Effective Feature Fusion for Pattern Classification Based on Intra-Class and Extra-Class Discriminative Correlation Analysis

Wu Zuobin, Mao Kezhi
School of Electrical and Electronics Engineering
Nanyang Technological University
Singapore 639798
wu0004in@e.ntu.edu.sg

Gee-Wah Ng
Cognition and Fusion Laboratory
DSO National Laboratories
Singapore

Abstract—Information fusion aims to exploit truthful knowledge from various sources in a reliable and accurate way. Fusion of information can be conducted at three abstraction levels including feature level, score level and decision level. The feature fusion approaches have the advantages of preserving effective discriminative structure underlying various features. In this paper, we propose an effective feature fusion algorithm based on intra-class and extra-class discriminative correlation analysis (IEDCA), aiming to eliminate between-class correlation and retain enough feature dimension for correlation analysis. IEDCA explores the intra-class correlation including both the pairs-wise correlation like CCA-based feature fusion approaches and the correlation across different features within the same class. Our proposed method can be used in unimodal feature fusion as well as multimodal feature fusion, and extensive experiments have proved its effectiveness.

Keywords—Feature fusion; canonical correlation analysis; pattern classification

I. INTRODUCTION

Information fusion is the integration of information from various sources, aiming to achieve better performance than they were used as disparate sources [1]. The fusion of information has been used in various tasks such as opinion fusion in risk assessment [2], multisensory image fusion in remote sensing [3, 4], sensor fusion in robotics and autonomous system [5, 6], merging of different database [7, 8], target tracking and classification [9], logic-based fusion [10, 11] etc. Fusion of information can be conducted at three abstraction levels: feature level, score level, and decision level. For the feature fusion, the features are fused and input to the classifier as a whole. In score fusion, one matching score is generated to represent the proximity of each feature vector to the corresponding template vector [12]. The scores are then combined to make the final decision. In decision fusion, techniques like majority voting [13] are used to combine the decisions obtained from various classifiers.

Compared with the score fusion and decision fusion, the feature fusion approach has several advantages. First, different features reflect different characteristics of patterns, and feature-level fusion effectively preserves the discriminative structure underlying these different features [14]. Second, feature fusion utilizes the correlation of multiple features at an early stage to enhance task accomplishment, and learning phase is conducted only once on the fused feature vector [15].

The classical feature fusion approaches include serial fusion [16] and parallel fusion [17]. Serial fusion concatenates two or more feature sets into a union vector. Take two feature sets X and Y as an example. If feature set X has p dimensions and Y has q dimensions, serial fusion will form a (p + q)-dimensional combined feature vector. For parallel fusion, a complex vector would be introduced for combining the two sets of features rather than a union vector. If dimensions of the two feature sets are not equal, zero padding is applied to the lower dimensional features.

Feature fusion by Canonical Correlation Analysis (CCA) [18] is frequently adopted in pattern recognition tasks where heterogeneous sources are employed. CCA-based feature fusion aims to find the linear correlation of two features by maximizing the correlation criterion. Multiset canonical correlation analysis is proposed as an extension of CCA in order to deal with multiple sets of features [19, 20]. However, such extension [19, 20] cannot infer the presence of latent correlations when the sample size is less than the feature dimension. Nicholas [21] proposed an informative multiset canonical correlation analysis (IMCCA) to solve this problem. IMCCA forms the block-structured matrix using a smaller subset of singular vectors. Since multiset canonical correlation analysis is essentially an unsupervised learning method, discriminant information may be overlooked when features are projected into canonical subspaces.

By taking class information into consideration, some supervised CCA-based approaches [22-27] have been proposed to enhance the discriminating power of learned features. Multiset Discriminant Correlation Analysis (MDCA) [27] learns discriminating correlation features through the maximization of pairwise correlations across different feature sets and the elimination of between-class correlations. Multiview Supervised CCA (MSCCA) [26] considers the class information of the training samples from both within-view and between-view.
With the development of mathematics theory, sparse representation has gained success in pattern recognition. Sparse representation assumes that there exists a linear combination of training samples to sufficiently represent a query sample. In sparse representation, the coefficients of the linear representation are computed, followed by the calculation of the reconstruction residuals. The query sample will be classified according to the minimum reconstruction residual. Joint sparse representation (JSR) [28] is regarded as a multimodal feature fusion approach, where multiple dictionaries are created from training samples of each individual modality. The objective of JSR is to find joint sparse representation coefficients that have the sparsity patterns in common. The query sample is then reconstructed by the multimodal training samples belonging to the same class. Multimodal task-driven dictionary learning [29] is developed to enforce the collaboration of multiple sources of information.

