Incremental Discriminative-Analysis of Canonical Correlations for Action Recognition

Xinxiao Wu   Wei Liang   Yunde Jia
Beijing Laboratory of Intelligent Information Technology, School of Computer Science
Beijing Institute of Technology, Beijing 100081, PR China
{wuxinxiao, liangwei, jiayunde }@bit.edu.cn

Abstract

Human action recognition is a challenging problem due to the large changes of human appearance in the cases of partial occlusions, non-rigid deformations and high irregularities. It is difficult to collect a large set of training samples with the hope of covering all possible variations of an action. In this paper, we propose an online recognition method, namely Incremental Discriminant-Analysis of Canonical Correlations (IDCC), whose discriminative model is incrementally updated to capture the changes of human appearance and thereby facilitates the recognition task in changing environments. As the training sets are acquired sequentially instead of being given completely in advance, our method is able to compute a new discriminant matrix by updating the existing one using the eigenspace merging algorithm. Experimental results on both Weizmann and KTH action data sets show that our method performs better than state-of-the-art methods on both accuracy and efficiency. Moreover, the robustness of our method is demonstrated on the irregular action recognition.

1. Introduction

Recognizing human actions has recently attracted increasing interests from computer vision for a wide range of promising applications, such as video indexing, visual surveillance, human-computer interaction, sports video analysis, and intelligent systems.

As motion speeds and body sizes are associated with individuals, the same action executed by different persons may exhibit large variations; while the environmental conditions such as lighting and view point may make the observations of different actions become similar. In order to reduce the variations of actions within the same class and suppress the environmental contributions to the similarities of actions in different classes, our method maximizes the canonical correlations of actions within the same class and minimizes the canonical correlations of actions between different classes. In the proposed algorithm for action recognition, each action is represented by an orthogonal linear subspace of sequential images and the similarity between two actions is defined by the canonical correlation of the corresponding two subspaces. We do not take into account the temporal dynamics of an action and in many cases several principal images even a single image is sufficient to recognize what a person is doing.

Another problem in action recognition rises from high irregularities of actions undergoing various non-stationary scenarios. Taking the walk action for example, people may walk with a dog, swinging a bag, carrying a briefcase or partially occluded by other objects. It is difficult to account for all the possible variations of an action during learning the discriminative model. In order to make the recognition task adapt to the changing image observations, we aim to find a discriminative model that can be online learned to describe the changes of human appearance and accurately classify the actions even in the irregular performances. Our method is capable of online updating the discriminative model with capturing images as the new training data, therefore the updated discriminative model can reflect the appearance variations. By merging eigenspace models [20], the proposed method updates the principal components of the total canonical correlations and between-class canonical correlations separately, and then computes the discriminant components directly from both updated principal component sets. To improve the computation efficiency of eigen-analysis, the sufficient spanning set [20] is adopted in the solution.

2. Introduction

2.1. Action recognition

Many approaches for human action recognition have been proposed in recent decades. Wang and Suter [1] used kernel principal component analysis to obtain the low-dimensional representation of human silhouette and introduced factorial conditional random field to model the motion. Their framework can effectively recognize human activities performed by different people with different body builds as well as different motion styles and speeds. Jia and Yeung [2] reported a new manifold embedding method to

2009 IEEE 12th International Conference on Computer Vision (ICCV)
978-1-4244-4419-9/09/$25.00 ©2009 IEEE

discover both the local spatial and temporal discriminant structures of human silhouette. They designed a two-stage recognition scheme to improve the recognition on low sensitivity to the temporal shape variation in the same action. Rodrigue et al. [3] introduced a template-based method for action recognition which is capable of capturing intra-class variability by synthesizing a single Action MACH filter for a given action class. Jhuang et al. [5] applied a biological model of motion processing to the action recognition by accounting for the dorsal stream of the visual cortex. Some other approaches [7-10, 12-14] extracted local spatio-temporal features to exploit rich and intrinsic representation and introduced statistical models to classify the action in large intra-class variations.

However, most of these approaches offline learn the recognition model and lack the adaptability to classify the different irregular actions which are not included in the training data. Our method can online efficiently update the discriminative model to learn the changes of human appearance with the superior adaptability to recognize high irregular actions. Moreover, to address the intra-class variation problem, our method maximizes the canonical correlations of within-class image sets and minimizes the canonical correlations of between-class sets.

