Incremental Multi-linear Discriminant Analysis Using Canonical Correlations for Action Recognition

Cheng-Cheng Jia[a], Su-Jing Wang[a], Xu-Jun Peng[b], Can-Yan Zhang[c], Chun-Guang Zhou[a], Zhe-Zhou Yu[a,*]

[a] *College of Computer Science and Technology, Jilin University, Changchun 130012, China*
[b] *Raytheon BBN technologies, Boston, MA, 02138, USA*
[c] *College of Computer Science and Technology, Harbin Engineering University, Harbin 150000, China*

Abstract

Canonical correlations analysis (CCA) is often used for feature extraction and dimensionality reduction. However, the image vectorization in CCA breaks the spatial structure of the original image, and the excessive dimension of vector often brings the curse of dimensionality problem. In this paper, we propose a novel feature extraction method based on CCA in multi-linear discriminant subspace by encoding an action sample as a high-order tensor. An optimization approach is presented to iteratively learn the discriminant subspace by unfolding the tensor along different tensor modes. It retains most of the underlying data structure including the spatio-temporal information, and alleviates the curse of dimensionality problem. At the same time, an incremental scheme is developed for multi-linear subspace online learning, which can improve the discriminative capability efficiently and effectively. The nearest neighbor classifier (NNC) is exploited for action classification. Experiments on Weizmann database showed that the proposed method outperforms the state-of-the-art methods in terms of accuracy. The proposed method has low time complexity and is robust against partial occlusion.

Keywords: Canonical Correlations Analysis, Multi-linear Subspace, Discriminant Information, Incremental Learning, Action Recognition

*Corresponding author
Email address: `yuzz@jlu.edu.cn` (Zhe-Zhou Yu)

1. Introduction

Nowadays, many feature extraction methods have been used in recognition related task, such as action recognition [1] [2] [3] and face recognition [4] [5]. Most traditional algorithms, such as principal component analysis (PCA) [6] [7] and linear discriminant analysis (LDA) [8] [9], represent an object as a 1-dimension vector. Canonical correlations analysis (CCA) [10] [11], which reflects the degree of similarity of two image sets in orthogonal subspaces, has received more increasing attention for recognition recently. Kim *et al.* [12] proposed an optimal discriminant function of canonical correlations (DCC) to transform image sets, so that the similarity of intra-class sets is maximized while the similarity of inter-class sets is minimized. Wu *et al.* [13] proposed an incremental learning scheme to update the discriminant matrix for the analysis of canonical correlations (IDCC), which does not require a complete re-training when training samples are available incrementally, resulting in reduced computational cost. However, the image vectorization of these methods has broken the original spatial structure and often leads to the curse of dimensionality problem.

For the sake of overcoming this limitation, a number of multi-linear subspace analysis (MSA) methods [14] [15] [16] have been suggested for recognition. The discriminant analysis with tensor representation (DATER) proposed in [17] captures most of the discriminatory information by maximizing a tensor-based scatter ratio criterion. The incremental tensor biased discriminant analysis (ITBDA) [18] is suitable for distinguishing and tracking the objects by online learning the tensor biased discriminant subspace. However, most of the MSA methods work directly on a single sample, without considering the canonical correlations between different samples.

In this paper, we propose a novel CCA-based feature extraction method, called multi-linear discriminant analysis of canonical correlations (MDCC), to iteratively learn the multi-linear discriminant subspace using canonical correlations between different samples. We develop an online learning scheme for the MDCC named as incremental multi-linear discriminant-analysis of canonical correlations (IMDCC). The added samples incrementally update the discriminant information, which can maximize the canonical correlations of the intra-class samples while minimizing the canonical correlations of the inter-class samples. We summarize the advantages of our algorithm IMDCC as follows.

1. IMDCC operates on each mode of the training tensors separately to

alleviate the curse of dimensionality problem.

2. The optimization algorithm IMDCC converges as discussed illustratively in this paper.

3. IMDCC shows the high computational efficiency of tensor subspace learning.

The rest of the paper is organized as follows. In Section 2, we introduce the tensor algebra and DCC algorithm. In Section 3, we present the MDCC and IMDCC algorithms and discuss the convergence performance of IMDCC. In Section 4, we compare the experimental results and the computational cost with other methods. Conclusions are drawn in Section 5.

2. Related works

2.1. Multi-linear algebra

A tensor is a multi-dimensional array. In this paper, scalers are denoted by lowercase letters, e.g., a. Vectors (*1-order* tensor) are denoted by bold lowercase letters, e.g., $\mathbf{a}$. Matrices (*2-order* tensor) are denoted by bold uppercase letters, e.g., $\mathbf{A}$. Higher-order tensors (*3-order* or higher) are denoted by calligraphic uppercase letters, e.g., $\mathcal{A}$.

An *N-order* tensor is represented as $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_n \times \ldots \times I_N}$, where I_n is the dimensions of *mode-n*, $(1 \leqslant n \leqslant N)$. An element of $\mathcal{A}$ is denoted as $\mathcal{A}_{i_1 i_2 \ldots i_n \ldots i_N}$, $(1 \leqslant i_n \leqslant I_n)$. The *mode-n* vectors of $\mathcal{A}$ are the vectors in $\mathbb{R}^{I_n}$ while keeping the vectors of other modes fixed. For example, a matrix is taken as a *2-order* tensor, the column vectors in the matrix are the *mode-1* vectors and the row vectors in the matrix are the *mode-2* vectors.

Slices of $\mathcal{A}$ are two-dimensional sections of a tensor. The left-hand side of Figure 1 shows the frontal, lateral, and horizontal slices of a *3-order* tensor $\mathcal{A}$, respectively.

Definition 1. (**Mode unfolding**). The *mode-n* unfolded matrix of $\mathcal{A}$, denoted by $\mathbf{A}_{(n)} \in \mathbb{R}^{I_n \times (I_1 I_2 \ldots I_{n-1} I_{n+1} \ldots I_N)}$ is obtained by spreading the slices side by side along a given direction. The column vectors of $\mathbf{A}_{(n)}$ are the *mode-n* vectors of $\mathcal{A}$ while keeping the vectors of other modes fixed.

Taking a *3-order* tensor $\mathcal{A}$ for example, the *mode-n* unfolded matrix $\mathbf{A}_{(n)} (1 \leqslant n \leqslant 3)$ is shown in Figure 1.

