, Fig. 1(b) shows the covariance between samples in a transformed feature set ($\overset{*}{X}^T\overset{*}{X}$). Being a block diagonal matrix, Fig. 1(b) clearly shows that the samples have higher correlation with only the ones in the same class.

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. (4) and (5). However, the summation method has the advantage of lower number of dimensions, while the change in recognition results is very small. In our experiments, we use the summation method, shown in Eq. (5), for both CCA and DCA approaches.

B. Multiset Discriminant Correlation Analysis

Multiset Discriminant Correlation Analysis (MDCA) generalizes DCA to be applicable to more than two sets of features. Here, we assume that we have m sets of features, $X_i \in \mathbb{R}^{p_i\times n}, i = 1,2,\ldots,m$, which are sorted by their rank, that is $rank(X_1) \geq rank(X_2) \geq \ldots \geq rank(X_m)$. MDCA applies DCA on two sets of features at a time. Based on the approach presented in the previous section, the maximum length of the fused feature vector is $min\left(c-1, rank\left(X_i\right), rank\left(X_j\right)\right)$. In

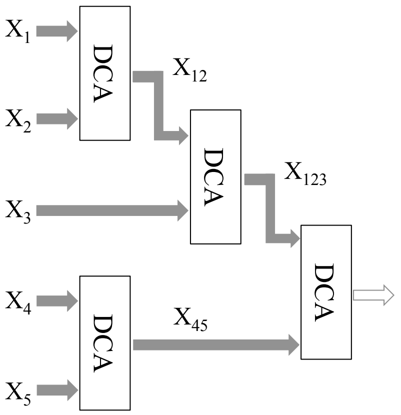


Fig. 2. Multiset discriminant correlation analysis techniques for five sample sets with $rank(X_1) > rank(X_2) > rank(X_3) > rank(X_4) = rank(X_5)$.

order to maintain the maximum possible length of the fused feature vector, in each step, the two feature sets with the highest ranks will be fused together. For example, in the first step, X_1 and X_2, which have the highest ranks, will be fused. The result of the fusion of X_1 and X_2 will be fused with the next highest rank feature set, *i.e.*, X_3, and so on. If there exists feature sets with equal ranks, they can be fused together at any time. We choose the length of the fused feature vector, r, to be equal to $min(c-1, rank(X_i), rank(X_j))$.

Fig. 2 shows an example framework of MDCA for five feature sets with $rank(X_1) > rank(X_2) > rank(X_3) > rank(X_4) = rank(X_5)$. In the first step of MDCA, we fuse X_1 and X_2, which have the highest ranks. X_4 and X_5, which have equal ranks, will be also fused together The length of the X_{12} is expected to be greater than the length of the X_{45}. Therefore, in the next step, X_3 is fused with X_{12}. In this way, we keep the maximum possible length for the fused feature vector in every step. The expected, possibly shorter, feature vector length can be determined in the final step, $r \leq min\left(c-1, rank\left(X_{123}\right), rank\left(X_{45}\right)\right)$.

IV. Experiments and Analysis

In this paper, we present several sets of experiments to demonstrate the performance of our proposed feature level fusion technique. We devise experiments for combining different features extracted from a single modality as well as combining feature vectors extracted from different biometric modalities. Section IV-A shows experimental results for combining different feature vectors extracted from a single modality. Additionally, Sections IV-B, IV-C, and IV-D present experiments on the fusion of different biometric modalities. In Section IV-B, experiments are performed on fusing features from frontal/near-frontal face, profile/near-profile face, and ear modalities extracted from West Virginia University (WVU) database [31]. Similarly, in Section IV-C, experiments are conducted on fingerprint and iris modalities from Multimodal Biometric Dataset Collection, BIOMDATA [32]. Section IV-D presents experiments on fusing information from weak biometric modalities, *i.e.*, periocular, mouth, and nose regions, extracted from face images in AR face database [33]. Section IV-E evaluates the scalability of the proposed DCA method in dealing with new subjects that are not seen during the training. Finally, as an example of the applicability of the proposed approach to other applications, Section IV-F shows how the proposed method helps to improve the accuracy of sketch to mugshot matching.

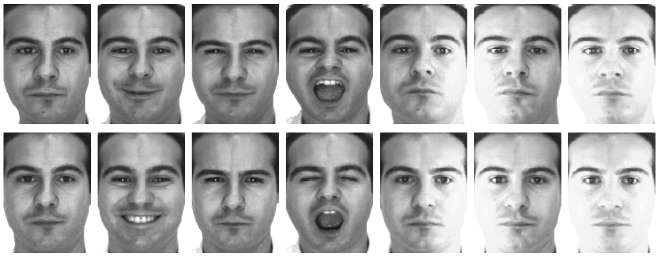


Fig. 3. Sample face images of a subject in AR database (Top row: first session; Bottom row: second session).

A. Unimodal Multi-Feature Fusion

In this section, we present experiments to show the effectiveness of the proposed method in combining feature sets extracted from a single modality. We evaluated our algorithms on a set of 100 subjects from AR face database [33], [34]. The AR face database consists of frontal face images with varying facial expressions and illumination. Fig. 3 shows sample images of one subject in the AR database. The face images are captured in two sessions. In this experiment, seven images of each subject from the first session are used for training and seven images from the second session are used for testing.

