Selecting Discriminative Features with Discriminative Multiple Canonical Correlation Analysis for Multi-Feature Information Fusion

Lei Gao1, Lin Qi1, Ling Guan1,2
1 School of Information Engineering Zhengzhou University
450001, Zhengzhou, China
email- iegaolei@gmail.com; ielqi@zzu.edu.cn
2 Electrical and Computer Engineering Ryerson University
M5B 2K3, Toronto, Canada
email- lguan@ee.ryerson.ca

Abstract: In this paper, it presents a novel approach for selecting discriminative features in multimodal information fusion based discriminative multiple canonical correlation analysis (DMCCA), which is the generalized form of canonical correlation analysis (CCA), multiple canonical correlation analysis (MCCA) and discriminative canonical correlation analysis (DCCA). The proposed approach identifies the discriminative features from the multi-feature in Fractional Fourier Transform (FRFT) domain, which are capable of simultaneously maximizing the within-class correlation and minimizing the between-class correlation, leading to better utilization of the multi-feature information and producing more effective pattern recognition results. The effectiveness of the introduced solution is demonstrated through extensive experimentation on a visual based emotion recognition problem.

1 Introduction

The effective utilization and integration of multiple information contents presented in different media sources are becoming an increasingly important research topic in many applications with the proliferation of multimedia and the advances in sensing technology. Since single information based pattern analysis and recognition systems only afford low level of performance due to the drastic variation and noisy nature of the acquired signals, it leads to insufficient and inaccurate pattern representation of the perception of interest. However, multiple data contains more information about the semantics presented in the media. The combination of multiple data may potentially provide a more complete and discriminatory description of the intrinsic characteristics of the pattern, and produce improved system performance compared with single information only [LYRYAM-2010].

Therefore, information fusion becomes an emerging and challenging research area in multimedia processing. The major difficulties lie in the identification of the inherent relationship between different information, and the design of a fusion strategy that can effectively utilize the complementary information presented in different channels. A wide variety of methods have been proposed in the literature to address the difficulties.

In general, there are three levels of information fusion: feature/data level, score level and decision level [AA-2003]. Compared with other methods, the advantage of the feature level fusion is as follows. As different feature vectors extracted from the same pattern tend to reflect different characteristics of the pattern, optimally combining these features not only keeps the effective discriminant information of multi-feature, but also eliminates the redundant information to certain degree, which is especially important to classification and recognition of large scale database in high dimensional feature space. It is also the focus of this paper.

Generally, there exist two traditional classes of feature fusion methods which are serial feature fusion and parallel feature fusion [JJDJ-2003]. Recently, there has been extensive interest in the analysis of correlation based approaches for multi-feature information
fusion such as canonical correlation analysis (CCA) [XK-1999], kernel canonical correlation analysis (KCCA) [CP-2001], discriminative CCA (DCCA) [TSZJ-2007] and multi-set canonical correlation analysis (MCCA) [A-2002], which have been applied to audiovisual based talking-face biometric verification [HG-2007], medical imaging analysis [NTYV-2010], handwriting recognition [QSYPD-2005], audio-visual synchronization [MYEA-2007], joint blind source separation [TWV-2009], blind single-input and multiple-output (SIMO) channels equalization [JID-2005]. However, as CCA, KCCA or DCCA method only could deal with the mutual relationships between two random vectors and it lacks discriminative character for MCCA. In order to address the mentioned problems, the approach of DMCCA is introduced in [LLEL-2012], which extracts more discriminative characteristics of multi-feature information. Nevertheless, one important yet not well studied problem in DMCCA is that there is not any reliable approach to select the discriminative features from multi-feature information to achieve better recognition results.

In this paper, we conduct a novel approach for selecting discriminative features in multimodal information fusion based discriminative multiple canonical correlation analysis (DMCCA). The proposed approach identifies the more discriminative features in FRFT domain on a visual based emotion recognition problem. The effectiveness of the introduced solution is demonstrated through extensive experimentation. The remainder of this paper is organized as follows: the conception of FRFT and 2D-FRFT is briefly introduced in Section 2. In Section 3, the analysis and derivation of the proposed approach is presented. The emotion recognition system and experimental results are given in Section 4. Conclusions are drawn in Section 5.