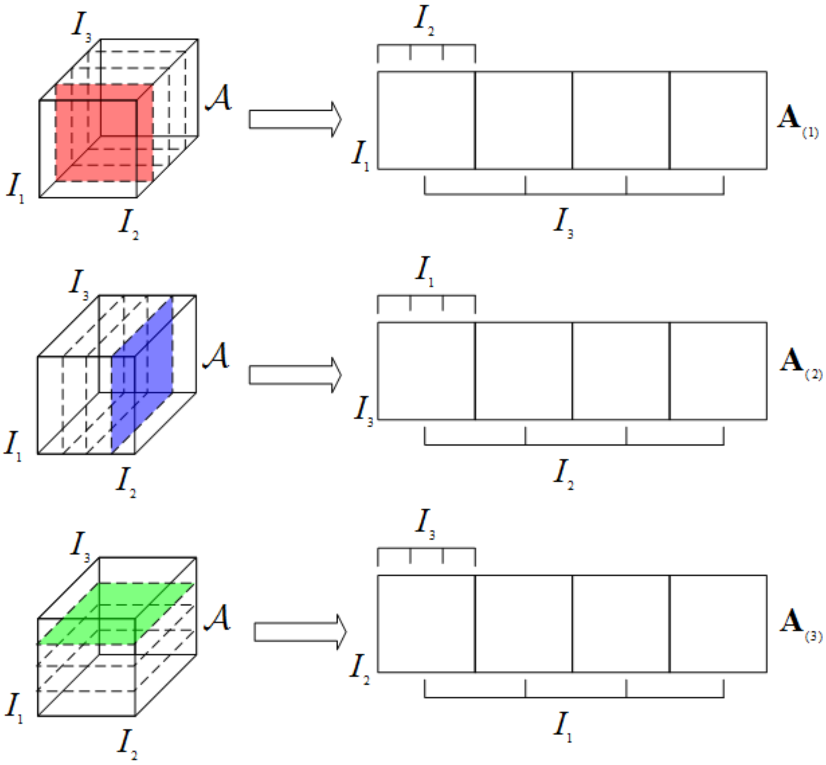


Figure 1: Illustration of *mode-n* unfolded matrix. Unfolding the tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times I_3}$ to the matrix $\mathbf{A}_{(1)} \in \mathbb{R}^{I_1 \times (I_2 I_3)}$, the matrix $\mathbf{A}_{(2)} \in \mathbb{R}^{I_2 \times (I_1 I_3)}$, and the matrix $\mathbf{A}_{(3)} \in \mathbb{R}^{I_3 \times (I_1 I_2)}$, respectively.

Definition 2. (**Mode product**). The *mode-n product* of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ and a matrix $\mathbf{U}_n \in \mathbb{R}^{I_n \times J_n} (J_n \leq I_n)$ is an $I_1 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots I_N$ tensor defined by

$$(\mathcal{A} \times_n \mathbf{U}_n^T)_{i_1 \ldots i_{n-1} j_n i_{n+1} \ldots i_N} = \sum_{i_n} a_{i_1 i_2 \ldots i_N} \mathbf{U}_{j_n i_n}, \tag{1}$$

where $\times_n$ denotes the mode-n product, $\mathbf{U}_n^T$ is the transposed matrix of $\mathbf{U}_n$, and $\mathbf{U}_{j_n i_n}$ is an element of $\mathbf{U}_n^T$. The *mode-n* product $\mathcal{Y} = \mathcal{A} \times_n \mathbf{U}_n^T$ can be expressed in terms of unfolded tensor $\mathbf{Y}_{(n)} = \mathbf{U}_n^T \mathbf{A}_{(n)}$.

Given two matrices $\mathbf{U} \in \mathbb{R}^{I_n \times J_n}$, $\mathbf{V} \in \mathbb{R}^{I_m \times J_m}$, the mode products can be represented as

$$\mathcal{A} \times_n \mathbf{U}^T \times_m \mathbf{V}^T = (\mathcal{A} \times_n \mathbf{U}^T) \times_m \mathbf{V}^T = (\mathcal{A} \times_m \mathbf{V}^T) \times_n \mathbf{U}^T, \quad (m \neq n), \tag{2}$$

which indicates that the order of the products is irrelevant for distinct modes in a series of products. If the modes are the same, then $\mathcal{A} \times_n \mathbf{U}^T \times_n \mathbf{V}^T = \mathcal{A} \times_n (V^T U^T)$.

Given an *N-order* tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$, the corresponding *mode-n* unfolded matrix is denoted as $\mathbf{A}_{(n)} \in \mathbf{R}^{I_n \times (I_1 \ldots I_{n-1} I_{n+1} \ldots I_N)}$. The tensor decomposition of $\mathcal{A}$ aims to seek for N orthonormal basis matrices $\mathbf{U}_n \in \mathbb{R}^{I_n \times J_n} (J_n < I_n, 1 \leqslant n \leqslant N)$, which are obtained by high-order SVD (HOSVD) in Algorithm 1. HOSVD is a convincing generalization of the matrix SVD, and performs on higher-order tensor $\mathcal{A}$ to compute the left singular matrix of $\mathbf{A}_{(n)}$ $(1 \leqslant n \leqslant N)$. Then the *mode-n* left singular matrix can be used as orthonormal basis matrix of $\mathbf{A}_{(n)}$ for decomposition.

Algorithm 1 HOSVD of tensor $\mathcal{A}$.

INPUT: $\mathcal{A}$.
for $n = 1$ to N **do**
  Do SVD on $\mathbf{A}_{(n)}$.
  $\mathbf{U}_n \leftarrow$ the left singular matrix of $\mathbf{A}_{(n)}$.
end for

$$\mathcal{C} = \mathcal{A} \times_1 \mathbf{U}_1^T \times_2 \mathbf{U}_2^T \ldots \times_N \mathbf{U}_N^T. \tag{3}$$

OUTPUT: $\mathcal{C}$, $\mathbf{U}_1, \mathbf{U}_2, \ldots, \mathbf{U}_N$

In Eq. (3), $\mathcal{C} \in \mathbb{R}^{J_1 \times J_2 \times \ldots \times J_N}$ is the core tensor, which governs the relationship among $\mathbf{U}_n (1 \leqslant n \leqslant N)$. The matrix representation of this decomposition can be obtained by

$$\mathbf{C}_{(n)} = \mathbf{U}_n^T \mathbf{A}_{(n)} (\mathbf{U}_{n+1} \otimes \mathbf{U}_{n+2} \otimes \ldots \otimes \mathbf{U}_N \otimes \mathbf{U}_1 \otimes \ldots \otimes \mathbf{U}_{n-1}), \tag{4}$$

where $\otimes$ denotes the Kronecker product, $\mathbf{C}_{(n)} \in \mathbb{R}^{J_n \times (J_1 J_2 \ldots J_{n-1} J_{n+1} \ldots J_N)}$ is the *mode-n* unfolded matrices of $\mathcal{C}$.

Definition 3. (**Tensor norm**). The *norm* of a tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is the square root of the sum of the squares of all its elements, represented by

$$\|\mathcal{A}\| = \sqrt{\langle \mathcal{A}, \mathcal{A} \rangle} = \sqrt{\sum_{i_1} \ldots \sum_{i_N} \mathcal{A}_{i_1 \ldots i_N}^2}, \tag{5}$$

which is analogous to the matrix Frobenius norm [19].

Definition 4. (**Scalar product**). The *scalar product* of two tensors $\mathcal{A}, \mathcal{B} \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is the sum of the products of their entries, represented by

$$\langle \mathcal{A}, \mathcal{B} \rangle = \sum_{i_1} \sum_{i_2} \ldots \sum_{i_N} \mathcal{A}_{i_1 i_2 \ldots i_N} \mathcal{B}_{i_1 i_2 \ldots i_N}, \tag{6}$$

when $\mathcal{A} = \mathcal{B}$, it follows that $\langle \mathcal{A}, \mathcal{A} \rangle = \|\mathcal{A}\|^2$.

More definitions of tensor can be found in [19].