Three different features are extracted from these images. These features include Gabor wavelet features [35], Histogram of Oriented Gradients (HOG) [36], and Speeded-Up Robust Features (SURF) [37]. We employ forty Gabor filters in five scales and eight orientations. Since the adjacent pixels in an image are usually correlated, the information redundancy can be reduced by downsampling the feature images that result from Gabor filters [35], [38]. In our experiments, the feature images are downsampled by a factor of five. HOG features, on the other hand, are extracted in 5×5 cells for nine orientations. We use the UOCTTI variant for the HOG presented in [39]. UOCTTI variant computes both directed and undirected gradients as well as a four dimensional texture-energy feature, but projects the result down to 31 dimensions (27 dimensions corresponding to different orientation channels, 9 contrast insensitive and 18 contrast sensitive, and 4 dimensions capturing the overall gradient energy in square blocks of four adjacent cells)[3]. Finally, we extract SURF features from 68 keypoints in every image. These points are the facial landmarks detected by fitting an Active Appearance Model (AAM) to the face images. A 64-dimensional feature vector is extracted from each point and the final feature vector is constructed by concatenating the feature vectors of all keypoints. A simple minimum distance classifier is used for classification, in which one minus the sample linear correlation between observations is used as the distance.

[3] VLFeat open source library is used to extract the HOG features [40].

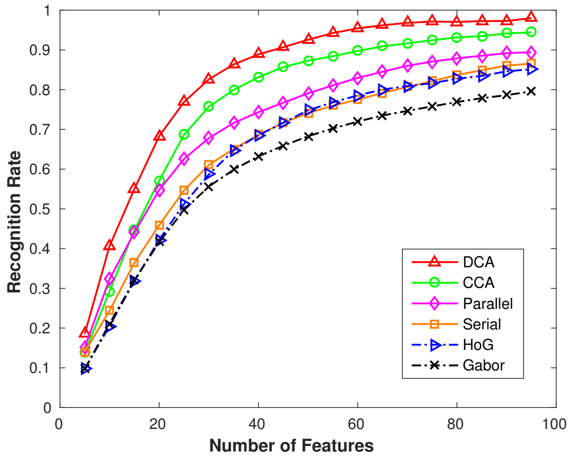


Fig. 4. Accuracy of the unimodal biometric systems using Gabor and HOG features on AR face database.

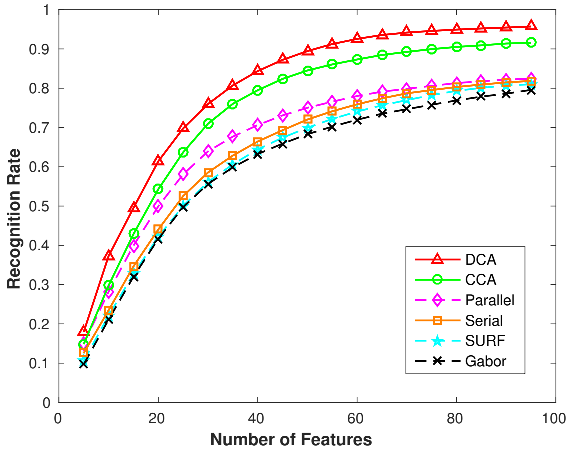


Fig. 5. Accuracy of the unimodal biometric systems using Gabor and SURF features on AR face database.

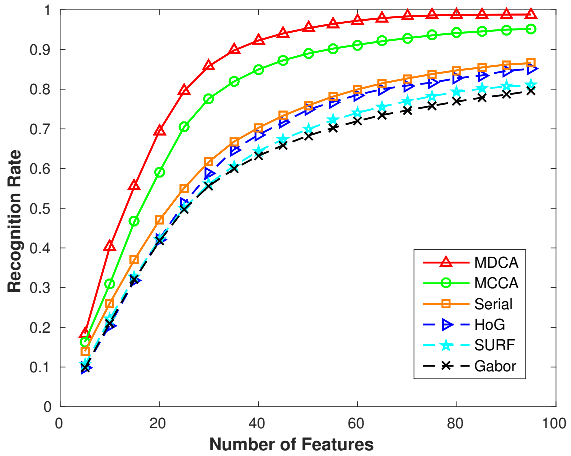


Fig. 6. Accuracy of the unimodal biometric systems using Gabor, SURF and HOG features on AR face database.

Figs. 4, 5, and 6 show the experimental results for combining different feature vectors. Table I shows the maximum rank-1 recognition rates, over the number of features, obtained using individual and fused feature vectors. As mentioned before, the goal is to combine relevant information from the two input feature vectors into a single vector, which is expected to be more discriminative than any of the input feature vectors.

TABLE I
MAXIMUM RANK-1 RECOGNITION RATES USING INDIVIDUAL AND FUSED FEATURE VECTORS (H: HOG, S: SURF, G: GABOR).

Method	H	S	G	HG	SG	HSG
Serial	85.14	81.00	79.57	86.57	81.86	86.57
Parallel	85.14	81.00	79.57	89.43	82.43	-
CCA/MCCA	85.14	81.00	79.57	94.43	91.57	95.14
DCA/MDCA	85.14	81.00	79.57	98.00	95.71	98.71

Therefore, a fusion method that decreases the correlations between features will be more effective.