2 FRFT and 2D-FRFT

FRFT is a generalized form of the FT, which can be interpreted as a rotation of the signal in the time-frequency plane [L-1994]. It contains simultaneity the time-frequency information of the signal, and is considered as a new tool for time-frequency analysis, especially in the area of image representation [HD-1993, A-1993]. In the area of facial expression recognition, some researches have shown its superiority with respect to other feature extraction tools [LLEXL-2010]. As different order features of 2D-FRFT contain different time-frequency information [CMH-2000], and the previous studies only focus on the single order FRFT features. Thus, it is reasonable to fuse the different features to improve the recognition results.

In this section, we introduce the definition and properties of the FRFT and the two-dimensional form 2D-FRFT.

Given a signal , its FRFT is defined as:

\[
X_u(t) = \left[F^\alpha[x(t)](u) \right] = \int_{-\infty}^{\infty} x(t) K_{\alpha}(t,u) \, dt
\]

(1)

\[
K_{\alpha}(t,u) = \begin{cases}
\delta(t-u), & \alpha=\pm\pi \\
\delta(t+u), & \alpha=\pm(2\pi) \\
\end{cases}
\]

(2)

where \(\alpha = (p\pi)/2\) is the rotation angle in FRFT domain and \(p\) is the transform order.

The FRFT can be extended to its two-dimensional form. For a two-dimensional signal \(x(s,t)\), its 2D-FRFT is defined as:
Given a signal, its FRFT is defined as:

\[ X_{\alpha}(u,v) = F_{\alpha}^{-1}[x(t,s)] \]  

(3)

where \( \alpha = (p^*\pi)/2, \beta = (q^*\pi)/2 \) are the rotation angles, and \( p, q \) are the transform orders in the 2D-FRFT. It satisfies the property of additivity in 2D-FRFT [SH-1998]:

\[ X_{\alpha\beta}(u,v) = F_{\alpha\beta}^{-1}\left[ X_{\alpha}(u,v) \right] \]  

(4)

In the fields of digital image processing, the two dimensional discrete FRFT can be implemented by row-column computation as shown in [SH-1998].

### 3 The Approach of Selecting Discriminative Feature for DMCCA

Let \( X_1, X_2, \ldots, X_Q \) be \( Q \) sets-zero-mean random samples as:

\[ X_i = \left[ x_{i1}^{(1)}, x_{i2}^{(1)}, \ldots, x_{im}^{(1)}, \ldots, x_{i1}^{(c)}, x_{i2}^{(c)}, \ldots, x_{in}^{(c)} \right]^{\top} \in \mathbb{R}^{n \times N} \]  

(5)

where \( i \) is the number sequence of the random samples, \( x_{ij}^{(m)} \) denotes the \( j \)th sample in the \( m \)th class, respectively, and \( n_i \) is the number of samples in the \( i \)th class of \( X_i \) set.

\[ \sum_{i=1}^{Q} n_i = N \]  

(6)

where \( c \) is the total number of classes.

The aim of DMCCA is to seek the projection vectors \( \omega = [\omega_1, \omega_2, \ldots, \omega_N] \) for feature extraction so that the within-class correlation is maximized and the between-class correlation is minimized. With the definition of DMCCA in [LLEL-2012], it can be written as:

\[ \frac{1}{N-1}(C - D)\omega = \rho D\omega \]  

(7)

where

\[
C = \begin{bmatrix}
X_1X_1^\top & \cdots & X_1X_N^\top \\
\vdots & \ddots & \vdots \\
X_NX_1^\top & \cdots & X_NX_N^\top
\end{bmatrix}
\]

\[
D = \begin{bmatrix}
X_1X_1^\top & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & X_NX_N^\top
\end{bmatrix}
\]

\[
A = \begin{bmatrix}
I_{n_1 \times n_1} & \cdots & 0 \\
\vdots & \ddots & \vdots \\
0 & \cdots & I_{n_N \times n_N}
\end{bmatrix}
\]

\[ R^{n \times n} \]

with \( I = [1,1,\ldots,1]^{\top} \in \mathbb{R}^n \) and \( \omega = [\omega_1^{\top}, \omega_2^{\top}, \ldots, \omega_N^{\top}]^{\top} \).