2.2. Discriminant analysis of canonical correlations (DCC)

An action sample includes N frames. Each frame is vectorized to be a D-dimensional vector. These frames can be represented as a matrix in space $\mathbb{R}^{D \times N}$, where each column vector is a vectorized frame. The matrix is called an image set. The purpose of the DCC [12] is to find the discriminative transformation matrix using canonical correlations for image sets classification. Assume m image sets coming from C classes: $\{\mathbf{X}_1^1, \ldots, \mathbf{X}_{m_1}^1, \mathbf{X}_1^2, \ldots, \mathbf{X}_{m_2}^2, \ldots, \mathbf{X}_1^C, \ldots, \mathbf{X}_{m_C}^C\}$, where $\mathbf{X}_i^c \in \mathbb{R}^{D \times N}$ means the i-th image set in the c-th class, m_c is the number of image sets in the c-th class, and $m_1 + m_2 + \ldots + m_C = m$ is satisfied.

First of all, a d-dimensional subspace of $\mathbf{X}_i^c$ is represented by an orthonormal basis matrix $\mathbf{P}_i^c \in \mathbb{R}^{D \times d}$ *s.t.* $\mathbf{X}_i^c \mathbf{X}_i^{cT} = \mathbf{P}_i^c \Lambda_i^c \mathbf{P}_i^{cT}$, where Λ_i^c, $\mathbf{P}_i^c$ are the matrices of eigenvalues and eigenvectors of the d largest eigenvalues, respectively. Given an initialized identity matrix $\mathbf{T} \in \mathbb{R}^{D\times D}$, $\mathbf{T}^T\mathbf{P}_i^c$ is orthonormalized to $\mathbf{T}^T\mathbf{P}_i^{c'}$ by QR-decomposition, the details are discussed in [12]. For every pair of $\mathbf{P}_i^{c'}$ and $\mathbf{P}_j^{r'}$ $(1 \leqslant c, r \leqslant C)$, let the SVD of $\mathbf{P}_i^{c'T}\mathbf{T}\mathbf{T}^T\mathbf{P}_j^{r'}$ be

$$\mathbf{P}_i^{c'T}\mathbf{T}\mathbf{T}^T\mathbf{P}_j^{r'} = \mathbf{Q}_{ij}\Lambda\mathbf{Q}_{ji}{}^T, \quad s.t. \quad \Lambda = diag(\rho_1, \ldots, \rho_n), \tag{7}$$

where $\mathbf{Q}_{ij}$, $\mathbf{Q}_{ji}$ are the rotation matrices; $\rho_1, \ldots, \rho_n$ are the canonical correlations.

The similarity of two image sets is defined as the sum of canonical correlations:

$$\mathbf{F}_{ij} = \underset{\mathbf{Q}_{ij},\mathbf{Q}_{ji}}{max}\, tr\{\mathbf{T}^T\mathbf{P}_j^{r'}\mathbf{Q}_{ji}\mathbf{Q}_{ij}{}^T\mathbf{P}_i^{c'^T}\mathbf{T}\}, \tag{8}$$

where $\mathbf{T}$ is obtained to maximize the similarities of intra-class sets and minimize the similarities of inter-class sets, which is obtained by

$$\mathbf{T} = \arg\max_{\mathbf{T}} \frac{\sum_{i=1}^{m}\sum_{k\in W_i}\mathbf{F}_{ik}}{\sum_{i=1}^{m}\sum_{l\in B_i}\mathbf{F}_{il}}, \tag{9}$$

where $W_i = \{k|C_k = C_i\}$ and $B_i = \{j|C_j \neq C_i\}$ denote the intra-class image sets and inter-class image sets, respectively. C_k, C_j are the class label of $\mathbf{X}_k$ and $\mathbf{X}_j$, respectively. C_i is the class label of a given $\mathbf{X}_i$. The discriminant transformation matrix $\mathbf{T}$ is rewritten as

$$\mathbf{T} = \arg\max_{\mathbf{T}} \frac{tr(\mathbf{T}^T\mathbf{S}_b\mathbf{T})}{tr(\mathbf{T}^T\mathbf{S}_w\mathbf{T})}, \tag{10}$$

where

$$\mathbf{S}_b = \sum_{i=1}^{m}\sum_{c\neq r}(\mathbf{P}_j^{r'}\mathbf{Q}_{ji} - \mathbf{P}_i^{c'}\mathbf{Q}_{ij})(\mathbf{P}_j^{r'}\mathbf{Q}_{ji} - \mathbf{P}_i^{\prime c}\mathbf{Q}_{ij})^T \tag{11}$$

is the inter-class scatter matrix, $(1 \leqslant c, r \leqslant C)$,

$$\mathbf{S}_w = \sum_{i=1}^{m}\sum_{c=h}(\mathbf{P}_j^{h'}\mathbf{Q}_{ji} - \mathbf{P}_i^{c'}\mathbf{Q}_{ij})(\mathbf{P}_j^{h'}\mathbf{Q}_{ji} - \mathbf{P}_i^{\prime c}\mathbf{Q}_{ij})^T \tag{12}$$

is the intra-class scatter matrix, $(1 \leqslant c, h \leqslant C)$. Then $\mathbf{T}$ can be updated by the eigen-decomposition of $(\mathbf{S}_w)^{-1}\mathbf{S}_b$. The details are discussed in [12].

3. Incremental multi-linear discriminant analysis of canonical correlations

An action sample is naturally represented by an *N-order* tensor. The purpose of IMDCC method is to define the discriminant transformation matrix (DTM) $\mathbf{T}_n \in \mathbb{R}^{I_n \times J_n}$ $(J_n < I_n, 1 \leqslant n \leqslant N)$ which maps the original multi-linear space $\mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ to $\mathbb{R}^{J_1 \times J_2 \times \ldots \times J_N}$, using canonical correlations of incremental tensors. Assuming that m tensor samples come from C classes: $\{\mathcal{A}_1^1, \ldots, \mathcal{A}_{m_1}^1, \mathcal{A}_1^2, \ldots, \mathcal{A}_{m_2}^2, \ldots, \mathcal{A}_1^C, \ldots, \mathcal{A}_{m_C}^C\}$, where $\mathcal{A}_i^c \in \mathbb{R}^{I_1 \times I_2 \times \ldots \times I_N}$ is the i-th *N-order* tensor in the c-th class, m_c is the number of tensors in the c-th class, and $m_1 + m_2 + \ldots + m_C = m$ is satisfied. We seek N discriminant transformation matrices $\mathbf{T}_n \in \mathbb{R}^{I_n \times J_n} (1 \leqslant n \leqslant N)$ for projection

$$\mathcal{D}_i^c = \mathcal{A}_i^c \times_1 \mathbf{T}_1^T \times_2 \mathbf{T}_2^T \ldots \times_N \mathbf{T}_N^T, \quad 1 \leqslant i \leqslant m_c, \quad 1 \leqslant c \leqslant C, \tag{13}$$

where $\mathcal{D}_i^c \in \mathbb{R}^{J_1 \times \ldots \times J_N}$ is the dimensional reduced tensor belongs to c-th class, which is used for classification by NNC method.