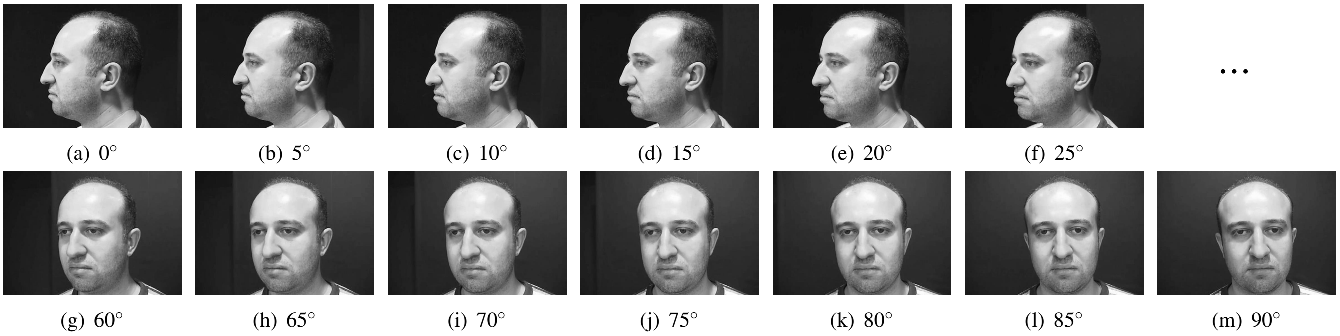
As it is clearly seen from the results, serial feature fusion [6] is not always successful in this regard, and in some cases, the fused feature vector has even less discriminative power than the input feature vector. Parallel feature fusion [7], [8] does not show a more discriminative feature either and in case of Gabor-SURF fusion, the fused feature vector works almost similar to the SURF feature vector. Note that the parallel feature fusion method cannot be applied on more than two sets of variables; therefore, it is excluded in the third experiment. For the cases of more than two feature sets, in this paper, we use Multiset Canonical Correlation Analysis (MCCA) [41] and MDCA methods.

The CCA-based feature fusion [9] and the proposed DCA feature fusion methods, on the other hand, work very well in combining different feature vectors. The reason might be the fact that these methods reduce the redundant information between two input feature vectors. Incorporating the class associations in its analysis, DCA provides a more powerful feature vector than CCA for the recognition purposes. The experimental results verify the effectiveness of our proposed method in comparison with serial, parallel and CCA-based feature fusion techniques. As mentioned in Section I, the JSRC [22] and SMDL [23] methods does not combine feature vectors extracted from multiple modalities into a single fused feature vector that can be used by any classifier. Therefore, these methods are not included in this experiment; however, they will be evaluated in the other experiments presented in Sections IV-B, IV-C, and IV-D.

B. *Multimodal Fusion: WVU Database*

1) Experimental Setup: In this set of experiments, we evaluate the performance of the proposed algorithm in combining feature vectors extracted from different biometric modalities on the WVU database [31]. This database consists of almost 110 seconds long video clips with rates of thirty frames per second, captured with a camera that rotates around the face. There are 402 subjects in the database. This database has 55 subjects with eyeglasses, 42 subjects with earrings, 38 subjects with partially occluded ears, and 2 subjects with fully occluded ears [42]. For subjects #239, #302, the ears are fully occluded with the hair, and for subject #308, just small portions of the ears are visible. Therefore, we exclude these three subjects and use the remaining 399 subjects in our experiments.

The video clips are captured by rotating a camera around the face; it starts from the left profile image of the face and ends at the right profile image. If we assume that the rotation for


(a) 0° (b) 5° (c) 10° (d) 15° (e) 20° (f) 25° ...

(g) 60° (h) 65° (i) 70° (j) 75° (k) 80° (l) 85° (m) 90°

Fig. 7. Different frames of the subject #1 from profile to frontal equally distanced by 5°.

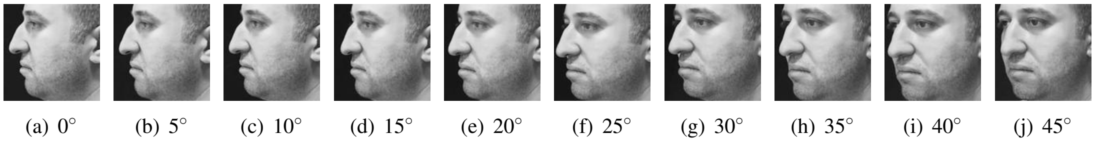

(a) 0° (b) 5° (c) 10° (d) 15° (e) 20° (f) 25° (g) 30° (h) 35° (i) 40° (j) 45°

Fig. 8. Profile/near-profile face images detected in different frames.

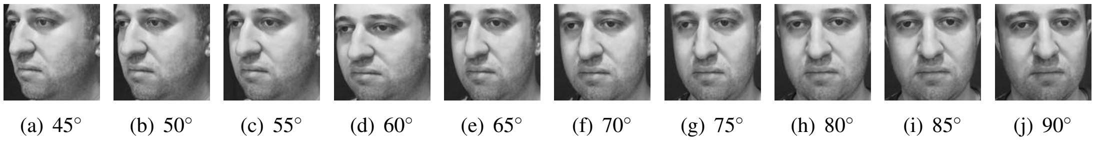

(a) 45° (b) 50° (c) 55° (d) 60° (e) 65° (f) 70° (g) 75° (h) 80° (i) 85° (j) 90°

Fig. 9. Frontal/near-frontal face images detected in different frames.

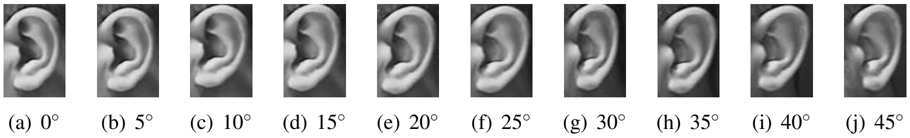

(a) 0° (b) 5° (c) 10° (d) 15° (e) 20° (f) 25° (g) 30° (h) 35° (i) 40° (j) 45°

Fig. 10. Ear images detected in different frames.

the left profile image is 0° and the rotation for the right profile image is 180°, the frontal image of the face is in the middle of the clip, *i.e.*, 90° of rotation. For our experiments, we choose frames that are five degrees of rotation apart. Figure 7 shows a sample of these frames in the range of 0° to 90°. We extract three different biometric modalities (frontal/near-frontal face, ear, and profile/near-profile face) from the above-mentioned frames. The best exposure of the profile face and the ear is at 0° while the best exposure of the frontal face is at 90°. For each modality, we choose ten images with up to 45° of rotation from their best exposure.