In this paper, we propose a feature fusion scheme called Intra-Class and Extra-Class Discriminative Correlation Analysis (IEDCA), with the goal of eliminating between-class correlation and retaining enough dimensions of feature for correlation analysis. In each feature set, class separability is enhanced by the discriminative analysis through the intra-class and extra-class nearest neighbor of each sample in the feature set. IEDCA also explores the correlation across different features within the same class. The discriminating structure is inherited by incorporating class matrix. The proposed method overcomes the problem of small sample size encountered in CCA and is extended to deal with multiset features by sequential forward selection in IEDCA.

The rest of the paper is organized as follows: In Section II, we review CCA-based feature fusion and linear discriminant analysis (LDA). In Section III, we present our proposed intra-class and extra-class discriminative correlation analysis. Section IV provides the experiments and discussion. The paper is concluded in Section V.

II. RELATED WORK

A. Feature Fusion by Canonical Correlation Analysis

Suppose there are two feature sets from the same n objects denoted by \( X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{p \times n} \) and \( Y = [y_1, y_2, \ldots, y_n] \in \mathbb{R}^{q \times n} \), respectively, where \( x_i \) (or \( y_i \)) can be treated as a sample. CCA is used to find optimal pairs of correlation projection matrices \( W_x \) and \( W_y \) so that \( W_x^T X \) and \( W_y^T Y \) have maximum correlation with each other. The following correlation function is maximized to find \( W_x \) and \( W_y \):

\[
\text{corr}(U, V) = \frac{\text{cov}(U, V)}{\text{var}(U) \cdot \text{var}(V)}
\]

where \( U = W_x^T X \) , \( V = W_y^T Y \) , \( \text{cov}(U, V) = W_x^T C_{xy} W_y \) , \( \text{var}(U) = W_x^T C_{xx} W_x \) and \( \text{var}(V) = W_y^T C_{yy} W_y \). \( C_{xy} \in \mathbb{R}^{p \times q} \) is the cross-set covariance matrix, while \( C_{xx} \in \mathbb{R}^{p \times p} \) and \( C_{yy} \in \mathbb{R}^{q \times q} \) are the within-set covariance matrices. Maximization of above-mentioned criterion can be an optimization problem of maximizing \( \text{cov}(U, V) \) with constraints \( \text{var}(U) = \text{var}(V) = 1 \). \( W_x \) and \( W_y \) are obtained through solving the following generalized eigenvalue problem [30]:

\[
\begin{align*}
C_{xy}^{-1} C_{xx} W_x &= \lambda^2 C_{xx} W_x \\
C_{yy}^{-1} C_{xy} W_y &= \lambda^2 C_{yy} W_y
\end{align*}
\]

The correlation projection matrices \( W_x \in \mathbb{R}^{p \times d} \) and \( W_y \in \mathbb{R}^{q \times d} \) correspond to top \( d \) eigenvalues. The projected correlation features can be extracted in the form of \( W_x^T X \) and \( W_y^T Y \), and concatenated into a fused feature vector.

Feature fusion by CCA encounters certain limitations. If the dimension of features exceeds the number of samples \( (p > n \) or \( q > n \) ), covariance matrices \( C_{xx} \) and \( C_{yy} \) become singular and non-invertible. This is often referred to small sample problem. To overcome this problem, one solution is to reduce the feature dimension using principal component analysis (PCA) before performing CCA. However, the negligence of class label information in PCA may deteriorate the classification performance. In this sense, LDA may be a better choice.