2.2. Incremental learning

Recently, a number of incremental learning approaches have been proposed and applied to computer vision. Hall et al. [15] proposed Incremental Principle Component Analysis (IPCA) based on the update of covariance matrix through a residue estimating procedure. Then they improved their method by merging and splitting eigenspace models that allow a chunk of new samples to be learned in a single step [20]. Pang et al. [16] proposed an Incremental Linear Discriminant Analysis (ILDA) in two forms: sequential ILDA and chunk ILDA. The discriminant eigenspace is updated for classification when bursts of data are added to an initial discriminant eigenspace in the form of random chunks. As an improvement of ILDA, Kim et al. [17] applied the concept of the sufficient spanning set approximation in updating the between-class scatter matrix, the projected data matrix as well as the total scatter matrix. Lin et al. [18] handled with the online update of discriminative models for tracking objects undergoing large pose and lighting changes. In the image set-based recognition, Kim et al. [22] proposed an incremental method of learning orthogonal subspace. With the concept of the sufficient spanning set, the algorithm separately updates the principal components of the class correlation and total correlation matrices, and then computes the orthogonal components of the updated few principal components.

Many of these methods just combine all examples of a class together and do not exploit the concept of multiple sets in a single class. Our method maximizes the canonical correlations between multiple sets within the same class and is more robust to the intra-class changes.

3. Background

Table 1 demonstrates the important notations used throughout the paper.

Notations	Descriptions
X_i	i-th image set with each column describing an image
C_i	class label of X_i
P_i	orthonormal basis matrix representing the linear subspace of X_i
T	discriminant transformation matrix
S_b , S_t	transformed canonical correlations of between-class sets and total sets
V , Σ	eigenvector and eigenvalue matrices of S_b
U , Δ	eigenvector and eigenvalue matrices of S_t
m	number of image sets

Table1: Notations.

By analogy to the optimization concept of LDA [21], Discriminant-Analysis of Canonical Correlations (DCC) [19] introduces a linear discriminative function to maximize canonical correlations of within-class sets and minimize canonical correlations of between-class sets. Assume m image sets are given as $\{X_1, X_2, ..., X_m\}$, here X_i represents a matrix with each column describing an image. X_i belongs to one action class denoted by C_i. A d-dimensional linear subspace of X_i is represented by an orthonormal basis matrix $P_i \in R^{N\times d}$ s.t. $X_iX_i^T = P_i\Lambda_iP_i^T$. Λ_i and P_i are the eigenvalues and eigenvector matrices of the d largest eigenvalues, and N is the dimension of column vector. The discriminant transformation matrix $T = [t_1, ..., t_n] \in R^{N\times n}$ is defined by $Y_i = T^TX_i$ to make the transformed image sets more discriminative using canonical correlations. Orthonormal basis matrices of the subspaces of the transformed data are given by

$$Y_iY_i^T = (T^TX_i)(T^TX_i)^T = (T^TP_i)\Lambda_i(T^TP_i)^T. \qquad (1)$$

Canonical correlations are only defined for orthonormal basis matrices of subspace. Because T^TP_i is not generally orthonormal, the matrix P_i is normalized to $P_i^{'}$ so that the columns of $T^TP_i^{'}$ are orthonormal. By the SVD computation $(T^TP_i^{'})^T(T^TP_j^{'}) = Q_{ij}\Lambda Q_{ji}^T$, the similarity of two transformed data sets is defined as the sum of canonical correlations:

$$F_{ij} = \max_{Q_{ij},Q_{ji}} tr\{T^T P_j^{'} Q_{ji} Q_{ij}^T P_i^{'T} T\} \quad . \qquad (2)$$

T is determined to maximize the similarities of any pairs of within-class sets and minimize the similarities of pair-wise sets of different classes

$$T = \arg\max_T \frac{\sum_{i=1}^m \sum_{k\in W_i} F_{ik}}{\sum_{i=1}^m \sum_{l\in B_i} F_{il}} \quad . \qquad (3)$$