The face detection method proposed in [43] is used to automatically extract frontal and profile faces in each frame. For each subject, we extract ten profile and near-profile faces spanning between 0° and 45°, and ten frontal and near-frontal face images spanning from 45° to 90° degrees of rotation. Figures 8 and 9 show the sample profile/near-profile face and frontal/near-frontal face images extracted from the corresponding frames shown in Fig. 7.

On the other hand, the ear detection method proposed in [44] is used to automatically extract the ear regions. The ear detection method uses the deformable part model to find 17 landmarks on the ear helix and anti-helix. Figure 11(a) shows these landmarks on a sample ear image. We use the two green landmarks, the Triangular Fossa and Incisure Intertragica, to normalize the ear for in-plane pose variations. The normalized ear is shown in Fig. 11(b). For each subject, we extract ten ear images spanning between 0° and 45°. Figure 10 shows the sample ear images extracted from the corresponding frames shown in Fig. 7.

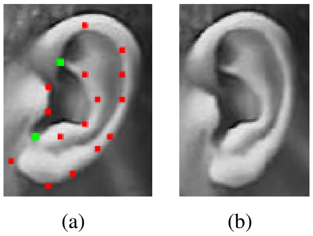

(a) (b)

Fig. 11. Ear normalization for in-plane rotations. (a) Detected Landmarks. (b) Normalized Ear.

In our experiments, all the face images are normalized to 120×120 pixels and all ear images are normalized to 120×80 pixels. For feature extraction, Gabor features are extracted in five scales and eight orientations, and similar to the setting described in Section IV-A, the feature images are downsampled by a factor of five. The most important advantage of Gabor filters is their invariance to rotation, scale, and translation. Furthermore, they are robust against photometric disturbances, such as illumination change and image noise [45], [46].

We perform three multimodal experiments using WVU database. These experiments include the fusion of (a) frontal face and ear, (b) profile face and ear, and (c) all three modalities. For the first experiment, ten face images of each subject are randomly paired with ten ear images of the same subject to create a multimodal dataset of face-ear pairs. Five randomly chosen pairs are used for training and the remaining

TABLE II
RANK-1 RECOGNITION RATES OBTAINED BY A KNN CLASSIFIER (K=1) USING INDIVIDUAL MODALITIES IN WVU DATABASE.

Modality	Face	Ear	Profile Face
Recognition Rate	82.59	79.66	81.71

five are used for testing. In order to validate the robustness of the experiments, repeated random sub-sampling validation is applied and the results are averaged over 10 iterations. The same setting is used for the second and third experiments using ear-profile pairs and face-ear-profile trios, respectively.

2) Comparison of Methods: The performance of the proposed feature level fusion algorithm is compared with that of several state-of-the-art feature level, matching score level and decision level fusion algorithms. The feature level fusion techniques include the serial feature fusion [6], the parallel feature fusion [8], the CCA-based feature fusion [9], [41], and the most recently published JSRC [22] and SMDL [23] methods. In order to prevent the small sample size problem in the CCA-based approach, dimensionality reductions based on Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) are applied [10], [29]. PCA and LDA are also used for dimensionality reduction and discriminant analysis of the results of the serial and parallel methods. Except for the JSRC and SMDL methods, which are restricted to work with a sparse representation classifier, all other feature level techniques use a simple KNN classifier with $K = 1$, *i.e.*, a minimum distance classifier, for classification. Here, one minus the sample linear correlation between observations is used as the distance. Note that in case of more than two modalities (Face+Ear+Profile), the parallel feature fusion method cannot be applied and Multiset-CCA [41] and Multiset-DCA are used.

For matching score level fusion and decision level fusion, we use Sparse Logistic Regression (SLR) [47] and SVM [48] techniques. For matching score level fusion, the probability outputs for each modality of the query samples are added together to produce the final score values, which are used for classification. For decision level fusion, on the other hand, the subject chosen by the maximum number of modalities was taken to be from the correct class. Following the notation in [22] and [23], we denote the score level fusion of these methods as SLR-Sum and SVM-Sum, and the decision level fusion as SLR-Major and SVM-Major. Moreover, we compare with the multiclass implementation of the Multiple Kernel Learning (MKL) algorithm [49].

Table II shows the rank-1 recognition rate for the individual modalities of face, ear and profile face, and Table III shows the multimodal fusion results. It is clear that the proposed DCA technique outperforms the other fusion methods. It is also shown that the combination of LDA + CCA is not effective for separating the classes due to the fact that the transformation applied by the CCA does not preserve the properties achieved by the LDA.

The complexity of the above-mentioned feature level fusion algorithms are compared using their run-time values. Table IV shows the average computation time for each algorithm. Note that the run-time values are for recognition of one multimodal face-ear pair in WVU database averaged over multiple runs. Note that the serial, parallel, CCA and DCA algorithms are very fast because they only apply the transformations obtained from the training process. Parallel feature fusion method is slightly more time consuming because it deals with complex feature vectors. JSRC and SMDL algorithms, on the other hand, are very time consuming and cannot be used in real-time applications.

TABLE III
RANK-1 RECOGNITION RATES FOR MULTIMODAL FUSION OF FACE, EAR AND PROFILE FACE BIOMETRICS IN WVU DATABASE.