The discriminant function $J(\mathbf{T}_n)$ is defined as the ratio of the similarities of any pairs of intra-class samples and the similarities of pairwise inter-class samples. According to Eqs. (8) $\sim$ (10), $J(\mathbf{T}_n)$ is defined by

$$J(\mathbf{T}_n) \stackrel{\text{def}}{=} \frac{tr(\mathbf{T}_n^T \mathbf{S}_b^{(n)} \mathbf{T}_n)}{tr(\mathbf{T}_n^T \mathbf{S}_w^{(n)} \mathbf{T}_n)}, \quad 1 \leqslant n \leqslant N, \tag{14}$$

where $\mathbf{T}_n$ is defined by

$$\mathbf{T}_n = \arg\max_{\mathbf{T}_n} J(\mathbf{T}_n) = \arg\max_{\mathbf{T}_n} \frac{tr(\mathbf{T}_n^T \mathbf{S}_b^{(n)} \mathbf{T}_n)}{tr(\mathbf{T}_n^T \mathbf{S}_w^{(n)} \mathbf{T}_n)}, \quad 1 \leqslant n \leqslant N, \tag{15}$$

which is found to maximize the discriminant function $J(\mathbf{T}_n)$; $\mathbf{S}_b^{(n)}$ is the *mode-n* inter-class scatter matrix; $\mathbf{S}_w^{(n)}$ is the *mode-n* intra-class scatter matrix. $\mathbf{T}_n (1 \leqslant n \leqslant N)$ ensures that the transformed tensors from the same class are distributed as close as possible, while the transformed tensors from different classes are distributed as far away as possible. The process of calculating $\mathbf{T}_n$ is represented as follows:

3.1. Multi-linear discriminant analysis of canonical correlations (MDCC)

First of all, the *N-order* tensor $\mathcal{A}_i^c$ is transformed by

$$\mathcal{A}_i^c \leftarrow \mathcal{A}_i^c \times_1 \mathbf{T}_1^T \ldots \times_{n-1} \mathbf{T}_{n-1}^T \times_{n+1} \mathbf{T}_{n+1}^T \ldots \times_N \mathbf{T}_N^T, \tag{16}$$

where $\leftarrow$ is the assignment operation. The updated $\mathcal{A}_i^c \in \mathbb{R}^{J_1\times\ldots\times J_{n-1}\times I_n\times J_{n+1}\times\ldots\times J_N}$ is *mode-n* unfolded to be a matrix $\mathbf{A}_{i(n)}^c \in \mathbb{R}^{I_n\times(J_1J_2\ldots J_{n-1}J_{n+1}\ldots J_N)}$, which is used for *mode-n* transformation by $\mathbf{T}_n \in \mathbb{R}^{I_n\times J_n}$. Since $\mathbf{T}_n$ depends on the discriminant transformation matrices of the other $(N-1)$ modes, it is hardly to find a closed form for $\mathbf{T}_n(1 \leqslant n \leqslant N)$. Thus it requires repeating the procedure for a certain number of iterations. The iterative optimization of $\mathbf{T}_n$ is described as follows:

After dimensional reduction of $\mathcal{A}_i^c$ in Eq. (16), eigen-decomposition is performed on $\mathbf{A}_{i(n)}^c{\mathbf{A}_{i(n)}^c}^T$, *s.t.*

$$\mathbf{A}_{i(n)}^c{\mathbf{A}_{i(n)}^c}^T = \mathbf{P}_i^{c(n)}\Lambda_i^{c(n)}{\mathbf{P}_i^{c(n)}}^T, \tag{17}$$

where the vectors of $\mathbf{P}_i^{c(n)} \in \mathbb{R}^{I_n\times J_n}$ are the eigenvectors of the J_n largest eigenvalues; $\mathbf{P}_i^{c(n)}$ is the *mode-n* linear subspace of $\mathbf{A}_{i(n)}^c$. Given an identity matrix $\mathbf{T}_n \in \mathbb{R}^{I_n\times I_n}$ for initialization, QR-decomposition of $\mathbf{T}_n^T\mathbf{P}_i^{c(n)}$ is performed *s.t.* $\mathbf{T}_n^T\mathbf{P}_i^{c(n)} = \Phi_i^{c(n)}\Delta_i^{c(n)}$, where $\Phi_i^{c(n)} \in \mathbb{R}^{I_n\times J_n}$ is the orthonormal matrix and $\Delta_i^{c(n)} \in \mathbb{R}^{J_n\times J_n}$ is the invertible upper-triangular matrix. Thus $\Phi_i^{c(n)} = \mathbf{T}_n^T(\mathbf{P}_i^{c(n)}(\Delta_i^{c(n)})^{-1})$. Assume $\mathbf{P}_i^{c'(n)} = \mathbf{P}_i^{c(n)}(\Delta_i^{c(n)})^{-1}$, so $\mathbf{T}_n^T\mathbf{P}_i^{c'(n)}$ is an orthonormal matrix. For every pair of $\mathbf{P}_i^{c'}$, $\mathbf{P}_j^{r'}$ $(1 \leqslant c, r \leqslant C)$, let the HOSVD of ${\mathbf{P}_i^{c'(n)}}^T T_n T_n^T \mathbf{P}_j^{r'(n)}$ be

$${\mathbf{P}_i^{c'(n)}}^T T_n T_n^T \mathbf{P}_j^{r'(n)} = \mathbf{Q}_{ij}^{(n)}\Lambda^{(n)}{\mathbf{Q}_{ji}^{(n)}}^T, \tag{18}$$

where $\Lambda^{(n)}$ is a singular matrix and $\mathbf{Q}_{ij}^{(n)}$, $\mathbf{Q}_{ji}^{(n)} \in \mathbb{R}^{J_n\times J_n}$ are orthogonal rotation matrices. Then the *mode-n* inter-class scatter matrix $\mathbf{S}_b^{(n)}$ and intra-class scatter matrix $\mathbf{S}_w^{(n)}$ can be calculated according to

$$\mathbf{S}_b^{(n)} = \sum_{i=1}^{m}\sum_{c\neq r}(\mathbf{P}_j^{r'(n)}\mathbf{Q}_{ji}^{(n)} - \mathbf{P}_i^{c'(n)}\mathbf{Q}_{ij}^{(n)})(\mathbf{P}_j^{r'(n)}\mathbf{Q}_{ji}^{(n)} - \mathbf{P}_i^{c'(n)}\mathbf{Q}_{ij}^{(n)})^T, \tag{19}$$

$$\mathbf{S}_w^{(n)} = \sum_{i=1}^{m}\sum_{c=h}(\mathbf{P}_j^{h'(n)}\mathbf{Q}_{ji}^{(n)} - \mathbf{P}_i^{c'(n)}\mathbf{Q}_{ij}^{(n)})(\mathbf{P}_j^{h'(n)}\mathbf{Q}_{ji}^{(n)} - \mathbf{P}_i^{c'(n)}\mathbf{Q}_{ij}^{(n)})^T, \tag{20}$$

where $1 \leqslant c, r, h \leqslant C$, $\mathbf{T}_n \in \mathbb{R}^{I_n\times J_n}$ are updated by eigen-decomposition of $(\mathbf{S}_w^{(n)})^{-1}\mathbf{S}_b^{(n)}$.