The two index sets $W_i = \{j | C_j = C_i\}$ and $B_i = \{j | C_j \neq C_i\}$ respectively denote the within-class and between-class sets for a given set of class C_i. By the simple linear algebra

$$T^T P_j^{'} Q_{ji} Q_{ij}^T P_i^{'T} T = I - T^T (P_j^{'} Q_{ji} - P_i^{'} Q_{ij})(P_j^{'} Q_{ji} - P_i^{'} Q_{ij})^T T/2 \ , \qquad (4)$$

the discriminative function is rewritten as

$$T = \arg\max_T tr(T^T S_b T) / tr(T^T S_w T) , \qquad (5)$$

where $S_b = \sum_{i=1}^m \sum_{l\in B_i} (P_l^{'} Q_{li} - P_i^{'} Q_{il})(P_l^{'} Q_{li} - P_i^{'} Q_{il})^T , B_i = \{j | C_j \neq C_i\}$,

$$S_w = \sum_{i=1}^m \sum_{k\in W_i} (P_k^{'} Q_{ki} - P_i^{'} Q_{ik})(P_k^{'} Q_{ki} - P_i^{'} Q_{ik})^T , \ W_i = \{j | C_j = C_i\} .$$

Finally the optimal T is computed by eigen-decomposition of $(S_w)^{-1} S_b$. Without losing generality, we assume in the rest of the paper all the P_i are normalized.

4. Incremental Discriminant-Analysis of Canonical Correlations (IDCC)

Two equivalent criterions to obtain the discriminant transformation matrix T are given by

$$\max_{\arg T} \frac{\sum_{i=1}^m \sum_{k\in W_i} F_{ik}}{\sum_{i=1}^m \sum_{l\in B_i} F_{il}} = \max_{\arg T} \frac{\sum_{i=1}^m \sum_{h\in T_i} F_{ih}}{\sum_{i=1}^m \sum_{l\in B_i} F_{il}} \ . \qquad (6)$$

$W_i = \{j | C_j = C_i\}$, $B_i = \{j | C_j \neq C_i\}$, $T_i = W_i \cup B_i$ indexes the total sets. $\sum_{i=1}^m \sum_{k\in W_i} F_{ik}$ and $\sum_{i=1}^m \sum_{l\in B_i} F_{il}$ respectively represent the canonical correlations of within-class sets and between-class sets. The total canonical correlations are represented by $\sum_{i=1}^m \sum_{h\in T_i} F_{ih} = \sum_{i=1}^m \sum_{k\in W_i} F_{ik} + \sum_{i=1}^m \sum_{l\in B_i} F_{il}$. In this paper, the algorithm uses the second criterion in Eq.6. By the simple linear algebra (Eq. 4), the discriminative function is

$$T = \max_T tr(T^T S_b T) / tr(T^T S_t T) , \qquad (7)$$

where $S_b = \sum_{i=1}^m \sum_{l\in B_i} (P_l Q_{li} - P_i Q_{il})(P_l Q_{li} - P_i Q_{il})^T , B_i = \{j | C_j \neq C_i\}$,

$$S_t = \sum_{i=1}^m \sum_{h\in T_i} (P_h Q_{hi} - P_i Q_{ih})(P_h Q_{hi} - P_i Q_{ih})^T , T_i = \{j \,|\, j = 1,..,m\} .$$

The solution of incremental discriminant-analysis canonical correlation includes three steps: updating the canonical correlations of the total sets, updating the canonical correlations of the between-class sets, and computing the discriminant transformation matrix. Note that S_b and S_t are respectively the linear algebra transformation (see Eq.4) of between-class canonical correlations and total canonical correlations, so the proposed algorithm actually involves the update of principal components of S_t , the update of principal components of S_b and the computation of T from the updated S_t and S_b.

4.1. Updating the total canonical correlations

Let the total canonical correlations of existing image sets be $\{U, \Delta, P_i, i = 1,2,...,m\}$, here U and Δ respectively denote the eigenvectors and eigenvalues of S_t , s.t. $S_t = U \Delta U^T$, and $P_i \in R^{N\times d}$ is a normalized orthonormal basis matrix of the i th existing set. Assume P_{m+1} is the normalized orthonormal basis matrix of a new data set, the update is defined as

$$\xi_1(U, \Delta, P_i, P_{m+1}) = (U^{'}, \Delta^{'}) \quad i = 1,2,..., m \ . \qquad (8)$$