B. Linear Discriminant Analysis

Linear discriminant analysis is a classical supervised technique for dimension reduction, and continues to be a popular tool. For a problem of \( c \) classes, LDA projects the data into a \((c-1)\)-dimension space through maximizing between-class scatter matrix and minimizing within-class scatter matrix. The between-class and within-class scatter matrices are given as follows:

\[
S_b = \sum_{i=1}^{c} n_i (m_i - m) (m_i - m)^T
\]

\[
S_w = \sum_{i=1}^{c} \sum_{j=1}^{n_i} (x_{ij} - m_i) (x_{ij} - m_i)^T
\]

where \( m_i \) is the mean vector of class \( i \), and \( m = \frac{1}{n} \sum_{i=1}^{n} x_i \) is the mean vector of the whole training set. Through maximizing the trace ratio of between-class scatter matrix and within-class scatter matrix, the projection matrix is obtained as:

\[
W = \arg \max_w \frac{\text{tr}(W^T S_b W)}{\text{tr}(W^T S_w W)}
\]

where \( \text{tr}(\cdot) \) represents the trace of a matrix. Instead of solving (5) directly, the trace differential solution is usually performed as [31].
Intra-Class and Extra-Class Discriminative Analysis

The basic idea of Intra-Class and Extra-Class Discriminative Correlation Analysis (IEDCA) is illustrated in Fig. 1. IEDCA attempts to find transforms where classes are separated within each feature set while the discriminating structure is inherited in the correlation analysis. Assume the samples are from \( c \) separate classes. Let \( x_i^j \) denotes feature vector of \( i^{th} \) sample in the \( j^{th} \) class. \( m_e(x_i^j) \) denotes the extra-class k-nearest neighbor (KNN) mean. \( w(i, l) \) denotes the weight corresponding to \( x_i^j \). The between-class scatter matrix is thus defined as:

\[
W = \arg \max_W tr(W^T(S_b - S_w)W)
\]

The transformation matrix \( W \) corresponds to top \( d \) eigenvalues.

III. INTRA-CLASS AND EXTRA-CLASS DISCRIMINATIVE CORRELATION ANALYSIS

A. Motivation

Linear Discriminant Analysis (LDA) can enhance class separability by using class label information. However, the discriminating structure from LDA may not be inherited by CCA since it is unsupervised in nature. The utility of CCA in fusing heterogeneous data for classification is limited since it cannot guarantee that learned correlated features are discriminative. Supervised CCA methods [26, 27] have been proposed to preserve the discriminating structure and can be applied together with LDA. But LDA reduces the original feature dimension to at most \( (c-1) \) discriminant features. This will lose some information required by the subsequent correlation analysis and eventually degrade the performance. Therefore, we propose to seek transformations that can retain enough dimension of features for class separation under each set of features, as well as inherit the discriminating structure in the learned correlation features.

B. Intra-Class and Extra-Class Discriminative Analysis

The basic idea of Intra-Class and Extra-Class Discriminative Correlation Analysis (IEDCA) is illustrated in Fig. 1. IEDCA attempts to find transforms where classes are separated within each feature set while the discriminating structure is inherited in the correlation analysis. Assume the samples are from \( c \) separate classes. Let \( x_i^j \) denotes feature vector of \( i^{th} \) sample in the \( j^{th} \) class. \( m_e(x_i^j) \) denotes the extra-class k-nearest neighbor (KNN) mean. \( w(i, l) \) denotes the weight corresponding to \( x_i^j \). The between-class scatter matrix is thus defined as:

\[
W = \arg \max_W tr(W^T(S_b - S_w)W)
\]

The transformation matrix \( W \) corresponds to top \( d \) eigenvalues.

\[
S_b^n = \sum_{i=1}^{c} \sum_{i=1}^{n_i} w(i, l) (x_i^j - m_e(x_i^j))(x_i^j - m_e(x_i^j))^T
\]

where the extra-class KNN mean \( m_e(x_i^j) \) is defined as:

\[
m_e(x_i^j) = \frac{1}{k} \sum_{p=1}^{k} \vec{N}(x_i^j)_{extra}
\]