Therefore, the solution obtained is the eigenvector associated to the eigenvalue of equation (7). That is:

\[ \frac{1}{N-1} \text{inv}(D)^*(C - D)\omega = \rho \omega \]  

(8)

where \( \text{inv} \) means inverse transform of matrix. However, unless the covariance matrices \( D \) have full rank, the block matrix in Eq. (8) will become singular. An approach [TMH-2003] to dealing with singular covariance matrices and to controlling complexity is to add a multiple of the identity matrix \( \lambda I \), \( \lambda > 0 \) to \( D \).
Thus, the general form of equation (8) is written:

$$\frac{1}{N-1} \text{inv}(D^+)*(C - D)\omega = \rho \omega$$  \hspace{1cm} (9)

where

$$D^+ = \begin{cases} D & \text{when } D \text{ is insingular matrix} \\ D + \lambda I & \text{when } D \text{ is singular matrix} \end{cases}$$

From equation (7), the $\rho$ is the criterion to seek the projection vectors for feature extraction. That is to say, the value of $\rho$ is the key parameter to the effect of selecting discriminative features, and the larger $\rho$ corresponds to the more discriminative features, which effectively maximize the within-class correlation and minimize the between-class correlation. While the smaller $\rho$ corresponds to the less discriminative features. Thus, it is reasonable to evaluate the final information fusion results by judgment $J(\rho)$:

$$J(\rho) = \sum_{i=1}^{N} \rho_i \hspace{1cm} (10)$$

where $\rho_i$ is the $i$th eigenvalue of equation (9).

Furthermore, since

$$\text{trance}\left(\frac{1}{N-1} \text{inv}(D^+)*(C - D)\right) = \sum_{i=1}^{N} \rho_i \hspace{1cm} (11)$$

where $\text{trance}(\ )$ represents the trace of the matrix. Thus, we can calculate $J(\rho)$ through trance of the matrix instead of the eigenvalue decomposition when the number of features is large, which could significantly speed up the calculation process.

4 Emotion Recognition System and Experimental Results Analysis

4.1 Emotion recognition system

In this paper, we demonstrate the effectiveness of the proposed method on a visual based emotion recognition problem with FRFT features. Figure 1 depicts a general block diagram of the proposed emotion recognition system.

For facial expression representation, a key image which corresponds to the frame with largest speech amplitude is first identified from the audiovisual. Then we detect the face region from the image frame using a color based method [YL-2008]. Then the face
region is normalized to an image of size of $112 \times 96$. As the large dimensionality of the coefficients, we downsample each subband to a size of $32 \times 32$ to enhance the recognition process. The extracted FRFT features in each of the time segments are then analyzed and selected using the proposed analysis approach. Subsequently, the newly generated features, which represent the multi-set information among different patterns, are concatenated into projected vectors for classification through the method of DMCCA. Then, the nearest-neighbor classifier is used for emotion recognition with visual features.

4.2 Experimental results analysis

To evaluate the performance of the proposed method, we conduct experiments on the RML audiovisual emotion database [YL-2008], which consists of video samples from eight human subjects, speaking six different languages (English, Mandarin, Urdu, Punjabi, Persian, and Italian). In the experiment, a total of 288 samples are selected from the RML audiovisual database, each belonging to one of the six universal emotional states: anger(AN), disgust(DI), fear(FE), sadness(SA), surprise(SU) and happiness(HA). Among the samples, 192 samples are chosen for training set and 96 are chosen for evaluation. As a benchmark, the performances of using single FRFT features in different transform orders (from 0.1 to 1.0) are first evaluated as Figure 2. From Figure 2, it is observed that recognition results are dissimilar in different transform orders due to different time-frequency information, which could be applied in the field of information fusion with expecting to reach higher recognition accuracy. In the experiment, we select the five single features ($\rho=0.3, 0.7, 0.8, 0.9, 1.0$) with better recognition accuracy from Figure 2 as the fused multi-feature. Based on equation (14) (15), the values of $J(\rho)$ fused with different 2D-FRFT features in DMCCA method are shown as Table 1, which mean the discrimination of the fused features.

![Figure 2: Results of emotion recognition with single FRFT feature](image)

Table 1. Results of $J(\rho)$ with different number of 2D-FRFT features

<table>
<thead>
<tr>
<th>FRFT Features</th>
<th>0.8, 0.9</th>
<th>0.8, 0.9, 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>$J(\rho)$</td>
<td>-0.24</td>
<td>0.12</td>
</tr>
<tr>
<td>FRFT Features</td>
<td>0.7, 0.8, 0.9, 1.0</td>
<td>0.3, 0.7, 0.8, 0.9, 1.0</td>
</tr>
<tr>
<td>$J(\rho)$</td>
<td>8.93</td>
<td>0.04</td>
</tr>
</tbody>
</table>