Assume $A = 2\sum_{i=1}^m (P_{m+1} Q_{m+1,i} - P_i Q_{i,m+1})(P_{m+1} Q_{m+1,i} - P_i Q_{i,m+1})^T$, then the updated S_t is computed by $S_t^{'} = S_t + A = U \Delta U^T + A$. We wish to calculate the eigenvectors $U^{'}$ and eigenvalues $\Delta^{'}$ of $S_t^{'}$, i.e. $S_t^{'} = U^{'} \Delta^{'} U^{'T}$. To reduce the dimension of eigenvalue problem, the concept of the sufficient spanning set [20] is used. Let the SVD of A be $A = W \Psi W^T$, here W and Ψ are respectively the eigenvectors and eigenvalues. The sufficient spanning set of $S_t^{'}$ can be calculated by $\Phi_t = h([U, W])$ with h an orthonormalization function. Then $U^{'}$ is written as $U^{'} = \Phi_t R_t$ and R_t is a rotation matrix. Thus, we solve a smaller eigen-problem to obtain R_t and $\Delta^{'}$:

$$\begin{aligned} & S_t^{'} = U^{'} \Delta^{'} U^{'T} = \Phi_t R_t \ \Delta^{'} R_t^T \Phi_t^T \\ & \Rightarrow \Phi_t^T (S_t + A) \Phi_t = \Phi_t^T (U \Delta U^T + A) \Phi_t = R_t \Delta^{'} R_t^T \end{aligned} \quad . \qquad (9)$$

Suppose that d_t and d_A are the number of eigenvectors of U and W respectively, the matrix $\Phi_t^T S_t^{'} \Phi_t$ has the reduced size $d_t^{'} = d_t + d_A$ and the eigen-analysis of $S_t^{'}$ takes only $O((d_t + d_A)^3)$ computations. Let m be the number of existing training sets, the eigen-analysis of A requires $O(m^3)$ and the total cost of our incremental method is $O((d_t + d_A)^3 + m^3)$. While the eigen-analysis of $S_t^{'}$ in batch mode requires $O(m^6)$. Typically, d_t and d_A are (much) less than m^2 and m respectively.

4.2. Updating the between-class canonical correlations

The between-class canonical correlations of existing data sets are represented by $\{V, \Sigma, P_i, C_i, i = 1,2,..., m\}$, here V and Σ are the eigenvectors and eigenvalues of S_b, s.t. $S_b = V\Sigma V^T$. P_i and C_i respectively represent the normalized orthonormal component matrix and class label of the i th set. Given a new image set represented by an orthonormal basis matrix P_{m+1} and the corresponding class label C_{m+1}, the update is described as

$$\xi_2(V, \Sigma, P_i, C_i, P_{m+1}, C_{m+1}) = (V^{'}, \Sigma^{'}) \quad i = 1,2,...,m. \tag{10}$$

The updated S_b is computed by $S_b^{'} = V\Sigma V^T + F$, where $F = 2\sum_{i \in E}(P_{m+1}Q_{m+1,i} - P_iQ_{i,m+1})(P_{m+1}Q_{m+1,i} - P_iQ_{i,m+1})^T$ and $E = \{j | C_j \neq C_{m+1}\}$. $Q_{m+1,i}$ and $Q_{i,m+1}$ are obtained by the SVD solution $P_i^T TT^T P_{m+1} = Q_{i,m+1}\Lambda Q_{m+1,i}^T$. Let Z be the eigenvectors of F obtained by SVD solution, the sufficient spanning set of $S_b^{'}$ can be given by $\Phi_b = h([V, Z])$. $V^{'} = \Phi_b R_b$ with R_b a rotation matrix. Accordingly, the new small dimensional eigen-problem is expressed by

$$\begin{aligned} &S_b^{'} = V^{'}\Sigma^{'}V^{'T} \\ &\Rightarrow \Phi_b^T S_b^{'} \Phi_b = \Phi_b^T(V\Sigma V^T + F)\Phi_b = R_b\Sigma^{'}R_b^T \end{aligned} \tag{11}$$

Let n_k be the number of sets belonging to class k, then the eigen-analysis of F costs $O((m - n_{C_{m+1}})^3)$. Suppose d_b and d_F are the number of eigenvectors of V and F respectively, the matrix $\Phi_b^T S_b^{'} \Phi_b$ has the reduced size $d_b^{'} = d_b + d_F$. The eigen-analysis of $S_b^{'}$ requires at most $O((d_b + d_F)^3)$, whereas the eigen-analysis of the new between-class canonical correlations in batch mode costs $O((m^2 - \sum_k n_k^2)^3)$ with $m = \sum_k n_k$ .Typically, d_b and d_F are respectively (much) less than $(m^2 - \sum_k n_k^2)$ and $(m - n_{C_{m+1}})$.