\( \vec{N}(x_i^j)_{extra} \) is the \( p \)-th nearest neighbor from extra-class to \( x_i^j \), and weight \( w(i, l) \) is defined as:

\[
w(i, l) = \min\{d(x_i^j, \vec{N}(x_i^j)_{intra}), d(x_i^j, \vec{N}(x_i^j)_{extra})\} \cdot \frac{d(x_i^j, \vec{N}(x_i^j)_{intra})}{d(x_i^j, \vec{N}(x_i^j)_{intra}) + d(x_i^j, \vec{N}(x_i^j)_{extra})}
\]

\( d(v_1, v_2) \) denotes the Euclidean distance between vector \( v_1 \) and \( v_2 \). \( \vec{N}(x_i^j)_{intra} \) is the \( k \)-th nearest neighbor from intra-class (class \( i \)) to \( x_i^j \). The weighting function emphasize the boundary information which is important for pattern classification. The between-class scatter matrix, \( S_b^n \) can be diagonalized as follows:

\[
Q^{-1}S_b^n Q = \Lambda
\]

where \( Q \) denotes the right eigenvectors and \( \Lambda \) denotes the diagonal matrix of descending eigenvalues. By choosing top \( r \) eigenvalues and their corresponding eigenvectors:

\[
Q^{-1}_{(r \times p)} S_b^n Q_{(p \times r)} = \Lambda_{(r \times r)}
\]
Instead of reducing the dimension directly to $c-1$, reasonable dimension of feature is kept before correlation analysis:

$$X'_{r 	imes n} = Q^T_{(r 	imes p)}X_{(p 	imes n)}$$ (12)

Similarly, $Y$ can be transformed to $Y'$. In this way, more information of classification structure is preserved.

**C. Intra-Class Correlation Analysis**

Now $X$ and $Y$ are transformed to $X'$ and $Y'$, respectively. Traditional correlation analysis only considers one-to-one correlation. The class structure which contains discriminative information is neglected. Intra-class correlation explores not only one-to-one correlation, but also the correlation across different features within the same class. To utilize intra-class correlation, we introduce a class matrix $L \in R^{n \times c}$ ($c$ denotes the number of the classes, while $n$ denotes the number of training samples). Each row of the class matrix $L$ represents the sample label. For instance, if there are three classes ($c=3$), $[1 \ 0 \ 0]$, $[0 \ 1 \ 0]$ and $[0 \ 0 \ 1]$ denote class 1, class 2 and class 3, respectively. The covariance matrix between the transformed feature sets with class correlation is:

$$C_L = (X'H)(Y'HL)^T$$ (13)

where the centering matrix $H \in R^{n \times n}$ is defined as $H = I - \frac{1}{n}ee^T$, and $e$ is a vector of all ones. To obtain the transformation that maximizes both the pair-wise correlation between different features and the correlation across different features within each class, and meanwhile preserves the class structure in the learned correlation features, we need to maximize the following criterion:

$$corr(U,V) = \frac{cov(UL, VL)}{\text{var}(U) \cdot \text{var}(V)}$$ (14)

where $U = W_x^T X'H$, $V = W_y^T Y'H$, $cov(UL, VL) = W_x^T C_L W_y$. It can be easily verified that $\text{var}(U) = W_x^T C_{xx} W_x$ and $\text{var}(V) = W_y^T C_{yy} W_y$.

$$\text{var}(U) = UU^T = \begin{pmatrix} W_x^T X'H(T) W_x^T X'H(T) \end{pmatrix}^T = W_x^T (X'H(T)X'H(T))^T W_x = W_x^T \left[ (I - \frac{1}{n}ee^T) \right]^T W_x = W_x^T (X' - \mu_x)(X' - \mu_x)^T W_x = W_x^T C_{xx} W_x$$ (15)

Maximization of (14) can be performed by Lagrange multiplier on the covariance between $UL$ and $VL$ subject to $\text{var}(U) = \text{var}(V) = 1$ .The transformation matrices, $W_x$ and $W_y$, are obtained through solving the following generalized eigenvalue equation:

$$\begin{pmatrix} C_L C_{yy}^{-1} C_L^T W_x = \lambda^2 C_{xx} W_x \\ C_L^T C_{xx}^{-1} C_L W_y = \lambda^2 C_{yy} W_y \end{pmatrix}$$ (16)