3.2. Incremental multi-linear discriminant analysis of canonical correlations (IMDCC)

Assuming $\mathcal{A}^c_{m+1}(1 \leqslant c \leqslant C)$ is a new added tensor to the training set, $\mathbf{P}^{c(n)}_{m+1}$ is the *mode-n* matrix of eigenvectors, $\mathbf{Q}^{(n)}_{m+1,i}$ is the rotation matrix between $\mathcal{A}^c_{m+1}$ and $\mathcal{A}^r_i$ $(1 \leqslant i \leqslant m, 1 \leqslant r \leqslant C)$, the incremental *mode-n* inter-class scatter matrix and intra-class scatter matrix are calculated according to

$$\Delta\mathbf{S}^{(n)}_b = \sum_{\substack{i=1 \\ c\neq r}}^{m} (\mathbf{P}^{c'(n)}_{m+1}\mathbf{Q}^{(n)}_{m+1,i} - \mathbf{P}^{r'(n)}_{i}\mathbf{Q}^{(n)}_{i,m+1})(\mathbf{P}^{c'(n)}_{m+1}\mathbf{Q}^{(n)}_{m+1,i} - \mathbf{P}^{r'(n)}_{i}\mathbf{Q}^{(n)}_{i,m+1})^T, \tag{21}$$

$$\Delta\mathbf{S}^{(n)}_w = \sum_{\substack{i=1 \\ c=h}}^{m} (\mathbf{P}^{c'(n)}_{m+1}\mathbf{Q}^{(n)}_{m+1,i} - \mathbf{P}^{h(n)}_{i}\mathbf{Q}^{(n)}_{i,m+1})(\mathbf{P}^{c'(n)}_{m+1}\mathbf{Q}^{(n)}_{m+1,i} - \mathbf{P}^{h(n)}_{i}\mathbf{Q}^{(n)}_{i,m+1})^T, \tag{22}$$

where $1 \leqslant h \leqslant C$, the updated *mode-n* inter-class scatter matrix and intra-class scatter matrix are calculated by

$$\mathbf{S}^{'(n)}_b = \mathbf{S}^{(n)}_b + \Delta\mathbf{S}^{(n)}_b, \tag{23}$$

$$\mathbf{S}^{'(n)}_w = \mathbf{S}^{(n)}_w + \Delta\mathbf{S}^{(n)}_w. \tag{24}$$

$\mathbf{T}_n$ is updated by eigen-decomposition of $(\mathbf{S}^{'(n)}_w)^{-1}\mathbf{S}^{'(n)}_b$. All the N $\mathbf{T}_n$ are initialized as an identity matrix, and are updated to realize optimization through an iterative learning with the incremental training samples. The computational procedure of $\mathbf{T}_n$ is represented by Eqs. (16) $\sim$ (24). The proposed method is summarized in Algorithm 2.

3.3. Discussion about convergence

The alternating optimization approach by iterative learning for $\mathbf{T}_n$ converges when the termination condition is satisfied as

$$Error(k) = \sum_{n=1}^{N} \| |\mathbf{T}^{(k)^T}_n \mathbf{T}^{(k-1)}_n| - I \| \leqslant \boldsymbol{\varepsilon},\ \textit{subject to}\ \mathbf{T}_n\mathbf{T}^T_n = I, \tag{25}$$

where $Error(k)$ and $Error(k-1)$ are the resulted error values from the k-th and (k-1)-th iteration, respectively. I is the identity matrix; $\boldsymbol{\varepsilon}$ is a pre-defined small threshold whose value is 0.1.

The convergence property of IMDCC in training procedure is shown in Fig. 2. In Fig. 2(a), the x-coordinate is the number of the training iterations k and the y-coordinate is the error value Error(k) between two neighboring training iterations, i.e., the Error(k) demonstrates how $\sum_{n=1}^{N} \||\mathbf{T}_n^{(k)^T}\mathbf{T}_n^{(k-1)}|-I\|$ changes in the training iterations with all mode features. From the sub-figure, we can see that as the number of training iterations increases, the change of Error(k) approaches to zero, which means the algorithm converges. In Fig. 2(b), the x-coordinate denotes the number of the training iterations and the y-coordinate is the discriminant function value (DFV) calculated by $\mathbf{T}_n$ in Eq. (14), i.e., how the DFV in each mode changes is shown in the training iterations. From the sub-figure, we can see that as the number of training iterations increases, the change of DFV in each mode approaches to zero. Since the DFV achieves its maximum and keeps stable, the optimal $\mathbf{T}_n$ can be obtained, as defined in Eq. (15). Both of the sub-figures demonstrate that the training procedure of IMDCC converges within 10 iterations.

Algorithm 2 IMDCC

INPUT: m labeled *N-order* original tensor samples $\Gamma_m = \{\mathcal{A}_1^1, \ldots, \mathcal{A}_{m_1}^1, \ldots, \mathcal{A}_1^C, \ldots, \mathcal{A}_{m_C}^C\}(m_1 + \ldots + m_C = m)$, L labeled incremental tensor samples $\Gamma_L = \{\mathcal{B}_1^1, \ldots, \mathcal{B}_{l_1}^1, \ldots, \mathcal{B}_1^C, \ldots, \mathcal{B}_{l_C}^C\}(l_1 + \ldots + l_C = L)$, class label $\{1, \ldots, C\}$, discriminant transformation matrices $\mathbf{T}_n, (1 \leqslant n \leqslant N)$.

OUTPUT: Updated $T_n, (1 \leqslant n \leqslant N)$.

Algorithm:

for $l = 1$ to L **do**

  Initialize $\forall \mathcal{A}_i^r \in \Gamma_m \bigcup \{\mathcal{B}_l^c \in \Gamma_L\}$, $1 \leqslant c, r \leqslant C$.

  repeat

    for $n = 1$ to N **do**

      Update $\mathcal{B}_l^c$ by Eq. (16), and calculate $\mathbf{P}_l^{c'(n)}$, $\mathbf{Q}_{li}^{(n)}(1 \leqslant i \leqslant m + l)$ by Eqs. (17) $\sim$ (18).

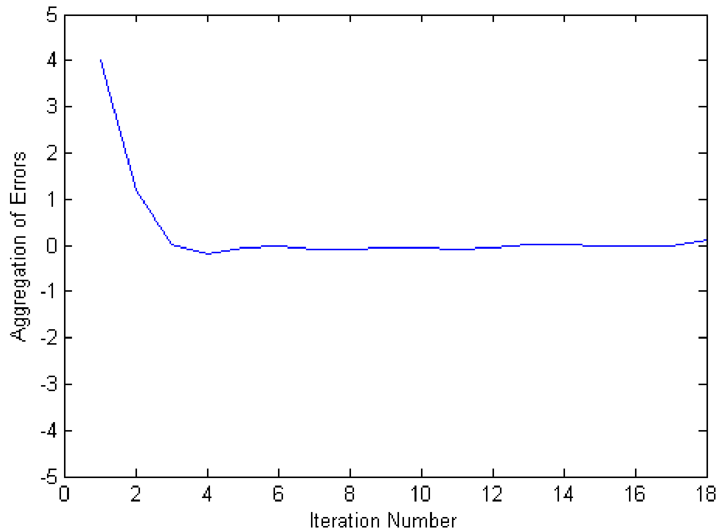
      Calculate $\mathbf{S}_b^{'(n)}$, $\mathbf{S}_w^{'(n)}$ by Eqs. (19) $\sim$ (24), and update T_n by eigen-decomposition of $(\mathbf{S}_w^{'(n)})^{-1}\mathbf{S}_b^{'(n)}$.