4.3. Updating the discriminant transformation matrix

The discriminant transformation matrix is computed using the updated total canonical correlations and between-class canonical correlations:

$$\xi_3(U^{'}, \Delta^{'}, V^{'}, \Sigma^{'}) = T^{'}. \tag{12}$$

In order to further reduce the computation complexity, we introduce new sufficient spanning set to change eigen-analysis into a smaller dimensional eigenvalue problem requiring cost of $O(d_b^{'3})$ rather than $O(d_t^{'3})$. Let $G = U^{'}\Delta^{'-1/2}$, then $G^T S_t^{'} G = I$. As the denominator of the second criterion in Eq.6 is the identity matrix, the problem is to find the discriminative components that maximize $G^T S_b^{'} G$, s.t. $G^T S_b^{'} G = H\Lambda H^T$. The final discriminant components are obtained by $T^{'} = GH$. The sufficient spanning set of the projection data can be constructed by $\Omega = h([G^T V^{'}])$ and the eigenvalue problem is

$$\begin{aligned} &G^T S_b^{'} G = \Omega R \Lambda R^T \Omega^T \\ &\Rightarrow \Omega^T G^T V^{'}\Sigma^{'}V^{'T} G\Omega = R\Lambda R^T \end{aligned}. \tag{13}$$

The updated discriminant matrix is given by

$$T^{'} = GH = G\Omega R. \tag{14}$$

```
Algorithm Incremental Discriminant-Analysis Canonical
          Correlations
Input: The total and between-class canonical correlations
eigen-models $\{U, \Delta, V, \Sigma, P_i, C_i, i = 1,2,...,m\}$ of the existing
data sets and the normalized orthonormal basis matrix $P_{m+1}$
of the new data set with its label $C_{m+1}$
Output: Updated discriminant matrix $T^{'}$

1. Update the total canonical correlations.
   Compute $A$ and $A = W\Psi W^T$. Set $\Phi_t$ by $\Phi_t = h([U, W])$.
   Compute the eigenvectors $R_t$ of $\Phi_t^T(U\Delta U^T + A)\Phi_t$.
   $U^{'} = \Phi_t R_t$.
2. Update the between-class canonical correlations.
   Compute $F$ and $F = Z\Pi Z^T$. Set $\Phi_b$ by $\Phi_b = h([V, Z])$.
   Compute the eigenvectors $R_b$ of $\Phi_b^T(V\Sigma V^T + F)\Phi_b$.
   $V^{'} = \Phi_b R_b$.
3. Update the discriminant matrix.
   Compute $G = U^{'}\Delta^{'-1/2}$, $\Omega = h([G^T V^{'}])$ and the eigenvectors
   $R$ of $\Omega^T G^T V^{'}\Sigma^{'}V^{'T} G\Omega$. $T^{'} = G\Omega R$.
```

Table2: Procedure of IDCC

Let $d_b^{'}$ be the number of eigenvectors $V^{'}$, the time for eigen-problem in Eq.13 takes $O(d_b^{'3})$. The dimension $d_t^{'}$ of $U^{'}$ is usually larger than $d_b^{'}$, so the computation efficiency of $T^{'}$ improves from $O(d_t^{'3})$ to $O(d_b^{'3})$. The procedure of the complete IDCC algorithm is listed in Table 2.

5. Experiments

We have conducted experiments to evaluate the performance of the proposed method on two publicly available datasets: KTH human dataset and Weizmann human dataset. For all experiments, the non-optimized Matlab codes run on a Dell PC with Intel Pentium D 3.4 GHz CPU and 1G RAM.