Since $C_{xx}$ and $C_{yy}$ are symmetric positive definite, we are able to decompose them as:

$$C_{xx} = R_{xx} R_{xx}^T$$ (17)

where $R_{xx}$ is a lower triangular matrix. Let $A_x = R_{xx}^T W_x$, the first part in (16) becomes:

$$(R_{xx})^{-1} C_L C_{yy}^{-1} C_L^T (R_{xx}^T)^{-1} A_x = \lambda^2 A_x$$ (18)

Which is a symmetric standard eigenproblem. $W_x$ can be obtained as $(R_{xx}^T)^{-1} A_x$, $W_y$ can be obtained as $C_{yy}^{-1} C_L^T W_y$. Since the number of nonzero generalized eigenvalues is at most $c-1$, the upper limit of $W_x$ and $W_y$ is $\min(c-1, \text{rank}(X'), \text{rank}(Y'))$. The whole transformation matrices are $W_x^T Q_x^T$ and $W_y^T Q_y^T$, respectively. Eventually, the transformed features are concatenated together into a fused feature vector.

The process of IEDCA is summarized below.

**IEDCA Algorithm.**

**Input:** training data matrix, $X$ and $Y$, label matrix $L$.

**Output:** Transformation matrix $W_x^T Q_x^T$ and $W_y^T Q_y^T$.

1: Construct between-class scatter matrix through the nearest neighbor from both intra-class and extra-class.
2: Diagonalize the between-class scatter matrix $Q \leftarrow$ eigenvectors of (11) corresponding to top $r$ eigenvalues.
3: $X' \leftarrow Q_x^T X$. Similarly, $Y$ is transformed into $Y'$
4: $H \leftarrow I - \frac{1}{n}ee^T$.
5: $C_L \leftarrow (X'HL)(Y'HL)^T$
6: Compute basis: $W_x, W_y \leftarrow$ eigenvectors of (16) corresponding to maximum $c-1$ nonzero generalized eigenvalues when seeking the maximization of correlation $UL$ and $VL$ subject to $\text{var}(U) = \text{var}(V) = 1$ as in (14). The entire transformation matrices are $W_x^T Q_x^T$ and $W_y^T Q_y^T$, respectively.
7: Concatenate transformed features

**D. Multiset Intra-Class and Extra-Class Discriminative Correlation Analysis**

In this section, the proposed intra-class and extra-class discriminative correlation analysis is extended to multiple sets of features. Inspired by sequential forward selection method [32], we develop the following procedure for multiset IEDCA:

1) First, all sets of features are ranked in the decreasing order based on their individual classification accuracy. The first set of feature which has the highest classification accuracy is selected among all the sets.
2) Second, the next set of feature is selected among all the unselected sets features which leads to the highest
classification accuracy when fusing with the selected set in the previous step.

3) Third, the fused feature is considered as a new set of features. Repeat the above process until enough sets of features are selected, or until the classification accuracy is good enough.

IV. EXPERIMENTS AND DISCUSSIONS

To test the effectiveness of our proposed model, experiments are performed on D-case dataset [33], SDUMLA-HMT database [34] and AR database [35] respectively. Experiments are designed to fuse a variety of features from single modality as well as multiple modalities. Part A demonstrates the experiment of unimodal feature fusion, while Part B shows the experiment of multimodal feature fusion. Part C evaluates the effect of different dimensions kept in IEDCA.

A. Unimodal Feature Fusion

In this experiment, the effectiveness of IEDCA is demonstrated in the case of fusing various features from single modality using D-case dataset [33]. The D-case dataset is provided by IEEE AASP challenge involving acoustic scene classification and event detection. Here, we use the scene classification dataset to evaluate our proposed model. The scene classification dataset consists of ten different scenes: tube, subway station, supermarket, restaurant, quiet street, park, open air market, office, busy street and bus. For each scene, ten audio clips are provided and each of them is 30 seconds long. All the audio clips are sampled at 44.1 kHz.