    end for

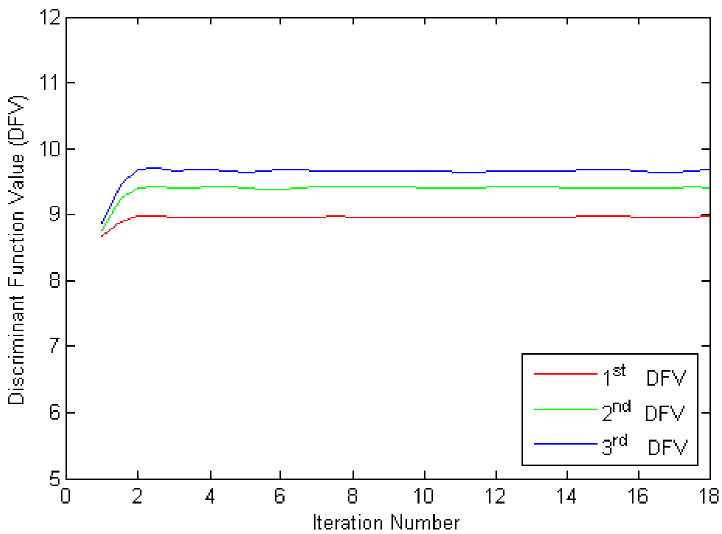
    Calculate $Error(k)$ by Eq. (25).

  until $Error(k) = \sum_{n=1}^{N} \||\mathbf{T}_n^{(k)^T}\mathbf{T}_n^{(k-1)}| - I\| \leqslant \boldsymbol{\varepsilon}$.

end for



(a)



(b)

Figure 2: Illustration of properties of IMDCC on the Weizmann database. (a) the aggregation of errors as the convergence check criterion Error(k), (b) the discriminant function value (DFV) in multi-linear subspace.

4. Experiment results

4.1. Actions from Weizmann database

The experiment was implemented on the Weizmann database, which is a commonly used database for human action recognition. There are 90 low-resolution (180×144, 25fps) videos which contain 10 action categories in the database.

We extracted 3500 samples from the 90 videos, where each sample consists of 20 successive frames and begins every other frame. We used 3000 samples for training and the remaining 500 samples for testing. Both the training set and testing set contain all the ten categories of actions. The training set was further partitioned into an initial set which was used for learning the initial discriminative model and the remaining samples were added consequently for re-training.

In order to represent the spatio-temporal feature of the samples, 20 successive frames of each action were taken to utilize the temporal feature. Each centered frame was normalized to the size of 64×48 pixels. Thus the tensor sample was represented in size of $64 \times 48 \times 20$ pixels. Fig. 3 shows a tensor sample of the bending action.

4.2. Comparative methods and parameter setting

In this section, we compare the performance of IMDCC/MDCC with other classical algorithms. These algorithms are discriminant analysis of canonical correlations (DCC) [12], incremental discriminant analysis of canonical correlations (IDCC) [13], and discriminant analysis with tensor representation (DATER) [17]. The DCC is a discriminative learning method for sets classification. The discriminant function is expected to maximize the canonical correlations of intra-class sets and minimize the canonical correlations of inter-class sets. The IDCC integrates incremental learning with the DCC framework, it performs the training process using increasing samples instead of using total training samples at one time. The DATER performed discriminant analysis to realize dimensionality reduction directly on tensors, without considering the canonical correlations between different tensors. The MDCC adapts DCC for tensor. The discriminant analysis using canonical correlations is performed in multi-linear subspace, and the discriminant transformation matrices are learnt by an iterative optimization approach. The IMDCC

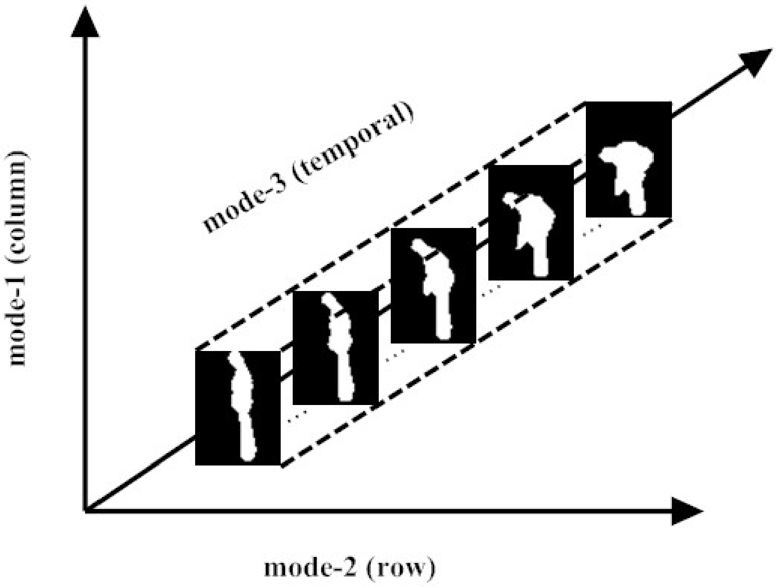


Figure 3: Example of the bending action in spatio-temporal domain from Weizmann database.

integrates incremental learning with the MDCC framework, it uses increasing tensors for training instead of using total training tensors for learning at one time.

To DCC and IDCC, the normalized silhouettes with 64×48 pixels were converted into 3072 dimensional vectors ($D = 3072$). Each image set had 20 vectors ($N = 20$) and was represented as a d-dimensional subspace ($d = 10$). To IMDCC, MDCC and DATER, the dimensions of original data were $I_1 = 64$, $I_2 = 48$, $I_3 = 20$, while the dimensions of subspaces were set to be $J_1 = 10$, $J_2 = 7$, $J_3 = 4$, respectively.

4.3. Computational Issues

Given an *N-order* tensor $\mathcal{A} \in \mathbb{R}^{I_1 \times \ldots \times I_N}$, which is dimensional reduced to be $\mathcal{B} \in \mathbb{R}^{J_1 \times \ldots \times J_N}$, where I_n, J_n are the dimensions of *mode-n* $\mathcal{A}$ and $\mathcal{B}$, respectively. For simplicity, it is assumed that $I_1 = \ldots = I_N = (\prod_{n=1}^{N} I_n)^{(1/n)} = I$, and $J_1 = \ldots = J_N = (\prod_{n=1}^{N} J_n)^{(1/n)} = J(J < I)$ likewise. From the computational complexity point of view, the most demanding steps of DATER [17] during training process are calculating the intra-class and inter-class scatter matrices, and the eigen-decomposition in tensor subspace. The com-

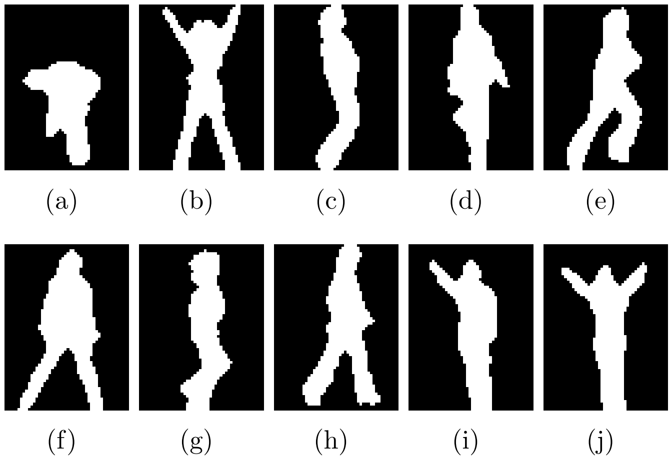


Figure 4: key silhouettes of ten actions from the Weizmann database.(a)bend, (b)jack, (c)jump, (d)pjump, (e)run, (f)side, (g)skip, (h)walk, (i)wave1, (j)wave2.