5.1. Weizmann action recognition

We have tested the proposed method on the Weizmann action dataset [11]. There are about 90 low-resolution (180 × 144, 25fps) video sequences showing nine different subjects, each performing 10 actions including bending (bend), jumping jack (jack), jumping-forward-on-two-legs (jump), jumping-in-place-on-two-legs (pjump), running (run), skipping (skip), galloping-sideways (side), walking (walk), waving-one-hand (wave1) and waving-two-hands (wave2). The centered silhouettes extracted in [11] are normalized to the same 64 × 48 dimension and converted into 3072 dimensional vectors in a raster-scan manner. The classification accuracy is evaluated under nine-fold cross validation. Each time we take the silhouette frames of eight subjects for training and use those of the remaining one subject for testing. The training dataset is further partitioned into an initial set which is used for learning the initial discriminative model and the remaining sets which are added successively for re-training.

The efficiency and accuracy of IDCC have been examined by comparing it with DCC [19], ILDA [16] and IPCA [15]. Particularly, we are interested in evaluating the discriminability and execution time of IDCC with the increasing datasets. In IDCC and DCC, the best dimension of the linear subspace of each image set is around 19 to represent 99 percent information and the Nearest Neighbor (NN) classification is utilized based on the similarity between subspaces. PCA is performed to learn the linear subspace of each set in IDCC and DCC. For ILDA and IPCA, the dimensions of eigenspace are set to 8 and 28 respectively, and the k-Nearest Neighbor (k=10) is used for classification. Figure 1 demonstrates the recognition accuracy of IDCC and the related methods with the increasing training data. IDCC achieves nearly the same accuracy as DCC, provided that enough components of the total and between-class canonical correlations are stored. 10NN-ILDA and 10NN-IPCA perform worse since they are based on single image matching without exploiting the multiple image sets. The comparison of computational costs between DCC and IDCC is illustrated in Figure 2. Whereas the execution time of DCC increases significantly with the training samples arriving successively, the time of the IDCC remains low. Table 3 concludes the mean and standard deviation of recognition accuracy for different methods. Both IDCC and DCC provide significant improvements on recognition accuracy over IPCA and ILDA. Table 4 demonstrates the recognition rates of IDCC as well as some art-of-state methods, and all these methods adopt the evaluation scheme of leaving one out cross validation. As shown in Table 4, our method is significantly superior to the previous methods.

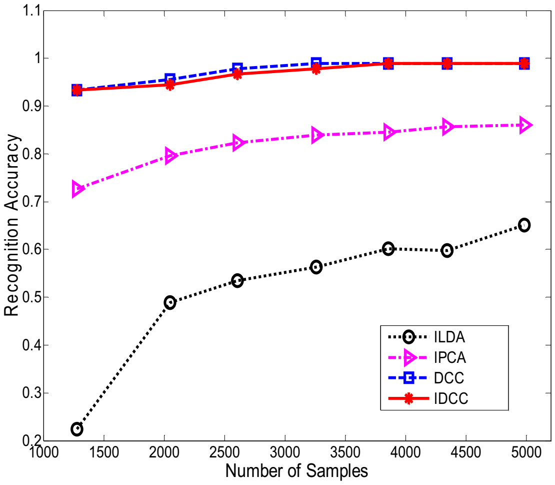

Figure1: Recognition accuracy of incremental solution.

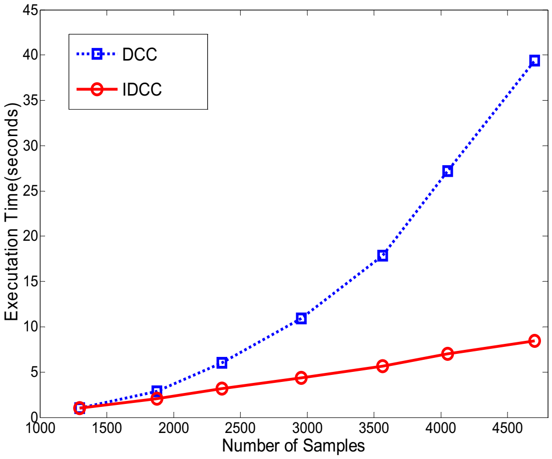

Figure 2: Computation efficiency of IDCC and DCC.