From Table 1, it can be observed that the value of $J(\rho)$ with four 2D-FRFT features($0.7, 0.8, 0.9, 1.0$) obtains the maximum 8.93, while the result with two 2D-FRFT features($0.8, 0.9$) arrives the minimum -0.24. Based on our analysis, as $J(\rho)$ means the discrimination of the fused features, the recognition result fused with four 2D-FRFT features($0.7, 0.8, 0.9, 1.0$) should outperform others. In order to demonstrate the effectiveness of the proposed method on 2D-FRFT features domain and our
mathematical analysis, the experiments on visual emotion recognition using the method of DMCCA performed with the different number of 2D-FRFT features are shown as follows. In the experiments, it is worthwhile to point out that, as there are 32*32 dimensions in the visual recognition problem, we adopt the first 50 projected dimensions by the approach of DMCCA. The overall recognition rates are shown in Figure 3. Furthermore, we also perform the highest recognition accuracy under different number of features, which can be shown as Table 2. Besides, in order to demonstrate the efficiency of the proposed method effectively, the relation between $J(\rho)$ and highest recognition results is shown as Figure 4. It can be seen that the experimental results comply with our mathematical analysis presented in the previous section.

![Figure 3 Emotion recognition experimental results of different number FRFT features](image)

**Table 2** Highest emotion recognition results of different FRFT features

<table>
<thead>
<tr>
<th>FRFT Features</th>
<th>0.8, 0.9</th>
<th>0.8, 0.9, 1.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Highest results(%)</td>
<td>83.33</td>
<td>85.42</td>
</tr>
<tr>
<td>FRFT Features</td>
<td>0.7, 0.8, 0.9, 1.0</td>
<td>0.3, 0.7, 0.8, 0.9, 1.0</td>
</tr>
<tr>
<td>Highest results(%)</td>
<td>90.67</td>
<td>84.38</td>
</tr>
</tbody>
</table>

![Figure 4 Relation between highest recognition results and values of $J(\rho)$](image)

Furthermore, in order to demonstrate the effectiveness of the proposed method in 2D-FRFT features domain, we also implemented feature fusion with the state-of-the-art methods such as CCA, MCCA and DCAA with different 2D-FRFT features for the purpose of comparison. The recognition rates with different projected dimensions are shown in Figure 5 and the highest recognition results with mentioned methods are shown as Figure 6. Based on Figure 5 and Figure 6, it can be seen that the highest recognition rates were achieved with the proposed method on selecting discriminative features in 2D-FRFT features with DMCCA.
mathematical analysis, the experiments on visual emotion recognition using the method of DMCCA performed with the different number of 2D-FRFT features are shown as follows. In the experiments, it is worthwhile to point out that, as there are 32*32 dimensions in the visual recognition problem, we adopt the first 50 projected dimensions by the approach of DMCCA. The overall recognition rates are shown in Figure 3. Furthermore, we also perform the highest recognition accuracy under different number of features, which can be shown as Table 2. Besides, in order to demonstrate the efficiency of the proposed method effectively, the relation between $J(\rho_i)$ and highest recognition results is shown as Figure 4. It can be seen that the experimental results comply with our mathematical analysis presented in the previous section.

![Figure 5 The overall emotion recognition experimental results with different feature fusion](image1)

![Figure 6 The highest emotion recognition experimental results with different feature fusion](image2)

5 Conclusion

This paper has introduced a new approach for effectively selecting discriminative features in multimodal information fusion based the method of DMCCA. The proposed solution identifies the discriminative features from multiple features for information fusion. The effective combination of multi-feature data potentially provides a more complete and discriminatory description of the intrinsic characteristics of the pattern, and produce improved system performance. Experimental results on a visual based emotion recognition problem in FRFT domain demonstrate the proposed method selects more discriminative features effectively, improves the recognition performance, and is computationally efficient. Although we focus on an emotion recognition problem in FRFT domain in this paper, the proposed solution can also be applied to other multimodal or multi-feature related multimedia analysis problems. Another interesting and challenging topic will be investigating kernelized version of the DMCCA based on the Kernel canonical correlation analysis (KCCA) in order to obtain a more effective method to solve nonlinear fusion problems which are more prevalent in information fusion.

6 Acknowledgement

This work is supported by the National Natural Science Foundation of China, No.61071211 and the Canada Research Chair Program.
References