putational cost of DATER is $O(MNI^{N+1} + NI^3)$, where M is the number of total training samples, each of which is an *N-order* tensor. While the total training computational cost of MDCC is $O(MNI^2J^{N-1} + NI^3)$, which is less than that of DATER because of $J < I$. Accordingly, the computational cost of IMDCC is $O(NI^2J^{N-1} + NI^3)$, which is far less than that of MDCC and DATER. The most demanding steps of DCC [12] training process are calculating the intra-class and inter-class scatter matrices, and eigen-decomposition. Considering the image vectorization, the training computational cost of DCC is $O(MI(I^{N-1})^2 + (I^{N-1})^3) = O(MI^{2N-1} + I^{3N-3})$. While the computational cost of IDCC [13] is $O(I^{2N-1} + I^{3N-3})$. When $N > 2$, the computational cost of IDCC is far more than that of IMDCC. In a word, the computational time of IMDCC is less than the other methods mentioned above.

4.4. Action recognition experiment

The 90 videos contain ten categories of actions including bending (bend), jacking (jack), jumping (jump), jumping in places (pjump), running (run), galloping-sideways (side), skipping (skip), walking (walk), single-hand waving (wave1), both-hands waving (wave2), which are performed by nine subjects. The centered key silhouettes normalized to the size of 64×48 pixels for each action are shown in Fig. 4.

Fig. 5 demonstrates the recognition accuracy of IMDCC/MDCC and the

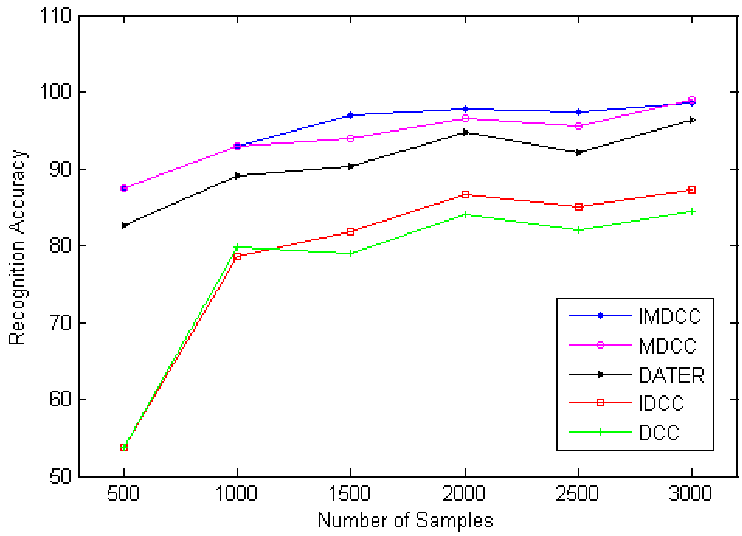


Figure 5: Recognition accuracy in the incremental stage.

other methods with the increasing training samples. The initial training set contained one-sixth of the total training samples with ten categories of actions. The remaining training set included five-sixth of the total training samples, one-sixth of which were added at each incremental stage for retraining. The IMDCC achieved approximate accuracy as MDCC, provided that enough components of the intra-class and inter-class canonical correlations were stored in multi-linear subspace. The IMDCC/MDCC methods had better accuracy than IDCC and DCC. The main reason is that IMDCC/MDCC have captured more effective information by preserving the original spatial structure of the samples. The performance of DATER was worse than IMDCC/MDCC, since DATER is based on single sample matching without exploiting the canonical correlations of multiple samples, and it hasn't updated the discriminant information by incremental learning.

Experiment on efficiency was performed between IMDCC and IDCC. Table 1 shows the experimental result, which lists how many seconds the training procedure cost using different number of samples. Fig. 6 illustrates the comparable results, which demonstrates that the computational cost is far lower on tensor data rather than matrix representation of data. Because it takes substantial time to calculate the discriminant transformation matrix of IDCC, we have chosen only a few samples for illustrating the training

Table 1: Computational time during training procedure.

Number / Time (s)	5	10	15	20
IMDCC	2	2	2.5	3
IDCC	120	240	450	545

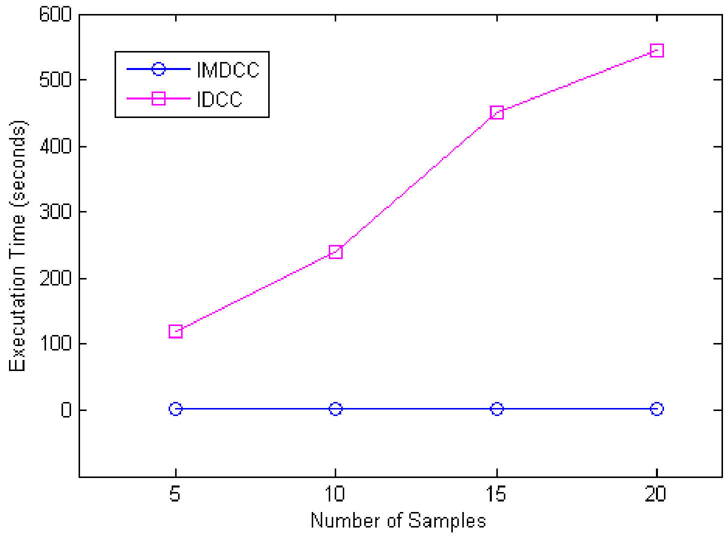


Figure 6: Computation efficiency of IMDCC and IDCC.

complexity comparison.

4.5. Robustness experiment

To test the robustness of the proposed method, an experiment was setup where the walking samples also came from the Weizmann database. In this experiment, a total of ten categories of walking actions were taken into consideration, including walking with a dog, swinging a bag, walking in a skirt, occluded feet, occluded by a "pole", moonwalk, limp Walk, walk with knees up, walk with a briefcase and normal walk. Fig. 7 shows samples of centered key silhouettes which were normalized to the size of 64×48 pixels for each actions.

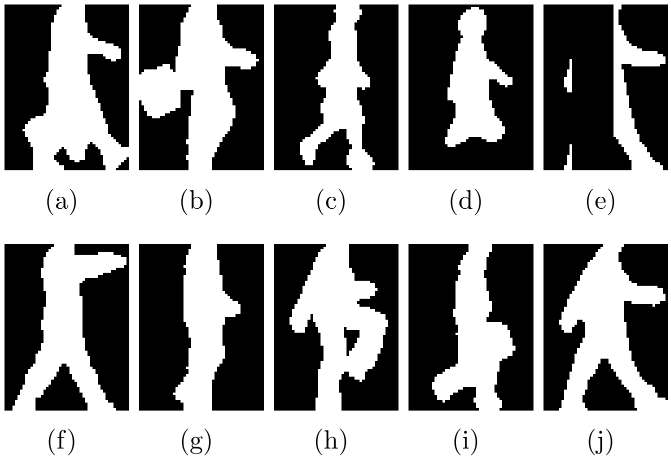


Figure 7: key silhouettes of walking from Weizmann database. (a) walking with a dog, (b) swinging a bag, (c) walking in a skirt, (d) occluded feet, (e) occluded by a "pole", (f) moonwalk, (g) limp Walk, (h) walk with knees up, (i) walk with a briefcase, (j) normal walk.