In our experiment, three types of feature extraction methods were employed. For each audio clips, the first type of features, Mel-frequency cepstral coefficients (MFCC) were extracted. Frame window is set to 25 ms, and step between successive windows is 10 ms. Frequency range is limited to 0-900 Hz. The periodogram is calculated for each frame. The Mel filter bank is applied, followed by taking the DCT of the calculated energies.

Recurrence Quantification Analysis (RQA) is the second type of features extracted. The basic idea of RQA is to quantify the patterns that emerge in the recurrence plots by certain metrics. The details can be found in [36]. RQA feature is a combination of the following metrics:

- Recurrence (REC), which represents the percentage of points in the threshold plot. It will be high for self-similar sound when the radius in RQA is fixed.
- Determinism, which denotes the proportion of points in the diagonal lines of RQA. It is used to identify periodic sounds.
- Laminarity (LAM), which represents the proportion of points in forming the vertical lines. It is used to identify stationary segments of sounds. Maximum length (Vmax) is used to characterize durations of stationary periods.
- The average diagonal length (LEN) and longest diagonal size (Lmax), which are used to characterize periodic repetitions.
- Trapping time, which represents the average vertical line length, while Entropy is the Shannon entropy of diagonal line length.

Texture windows of 40 MFCC frames representing 400ms of audio were used in computing the recurrence plots. In this way, the temporal evolution of the features, which is averaged to obtain a document level representation, is captured. A cosine similarity of 0.03 between MFCC frames is adopted.

Discrete Wavelet Transform (DWT) is the third type of features employed in this experiment. DWT provides frequency and time resolution which closely matches human ear. A six-level wavelet decomposition using the Daubechies4 (D4) wavelet was used. For each level, both the mean and the variance of the wavelet coefficients over each scene segment were computed and used as features.

We performed five-fold cross-validation for ten runs and averaged the recognition results. Several state-of-the-art feature fusion algorithms are compared with our proposed feature fusion algorithm: serial fusion, parallel fusion, conventional CCA, MDCA, and Multiview Supervised CCA (MSCCA). In the CCA-based technique, dimension reduction such as PCA and LDA are applied to overcome the small sample size problem. SVM is chosen as the classifier. When there are more than two feature sets, the method of parallel fusion cannot be applied and the conventional CCA is replaced by multiset-CCA.

Table I shows the classification accuracy of the individual type of features, while Table II shows the classification accuracy of feature fusion. As demonstrated in the results, serial fusion does not always work. In the case of DWT, the serial-fused feature is less discriminative than the individual feature. The parallel-fused feature is also less discriminative than the individual feature in the cases of MFCC, RQA, and DWT, and it can only be applied on the fusion of two feature sets. Most CCA-based feature fusion results are good since they maximize the correlation between different features. Supervised CCA-based feature fusion techniques perform even better due to the incorporation of class information into correlation analysis. Our proposed IEDCA not only inherits the class information in the correlation analysis but also retains enough dimensions of feature for correlation analysis.

<table>
<thead>
<tr>
<th>Feature Type</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>MFCC (Channel 1)</td>
<td>65.02</td>
</tr>
<tr>
<td>MFCC (Channel 2)</td>
<td>64.66</td>
</tr>
<tr>
<td>RQA (Channel 1)</td>
<td>55.38</td>
</tr>
<tr>
<td>RQA (Channel 2)</td>
<td>54.97</td>
</tr>
<tr>
<td>DWT (Channel 1)</td>
<td>64.84</td>
</tr>
<tr>
<td>DWT (Channel 2)</td>
<td>56.23</td>
</tr>
</tbody>
</table>
### TABLE II. CLASSIFICATION ACCURACIES FOR FEATURE FUSION OF MFCC, RQA AND DWT IN D-CASE DATABASE