Modality / Method	Face+Ear	Ear+Profile	Face+Ear +Profile
SVM-Major	85.09	85.31	87.59
SVM-Sum	94.18	94.42	95.12
SLR-Major	85.92	85.85	88.12
SLR-Sum	94.37	94.63	95.57
MKL	92.51	92.97	94.46
Serial + PCA + KNN	89.14	89.46	92.28
Serial + LDA + KNN	94.23	95.14	95.14
Parallel + PCA + KNN	90.71	90.61	-
Parallel + LDA + KNN	93.38	93.13	-
PCA + CCA/MCCA + KNN	94.10	94.34	97.74
LDA + CCA/MCCA + KNN	94.44	94.89	97.86
JSRC	96.20	97.74	98.74
SMDL	97.24	97.97	99.20
DCA/MDCA + KNN	98.56	99.38	99.85

TABLE IV
AVERAGE RUN-TIME VALUES OF DIFFERENT FEATURE LEVEL FUSION TECHNIQUES FOR RECOGNITION OF ONE MULTIMODAL FACE-EAR PAIR IN WVU DATABASE.

Method	Run Time (in milliseconds)
Serial + PCA + KNN	19
Serial + LDA + KNN	24
Parallel + PCA + KNN	39
Parallel + LDA + KNN	42
PCA + CCA + KNN	19
LDA + CCA + KNN	21
JSRC	8406
SMDL	7882
DCA + KNN	19

C. Multimodal Fusion: BIOMDATA Multimodal Biometric Dataset

In this set of experiments, we use the multimodal biometric dataset (BIOMDATA) collected in West Virginia University [32]. This dataset is a comprehensive collection of image and sound files for six biometric modalities: iris, face, voice, fingerprint, hand geometry, and palm print, from subjects of different ethnicity, gender, and age. It is a challenging data set, as many of the samples suffer from various artifacts such as blur, occlusion, shadows, and sensor noise, as shown in Fig. 12. Table V shows the number of subjects and samples available in each modality. Due to privacy issues related to identifying individuals, face data is not made available in combination with other modalities; therefore, it cannot be used in a multimodal experiment. Following the experimental setting in [22] and [23], we chose iris and fingerprint modalities for our

TABLE V
BIOMDATA MULTIMODAL BIOMETRIC DATASET.

Biometric Modality	# of Subjects	# of Samples
Iris	231	3043
Fingerprint	270	7136
Palm	263	673
Hand	219	2837
Voice	240	640
Face	205	1170

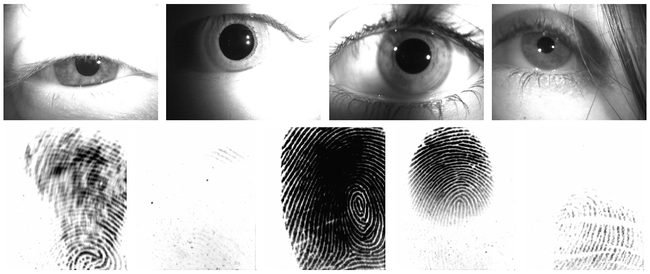
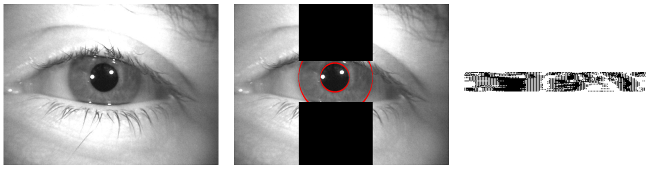
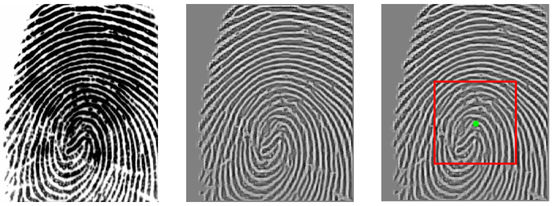


Fig. 12. Examples of challenging samples in BIOMDATA database. The images are corrupted with blur, occlusion, shadows, and sensor noise.



(a) (b) (c)

Fig. 13. Preprocessing for iris images. (a) Original iris image from BIOMDATA database. (b) Segmented iris area. (c) 25×240 binary iris template.



(a) (b) (c)

Fig. 14. Preprocessing for fingerprint images. (a) Original fingerprint image from BIOMDATA database. (b) Enhanced image using the method in [50]. (c) Core point of the fingerprint and the region of interest around it.

TABLE VI
RANK-1 RECOGNITION RATES OBTAINED BY A MINIMUM DISTANCE CLASSIFIER USING INDIVIDUAL MODALITIES ON 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 VII
RANK-1 RECOGNITION RATES FOR MULTIMODAL FUSION OF IRIS AND FINGERPRINT BIOMETRICS IN BIOMDATA DATABASE.

Modality Method	2 Irises	4 Fingerprints	All 6 Modalities
SVM-Major	62.30	90.14	92.24
SVM-Sum	71.03	93.43	97.51
SLR-Major	61.73	89.23	91.18
SLR-Sum	69.43	93.67	97.09
MKL	68.23	93.28	95.96
Serial + PCA + KNN	62.48	94.46	94.85
Serial + LDA + KNN	70.31	96.22	96.22
Parallel + PCA + KNN	68.22	-	-
Parallel + LDA + KNN	72.25	-	-
PCA + CCA/MCCA + KNN	78.51	96.32	97.20
LDA + CCA/MCCA + KNN	78.90	96.40	97.51
JSRC	78.20	97.60	98.60
SMDL	83.77	97.56	99.10
DCA/MDCA + KNN	84.16	98.71	99.60

experiments. All the evaluations are performed on a subset of 219 subjects having samples in both modalities. In total, there are two iris (left and right eye) and four fingerprint modalities (thumb and index fingers from both hands).