Methods	Weizmann Accuracy (%)
10NN-IPCA	86.09 ± 0.03(28)
10NN-ILDA	65.15 ± 0.08(8)
DCC	98.89 ± 0.02(19)
IDCC	98.89 ± 0.02(19)

Table 3: Mean and standard deviation of recognition accuracy for different methods. The number in parentheses represents the dimensionality of linear subspace.

Methods	Weizmann Accuracy (%)
Our method	98.9
Wang and Suter [1]	97.8
Zhang et al. [13]	92.9
Ali et al. [6]	92.6
Jia and Yeung [2]	90.9
Scovanner et al. [14]	84.2
Niebles and Li [7]	72.8

Table 4: Recognition accuracy of some related recognition approaches. All these approaches use the evaluation scheme of leaving one out cross validation

5.2. KTH action recognition

The KTH human action dataset [12] contains six types of human actions: walking, jogging, running, boxing, hand waving and hand clapping. These actions are performed several times by twenty-five subjects in four different scenarios: outdoors (s1), outdoors with scale variation (s2), outdoors with different clothes (s3) and indoors with lighting variation (s4). Some body tracking methods (e.g. [23]) can be applied to locate the areas and the geometric centers of human bodies in each frame, and the centered body region is normalized to the size of 50×50. Since the scenario s2 is only the scale variation of s1, the normalized human image of s2 are very similar to that of s1 and we conduct the experiment on s1, s3 and s4. Leave-one-out cross-validation is performed to test the proposed method, i.e. for each run the image sets of 24 subjects are used for training and the image sets of the remaining subject are for testing. Figure 3 demonstrates the recognition rates of incremental solution between IDCC, DCC, IPCA and ILDA. Table 5 shows the mean and standard deviation of recognition accuracy for different methods on s1, s3 and s4 datasets of KTH. Moreover, we compare the recognition rates between some related recognition methods in Table 6.

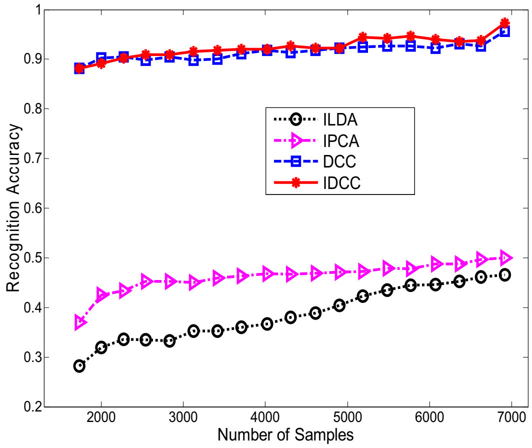

Figure 3: Recognition rates of incremental solutions.

	KTH s1	KTH s3	KTH s4	Avg
10NN-IPCA	51.46 ± 0.16	49.82 ± 0.14	50.69 ± 0.15	50.66 ± 0.15
10NN-ILDA	48.07 ± 0.13	39.77 ± 0.13	53.62 ± 0.08	47.15 ± 0.11
DCC	94.67 ± 0.05	89.33 ± 0.10	97.67 ± 0.02	93.89 ± 0.06
IDCC	96.00 ± 0.06	90.67 ± 0.10	98.67 ± 0.02	95.11 ± 0.06

Table 5: Comparisons of recognition rates. s1, s3, s4 corresponds to different conditions of the KTH database and Avg to the mean performance across the sets.

Methods	KTH Accuracy (%)
Our method	95.1
Zhang et al. [13]	91.3
Savarese et al. [8]	86.8
Wang et al. [4]	85.0
Niebles et al. [9]	81.5
Dollar et al. [10]	81.7

Table 6: Recognition accuracy on KTH dataset comparison between related recognition approaches. All these approaches use the evaluation scheme of leaving one out cross validation

5.3. Robustness test

To evaluate the adaptability and robustness of IDCC to the irregular actions in changing scenarios, we conduct the experiment on 10 video sequences of people walking in various difficult scenarios [11], including walking with a dog, walking when swinging a bag, walking in a skirt, walking with partially occluded legs, walking occluded by pole, sleepwalking, limping, walking with knees up, walking when carrying a briefcase, and normal walking.