<table>
<thead>
<tr>
<th>Method</th>
<th>Feature Type</th>
<th>MFCC (Channel 1&amp;2)</th>
<th>RQA (Channel 1&amp;2)</th>
<th>DWT (Channel 1&amp;2)</th>
<th>MFCC + RQA (Channel 1&amp;2)</th>
<th>MFCC + DWT (Channel 1&amp;2)</th>
<th>RQA + DWT (Channel 1&amp;2)</th>
<th>All Six Feature Sets</th>
</tr>
</thead>
<tbody>
<tr>
<td>Serial + PCA + SVM</td>
<td></td>
<td>65.74</td>
<td>55.82</td>
<td>62.61</td>
<td>65.49</td>
<td>65.91</td>
<td>60.48</td>
<td>61.94</td>
</tr>
<tr>
<td>Serial + LDA + SVM</td>
<td></td>
<td>66.54</td>
<td>56.50</td>
<td>63.36</td>
<td>66.17</td>
<td>66.72</td>
<td>62.39</td>
<td>63.51</td>
</tr>
<tr>
<td>Parallel + PCA + SVM</td>
<td></td>
<td>63.67</td>
<td>54.22</td>
<td>62.05</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Parallel + LDA + SVM</td>
<td></td>
<td>65.01</td>
<td>55.76</td>
<td>62.51</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>PCA + CCA/MCCA + SVM</td>
<td></td>
<td>67.13</td>
<td>56.68</td>
<td>63.47</td>
<td>67.28</td>
<td>67.03</td>
<td>63.58</td>
<td>65.28</td>
</tr>
<tr>
<td>LDA + CCA/MCCA + SVM</td>
<td></td>
<td>67.49</td>
<td>57.42</td>
<td>64.79</td>
<td>67.91</td>
<td>67.91</td>
<td>64.14</td>
<td>69.36</td>
</tr>
<tr>
<td>MSCCA + SVM</td>
<td></td>
<td>67.51</td>
<td>58.18</td>
<td>65.18</td>
<td>68.77</td>
<td>68.53</td>
<td>66.07</td>
<td>70.76</td>
</tr>
<tr>
<td>MDCA + SVM</td>
<td></td>
<td>67.72</td>
<td>60.51</td>
<td>64.91</td>
<td>72.95</td>
<td>68.91</td>
<td>66.75</td>
<td>74.95</td>
</tr>
<tr>
<td>Proposed Method + SVM</td>
<td></td>
<td>67.83</td>
<td>61.28</td>
<td>65.32</td>
<td>74.56</td>
<td>69.63</td>
<td>67.89</td>
<td>76.72</td>
</tr>
</tbody>
</table>

Therefore, IEDCA provides a more discriminative feature fusion scheme. Experiment results show that our proposed IEDCA outperforms serial, parallel, CCA-based and supervised CCA-based feature fusion methods.

### B. Multimodal Feature Fusion: SDUMLA-HMT Database

In this experiment, SDUMLA-HMT database is used. This database is a comprehensive collection of multiple modalities: finger vein, face, iris, gait and fingerprint from 106 individuals. Experimental settings follow those in [28] and [29]. Irises of both eyes and fingerprints from thumb and index of both hands are used.

Segmentation was performed on iris images using the technique in [37]. Valid part of the iris is illustrated in Fig. 2(b). The segmented area is normalized and a binary iris template [38] is generated from the segmented iris as shown in Fig. 2(c). Log-Gabor features are extracted from each binary iris template.

For fingerprints, they were enhanced following the method in [39]. After enhancement, the core point was detected [40] and convolution was performed on the enhanced fingerprints using Gabor filters which have eight orientations. Around the detected core point, circular tessellations consisting of 15 concentric bands were extracted. The entire preprocessing is shown in Fig. 3.

---

![Fig. 2. Iris preprocessing. (a) Sample image in SDUMLA-HMT database. (b) Valid region of iris after segmentation. (c) Generated binary iris template.](image)

---

![Fig. 3. Fingerprint preprocessing. (a) Sample fingerprint in SDUMLA-HMT database. (b) Enhanced fingerprint. (c) Detected core point and region of interest.](image)

---

We performed five-fold cross-validation for ten runs and averaged the recognition results. Our proposed feature fusion algorithm is compared with other feature fusion algorithms as in the previous experiment. Besides, we also compare with Joint Sparse Representation (JSR) and SMDL [29] as mentioned in Section I, which are used in combination with a sparse representation classifier. JSR and SMDL are not included in experiment A since they are used for multimodal fusion.