Fig. 13 shows the preprocessing steps for a sample iris image. We segmented the iris images using the method proposed in [51]. As shown in Fig. 13(b), the non-iris areas in the segmented region are removed as noise. Following the segmentation step, iris regions are 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 [52]. On the other hand, we enhanced the fingerprint images using the filtering methods described in [50]. Following the image enhancement step, the core points of the fingerprints are detected [53] and Gabor features in eight orientations are extracted around each detected core point. Fig. 14 shows the preprocessing steps for a sample fingerprint image.

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. As before, all experiments, except for the JSRC method, use a minimum distance classifier. One minus the sample linear correlation between observations is used as the distance.

Table VI shows the rank-1 recognition rate for the individual iris and fingerprint modalities, and Table VII shows the multimodal fusion results. We compare the proposed feature level fusion technique with several state-of-the-art feature level, matching score level and decision level fusion algorithms mentioned in Section IV-B2. Experimental results clearly show that the proposed DCA technique outperforms the other fusion methods.

D. Multimodal Fusion: AR Face Database

In this set of experiments, we show the applicability of the proposed MDCA 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. 15. It was shown that the periocular regions, nose and mouth can be considered as useful biometrics [54]–[56]; however, they are not as discriminative as the whole face [22].

We evaluated our algorithms on a set of 100 subjects from AR face database [33], [34] described in Section IV-A. Similar

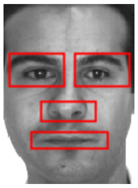


Fig. 15. Face mask used to crop out different modalities.

TABLE VIII

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

to the setup in [22], 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 VIII 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 performance with respect to that of the strong modality [57]. Table IX shows the recognition rates for different feature level fusion methods using combinations of different modalities. The results of fusing all five modalities with other methods including matching score level and decision level fusion techniques are presented in Table X. 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.

TABLE IX

RANK-1 RECOGNITION RATES FOR MULTIMODAL FUSION OF DIFFERENT MODALITIES IN AR DATABASE. MODALITIES INCLUDE 1. LEFT PERIOCULAR, 2. RIGHT PERIOCULAR, 3. NOSE, 4. MOUTH, AND 5. FACE.

Modality Method	{1,2}	{1,2,3}	{1,2,3,4}	{1,2,3,4,5}
Serial + PCA + KNN	85.57	88.71	90.42	90.71
Serial + LDA + KNN	89.43	92.14	92.86	93.57
PCA+CCA/MCCA+KNN	90.57	92.86	94.43	96.57
LDA+CCA/MCCA+KNN	91.28	92.57	93.71	97.00
JSRC	92.14	92.86	94.43	98.57
SMDL	92.29	92.86	95.14	98.85
DCA/MDCA + KNN	92.71	93.28	97.43	99.14

E. Scalability of DCA

In this section, we evaluate the scalability of the proposed method in dealing with new subjects that were not used for training. The goal is to examine if DCA is trained on a separate a population of subjects whether the transformation matrices will still perform well on new subjects. We use a population of subjects to train DCA and obtain the transformation matrices. Another population of subjects, which is not used for training, is used for evaluating the recognition performance.

For this purpose, we use the WVU database [31] with 399 subjects, introduced in Section IV-B. Similar to the experiment in Section IV-B, three different biometric modalities, *i.e.*, face (from frames between 45 and 90 degrees), ear and profile face (from frames between 0 and 45 degrees), are extracted from these frames. Each time we repeat the experiment, we randomly select a frame from the specified range for each modality for each subject to create the multimodal samples. A multimodal sample is a trio of a face, an ear, and a profile face image of a subject. Here, we have ten multimodal (face-ear-profile) samples per subject.

We divide the database into two populations with n_1 subjects for training the DCA and n_2 subjects for testing the performance. Five randomly selected multimodal samples from the first population, *i.e.*, training set, are used to calculate the transformation matrices of the DCA. The obtained transformation matrices are used to transform and fuse the feature sets of the second population, *i.e.*, testing set. We divide the second population into gallery and probe sets, which are used for the evaluation. Five randomly chosen multimodal (face-ear-profile) samples, for each subject, are used as gallery samples and the remaining five samples are used as probe.

Each time we repeat the experiment, we separate 99 randomly selected subjects from the database for the test population. Then, using the remaining subjects, we conduct three experiments with different number of training subjects in the first populations, $n_1 = 100, 200, 300$. In order to validate the robustness of the experiments, repeated random sub-sampling validation is applied and the results are averaged over 100 iterations. Fig. 16 shows the rank-1 recognition rate of the system with different number of training subjects n_1. Table XI shows the maximum recognition rate over the number of features in each case. The results show that the proposed algorithm is robust and it still performs well on new unseen subjects.

Since the maximum number of features is limited to $c-1$, c being the number of training subjects, the three diagrams shown in Fig. 16 have different domains. In case of $n_1 = 100$, we are only limited to 99 features and the maximum recognition rate achieved by these features is 99.32%. The other cases use more subjects for training; therefore, not only the training becomes more robust, but also the number of features increases, *i.e.*, 199 and 299. This helps achieve higher recognition accuracies, 99.89% and 99.98%. This phenomenon is clearly shown in the magnified part of Fig. 16.

F. Sketch to Mugshot Matching

In this section, we present an experiment that shows the applicability of DCA in improving the accuracy of a sketch to mugshot matching technique. Matching sketches to facial photographs is a challenging face recognition problem, which assists law enforcement to determine the identity of criminals [58]. Due to the large differences between sketches and photos and the unknown mechanism of sketch generation, it is difficult to match photos and sketches because they represent two different modalities. One way to solve this problem is to first transform a query sketch into a photo image and then

TABLE X
RANK-1 RECOGNITION RATES FOR MULTIMODAL FUSION OF ALL MODALITIES IN AR DATABASE.