We extracted 98 samples from the walking videos for robustness testing. Fig. 8 shows the robustness experimental accuracies of IMDCC/MDCC and the other methods in the incremental stage. It is easy to see that the accuracy of each method has descended compared with the first accuracy experiment. The performance of MDCC and the IMDCC were quite similar to each other initially because of the same initial samples. However, the IMDCC outperformed the MDCC in the last two incremental stage because the discriminant transformation matrices are more effective by being updated incrementally. The IDCC and DCC achieved much better accuracies than the DATER because the canonical correlations of multiple samples preserve more effective discriminatory information than a single sample does.

5. Conclusion

This paper proposes a novel CCA-based feature extraction method, which iteratively learns the multi-linear discriminant subspace using the canonical correlations between different samples, named MDCC. We develop an online learning scheme for the MDCC named IMDCC, which is an optimization approach. IMDCC incrementally updates the discriminant transformation matrices, which can maximize the canonical correlations of intra-class samples while minimize the canonical correlations of inter-class samples.

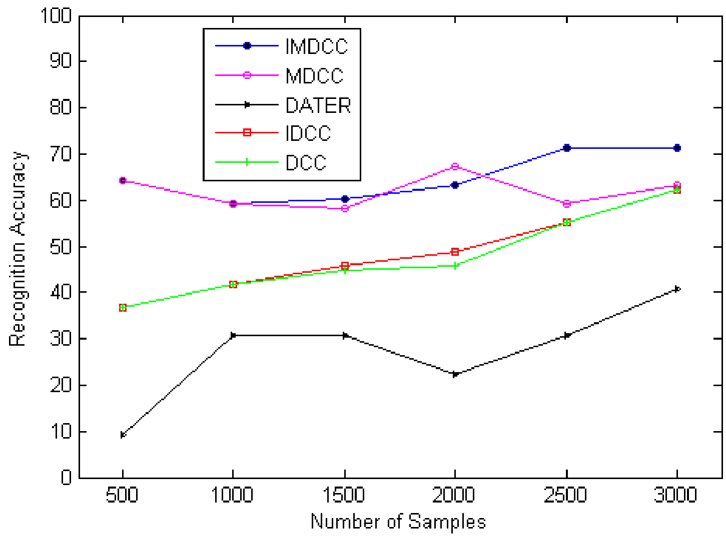


Figure 8: Robustness accuracy in the incremental stage.

The features of IMDCC combined with NNC significantly improved the accuracy of state-of-the-art action recognition methods. IMDCC is also practically appealing as it is robust against partial occlusion. Additionally, IMDCC was shown to be highly time efficient in training procedure, thus offering an attractive tool for recognition involving a large-scale database. Besides, IMDCC converges by iterative learning and also reduces the curse of dimensionality problem.

A sparse version of IMDCC will be proposed as future work, which reduces the chance of including unimportant variables in the canonical vectors and thus obtains better classification capability in multi-linear discriminant subspace. The action video will be used directly without extracting silhouettes, for the raw images reserve more sufficient spatial information.

Acknowledgements

This paper is supported by (1) the National Natural Science Foundation of China under Grant No. 60873146, 60973092, 60903097, (2) project of science and technology innovation platform of computing and software science (985 engineering), (3) the Key Laboratory for Symbol Computation and Knowledge Engineering of the National Education Ministry of China,

(4) the Natural Science Foundation of Jilin province of China under Grant No. 201115022.

[1] F. Zheng, L. Shao, Z. Song, A set of co-occurrence matrices on the intrinsic manifold of human silhouettes for action recognition, in: Proceedings of the ACM International Conference on Image and Video Retrieval, ACM, 2010, pp. 454–461.

[2] F. Zheng, L. Shao, Z. Song, Eigen-space learning using semi-supervised diffusion maps for human action recognition, in: Proceedings of the ACM International Conference on Image and Video Retrieval, ACM, 2010, pp. 151–157.

[3] L. Shao, X. Chen, Histogram of body poses and spectral regression discriminant analysis for human action categorization, in: British Machine Vision Conference (BMVC), Aberystwyth, UK, 2010.

[4] S. Wang, J. Yang, N. Zhang, C. Zhou, Tensor Discriminant Color Space for Face Recognition, IEEE Transactions on Image Processing. In Press, Accepted Manuscript.

[5] S. Wang, C. Zhou, N. Zhang, X. Peng, Y. Chen, X. Liu, Face recognition using second order discriminant tensor subspace analysis, Neurocomputing 74(12-13) (2011) 2142–2156.

[6] P. Belhumeur, J. Hespanha, D. Kriegman, Eigenfaces vs. fisherfaces: Recognition using class specific linear projection, Pattern Analysis and Machine Intelligence, IEEE Transactions on 19 (7) (1997) 711–720.

[7] C. Fraley, A. Raftery, Model-based clustering, discriminant analysis, and density estimation, Journal of the American Statistical Association 97 (458) (2002) 611–631.

[8] T. Shakunaga, K. Shigenari, Decomposed eigenface for face recognition under various lighting conditions, 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01) 1 (2001) 864–872.

[9] J. Tenenbaum, W. Freeman, Separating style and content with bilinear models, Neural Computation 12 (6) (2000) 1247–1283.

[10] A. Bissacco, A. Chiuso, Y. Ma, S. Soatto, Recognition of human gaits, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 2, IEEE, 2001, pp. II–52.

[11] F. Bach, M. Jordan, A probabilistic interpretation of canonical correlation analysis, in: University of California, Berkeley, Vol. 688, Citeseer, 2005.

[12] T. Kim, J. Kittler, R. Cipolla, Discriminative learning and recognition of image set classes using canonical correlations, IEEE Transactions on Pattern Analysis and Machine Intelligence (2007) 1005–1018.

[13] X. Wu, Y. Jia, W. Liang, Incremental discriminant-analysis of canonical correlations for action recognition, Pattern Recognition 43 (2010) 4190–4197.

[14] M. Vasilescu, D. Terzopoulos, Multilinear subspace analysis of image ensembles, 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition 2 (2003) 93.

[15] D. Tao, X. Li, X. Wu, S. Maybank, General tensor discriminant analysis and gabor features for gait recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence (2007) 1700–1715.

[16] D. Tao, X. Li, X. Wu, S. Maybank, Tensor Rank One Discriminant Analysis–A convergent method for discriminative multilinear subspace selection, Neurocomputing 71 (10-12) (2008) 1866–1882.

[17] S. Yan, D. Xu, Q. Yang, L. Zhang, X. Tang, H. Zhang, Discriminant analysis with tensor representation, in: Computer Vision and Pattern Recognition (CVPR'05), Vol. 1, IEEE Computer Society, 2005.

[18] J. Wen, X. Gao, Y. Yuan, D. Tao, J. Li, Incremental tensor biased discriminant analysis: A new color-based visual tracking method, Neurocomputing 73 (4-6) (2010) 827–839.

[19] T. Kolda, B. Bader, Tensor decompositions and applications, SIAM review 51 (3) (2009) 455–500.