Table III shows the classification accuracy for each individual modality, while Table IV provides the classification accuracy for multimodal fusion. Experimental results demonstrate our IEDCA outperforms other fusion methods.

### C. Effect of Dimensions Retained for Correlation Analysis

In this section, the effect of different dimensions retained for correlation analysis is evaluated on AR database [35]. This database contains images from 100 subjects under various illuminations and facial expressions. The subset of AR database without occlusion is chosen. In the experiment, seven images from each subject are randomly chosen for training, while the remaining are used for testing. Ten independent recognition tests are conducted.
TABLE III.
CLASSIFICATION ACCURACIES BY SVM FOR INDIVIDUAL MODALITIES OF SDUMLA-HMT DATABASE

<table>
<thead>
<tr>
<th>Modality</th>
<th>Classification Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Left Iris</td>
<td>61.05</td>
</tr>
<tr>
<td>Right Iris</td>
<td>55.79</td>
</tr>
<tr>
<td>Left Thumb Fingerprint</td>
<td>76.84</td>
</tr>
<tr>
<td>Left Index Fingerprint</td>
<td>88.53</td>
</tr>
<tr>
<td>Right Thumb Fingerprint</td>
<td>81.05</td>
</tr>
<tr>
<td>Right Index Fingerprint</td>
<td>89.58</td>
</tr>
</tbody>
</table>

Gabor features and Histogram of Oriented Gradients (HOG) are extracted. Convolution is performed on each image with a set of 40 Gabor filters with eight orientations and five scales, followed by a down-sampling factor of five. HOG features are extracted with nine orientations in $5 \times 5$ cells. UOCTTI variant [39] is adopted in calculating directed and undirected gradients of HOG. The result is projected down to 31 dimensions, which consist of 27-dimension orientation channels and 4-dimension gradient energy. Since we only explore the number of dimensions kept in IEDCA, a simple nearest neighbor classifier would be enough. Table V shows the average classification accuracies for various numbers of dimensions kept in the discriminative analysis, obtained by the individual feature and the fused feature.

As mentioned in Section II, to solve small sample problem, the dimension of feature usually reduces to at most $c-1$ ($c$ is the number of classes), followed by correlation analysis. In this experiment, $c=100$, the feature dimension is reduced to 99 and the average classification accuracies achieved by Gabor features and HOG features are 88.57% and 77.86%, respectively. Average classification accuracy achieved by IEDCA fusion of Gabor and HOG features is 96.29%. In this case, the dimension of feature is reduced to $c-1$, and certain information which is important for the correlation analysis between different features is lost. This will, in turn, affect the eventual performance.

In IEDCA, dimension reduction of features for correlation analysis is not restricted to $c-1$. In Table V, with the increase of dimensions kept in discrimination analysis (from 99 to 190), the average classification accuracy achieved by Gabor features drops from 88.57% to 84.14%. This is because part of the information is redundant when we increase dimensions kept in discrimination analysis. Meanwhile, the average classification accuracy achieved by HOG stays at 77.86%. The increase of dimensions kept for HOG has no effect in the recognition. Average classification accuracy achieved by IEDCA fusion of Gabor features and HOG features improves from 96.29% to 98.29% with the increase of dimensions kept in discrimination analysis. This is because part of redundant information underlying individual features contains correlation information which is important for feature fusion.

V. CONCLUSIONS

In this paper, we propose a feature fusion scheme named intra-class and extra-class discriminative correlation analysis (IEDCA). The proposed IEDCA inherits the discriminating structure in the learned correlation features by incorporating the class information and fully explores the intra-class correlation including both one-to-one pair-wise correlation and the correlation between different features within the same class. Moreover, more information is kept in the discriminative analysis to enhance the performance of learned correlation features. Various experiments ranging from the fusion of different features of single modality to multimodal feature fusion demonstrate the effectiveness of our proposed approach.

ACKNOWLEDGMENT

Supports from MINDEF-NTU-JPP/15/02/01 are acknowledged.

REFERENCES