SVM-Major	SVM-Sum	SLR-Major	SLR-Sum	MKL	Serial+LDA	LDA+MCCA	JSRC	SMDL	MDCA
85.71	92.85	86.85	93.71	93.00	93.57	97.00	98.57	98.85	99.14

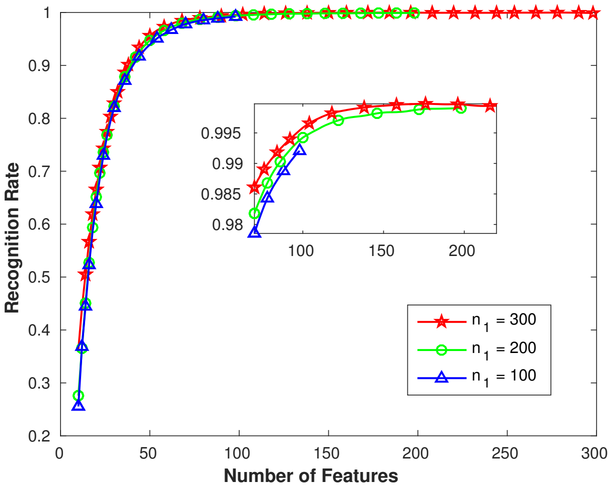


Fig. 16. Scalability of the proposed DCA algorithm using different number of training subjects and testing on unseen populations.

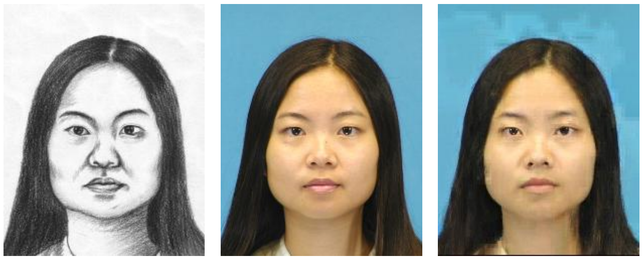
TABLE XI
MAXIMUM RANK-1 RECOGNITION RATES OVER THE NUMBER OF FEATURES IN FIG. 16.

n_1	100	200	300
Recognition Rate	99.32 ± 0.085	99.89 ± 0.051	99.98 ± 0.012

match the synthesized photo with real photos in the gallery [59].

In this experiment, we use the publicly available Chinese University of Hong Kong (CUHK) face photo-sketch dataset [59]. It includes 188 faces where for each face, there is a sketch drawn by an artist and a photo taken in frontal pose and neutral expression. In this database, 88 faces are preselected for training and the remaining 100 faces are used for testing. There is no identity overlap between the training and testing sets. Given a face sketch, we synthesize a pseudo-photo using a multiscale Markov Random Fields (MRF) model, which learns the face structure across different scales [59]. The MRF model is obtained using the training set of 88 photo-sketch pairs. Pseudo-photos are synthesized for the remaining 100 sketch images in the testing set of the CUHK database[4]. Fig. 17 shows a sample face photo-sketch pair and the synthesized pseudo-photo.

The projection matrices of DCA are obtained using the training set of 88 photo-sketch pairs. The remaining 100 real photos and the synthesized pseudo-photos are used as gallery and probe sets, respectively. Similar to the setting in Section IV-A, we extract Gabor and HOG features from these images and fuse them using DCA. A simple minimum distance classifier is used for recognition. Table XII shows the rank-1 recognition rate and compares the performance with that of [59] and the most recently published work [61]. The results show the advantages in fusing different features using DCA, as it significantly improves the sketch to photo matching accuracy.



(a) (b) (c)

Fig. 17. Photo synthesis result: (a) sketch drawn by the artist; (b) real photo; and (c) pseudo-photo synthesized from the sketch.

TABLE XII
RANK-1 RECOGNITION RATE FOR PHOTO-SKETCH MATCHING IN CUHK DATABASE.

Method	Ref. [59]	Ref. [61]	DCA
Recognition Rate	96.3	96	100

V. 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, uses the class associations of the samples in the 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. Moreover, DCA is computationally efficient and can be employed in real-time applications. Extensive experiments on various multimodal biometric databases demonstrated the efficacy of our proposed approach in the fusion of multimodal feature sets or different feature sets extracted from a single modality. In order to apply DCA for face recognition in unconstrained videos, more work needs to be performed to make sure that we obtain corresponding information from the different video clips. We will address this important problem in our future work.

REFERENCES

[1] A. Jain, K. Nandakumar, and A. Ross, "Score normalization in multimodal biometric systems," *Pattern recognition*, vol. 38, no. 12, pp. 2270–2285, 2005.

[2] A. Ross and A. Jain, "Information fusion in biometrics," *Pattern recognition letters*, vol. 24, no. 13, pp. 2115–2125, 2003.

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

[4] For synthesizing, we used the open-source code available from [60].

[4] M. M. Monwar and M. L. Gavrilova, "Multimodal biometric system using rank-level fusion approach," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, vol. 39, no. 4, pp. 867–878, 2009.

[5] X. Xu and Z. Mu, "Feature fusion method based on KCCA for ear and profile face based multimodal recognition," in *IEEE International Conference on Automation and Logistics (ICAL)*, 2007, pp. 620–623.

[6] 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.

[7] J. Yang and J.-y. Yang, "Generalized K–L transform based combined feature extraction," *Pattern Recognition*, vol. 35, no. 1, pp. 295–297, 2002.

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

[9] 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.

[10] N. M. Correa, T. Adali, Y.-O. 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.

[11] J. Yang and X. Zhang, "Feature-level fusion of fingerprint and finger-vein for personal identification," *Pattern Recognition Letters*, vol. 33, no. 5, pp. 623–628, 2012.

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

[13] W.-P. Li, J. Yang, and J.-P. Zhang, "Uncertain canonical correlation analysis for multi-view feature extraction from uncertain data streams," *Neurocomputing*, vol. 149, pp. 1337–1347, 2015.

[14] 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.

[15] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.

[16] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Transactions on Image Processing*, vol. 11, no. 3, pp. 293–305, 2002.

[17] Y.-H. Yuan, Q.-S. Sun, Q. Zhou, and D.-S. Xia, "A novel multiset integrated canonical correlation analysis framework and its application in feature fusion," *Pattern Recognition*, vol. 44, no. 5, pp. 1031–1040, 2011.

[18] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[19] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[20] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372–386, 2012.

[21] A. Ross and N. Poh, "Multibiometric systems: Overview, case studies, and open issues," in *Handbook of Remote Biometrics*. Springer, 2009, pp. 273–292.

[22] 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.

[23] S. Bahrampour, N. M. Nasrabadi, A. Ray, and W. K. Jenkins, "Multimodal task-driven dictionary learning for image classification," *IEEE Transactions on Image Processing*, vol. 25, no. 1, pp. 24–38, 2016.

[24] M. Haghighat, M. Abdel-Mottaleb, and W. Alhalabi, "Discriminant correlation analysis for feature level fusion with application to multimodal biometrics," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 1866–1870.

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

[26] 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.

[27] T.-K. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 1005–1018, 2007.

[28] Y. Ma, S. Lao, E. Takikawa, and M. Kawade, "Discriminant analysis in correlation similarity measure space," in *Proceedings of the 24th international conference on Machine learning (ICML)*. ACM, 2007, pp. 577–584.

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

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

[31] G. Fahmy, A. El-Sherbeeny, S. Mandala, M. Abdel-Mottaleb, and H. Ammar, "The effect of lighting direction/condition on the performance of face recognition algorithms," in *SPIE Conference on Biometrics for Human Identification*, 2006, pp. 188–200.

[32] 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.

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

[34] 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.

[35] 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.

[36] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2005, pp. 886–893.

[37] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008.

[38] 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.

[39] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2010.

[40] A. Vedaldi and B. Fulkerson, "VLFeat: An open and portable library of computer vision algorithms," http://www.vlfeat.org/, 2008.

[41] 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.

[42] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. S. Nixon, "A survey on ear biometrics," *ACM Computing Surveys (CSUR)*, vol. 45, no. 2, p. 22, 2013.

[43] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012, pp. 2879–2886.

[44] J. Lei, J. Zhou, and M. Abdel-Mottaleb, "Gender classification using automatically detected and aligned 3D ear range data," in *International Conference on Biometrics (ICB)*. IEEE, 2013, pp. 1–7.

[45] J.-K. Kamarainen, V. Kyrki, and H. Kalviainen, "Invariance properties of gabor filter-based features-overview and applications," *IEEE Transactions on Image Processing*, vol. 15, no. 5, pp. 1088–1099, 2006.

[46] 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.

[47] B. Krishnapuram, L. Carin, M. A. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 957–968, 2005.

[48] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.

[49] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

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

[51] 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.

[52] 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.

[53] 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.

[54] 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.

[55] A. Moorhouse, A. N. Evans, G. Atkinson, J. Sunf, and M. Smith, "The nose on your face may not be so plain: Using the nose as a biometric," pp. 1–6, 2009.

[56] 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.

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

[58] H. Han, B. F. Klare, K. Bonnen, and A. K. Jain, "Matching composite sketches to face photos: A component-based approach," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 191–204, 2013.

[59] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009.

[60] J. Xie, "MATLAB implementation of converting face to sketch and vice versa," https://github.com/ClaireXie/face2sketch, accessed: 2016-02-24.

[61] R. Srinivasan and A. Roy-Chowdhury, "Robust face recognition based on saliency maps of sigma sets," in *IEEE 7th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2015, pp. 1–6.

Mohammad Haghighat (S'10) received the B.Sc. and M.Sc. degrees in electrical engineering - communications from University of Tabriz, Tabriz, Iran, in 2008 and 2010, respectively. He is currently a Ph.D. candidate in the Department of Electrical and Computer Engineering at University of Miami. His research interests include image processing, computer vision, pattern recognition, cloud security, cryptography, and information theory.

Mohamed Abdel-Mottaleb (SM'03-F'11) received the Ph.D. degree in computer science from the University of Maryland, College Park, in 1993. He joined the University of Miami in 2001. Currently, he is a Professor and Chairman of the Department of Electrical and Computer Engineering. His research focuses on 3-D face and ear biometrics, dental biometrics, visual tracking, and human activity recognition. Prior to joining the University of Miami from 1993 to 2000, he was with Philips Research, Briarcliff Manor, NY, where he was a Principal Member of the Research Staff and a Project Leader. At Philips Research, he led several projects in image processing and content-based multimedia retrieval. He represented Philips in the standardization activity of ISO for MPEG-7, where some of his work was included in the standard. He holds 22 U.S. patents and more than 30 international patents. He published more than 120 journal and conference papers in the areas of image processing, computer vision, and content-based retrieval. He is an editorial board member for the Pattern Recognition journal. He is an IEEE fellow since January 2011.

Wadee Alhalabi received his Ph.D. degree in electrical and computer engineering from the University of Miami, Coral Gables, FL, in 2008. He joined the department of Computer Science at King Abdulaziz University in 2010 as an Assistant Professor. Also, in 2010 he became an active researcher at Effat University. Dr. Alhalabis research interest includes image processing and the application of virtual reality in the medical field. He published more than 40 journal and conference articles.