In this experiment, the testing action is recognized on a frame-by-frame basis and the recognition accuracy is measured in terms of the percentage of the correctly recognized frames among the whole sequence. At each frame, we collect its local temporal neighbors as test image set and acquire the class label by computing the canonical correlations between the test set and those training sets. With the time process, several recognized frames are accumulated to construct the new training set for online re-training the discriminative model. For the trade-off between computational efficiency and effectiveness, we update the discriminative model at each interval of several frames rather than each frame. Table 7 gives the comparison results of LSTDE [2], DCC and IDCC, from which we can see that IDCC indeed flexibly adapts to the irregular action recognition from ever-changing silhouettes via online updating the discriminative model.

Test sequence	Recognition accuracy (%)		
	LSTDE	DCC	IDCC
Walk with a dog	90.74	100	100
Swinging a bag	74.58	100	100
Walk in a skirt	92.16	80.49	85.37
Occluded feet	71.19	77.55	93.88
Occluded by pole	-	84.62	92.31
Moonwalk	56.06	78.57	94.64
Limp walk	83.96	87.50	90.63
Walk with knees up	66.02	64.52	72.10
Carry a briefcase	92.86	100	100
Normal walk	95.16	100	100

Table 7: Comparison of robustness test results between LSTDE, DCC and IDCC.

6. Conclusions

We have presented a novel Incremental Discriminant-Analysis Canonical Correlation (IDCC) method and its application to the online human action recognition in various changing scenarios. By efficiently updating the discriminative model, IDCC can adapt to the appearance variations of human and accurately recognize the action even in high irregular performance. Experiments on both regular and irregular actions have shown the superior discriminability in classification, significant adaptability to changing environments and high computational efficiency of learning.

7. Acknowledgements

This work was partially supported by the Natural Science Foundation of China (60675021), the 973 Program of China (2006CB303103), and the Chinese High-Tech Program (2009AA01Z323).

References

[1] L. Wang and D. Suter. Recognizing human activities from silhouettes: motion subspace and factorial discriminative graphical model. In CVPR, 2007.

[2] K. Jia and D. Y. Yeung. Human action recognition using local spatio-temporal discriminant embedding. In CVPR, 2008.

[3] M. D. Rodriguez, J. Ahmed and M. Shah. Action MACH: A spatio-temporal maximum average correlation height filter for action recognition. In CVPR, 2008.

[4] Y. Wang, P. Sabzmeydani and G. Mori. Semi-latent dirichlet allocation: A hierarchical model for human action recognition. In HUMO, 2007.

[5] H. Jhuang, T. Serre, L. Wolf and T. Poggio. A Biologically Inspired System for Action Recognition. In ICCV, 2007.

[6] S. Ali, A. Basharat and M. Shah. Chaotic invariants for human action recognition. In ICCV, 2007.

[7] J. C. Niebles and F. F. Li. A Hierarchical model of shape and appearance for human action classification. In CVPR, 2007.

[8] S. Savarese, A. DelPozo, J.C. Niebles and F.F. Li. Spatial-temporal correlations for unsupervised action classification. In IEEE Workshop on Motion and Video Computing, 2008.

[9] J.C. Niebles, H.C. Wang and F.F. Li. Unsupervised learning of human action categories using spatial-temporal words. In BMVC, 2006.

[10] P. Dollar, V. Raband, G. Cottrell and S. Belongie. Behavior recognition via sparse spatio-temporal features. In VS-PETS, 2005.

[11] M. Blank, L. Gorelick, E. Shechtman, M. Irani and R. Basri. Actions as space-time shapes. In ICCV, 2005.

[12] C. Schuldt, I. Laptev and B. Caputo. Recognizing human actions: A local SVM approach. In ICPR, 2004.

[13] Z.M. Zhang, Y.Q. Hu, S. Chan and L.T. Chia. Motion Context: A New Representation for Human Action Recognition. In ECCV, 2008.

[14] P. Scovanner, S. Ali and M. Shah. A 3-dimensional sift descriptor and its applications to action recognition. In ACM Multimedia, 2007.

[15] P. Hall and R. Martin. Incremental eigenanalysis for classification. In BMVC, 1998.

[16] S.N. Pang, S. Ozawa and N. Kasabov. Incremental linear discriminant analysis for classification of data streams. In IEEE Transactions on Systems, and Cybernetics-Part B: Cybernetics: vol.35, No.5, October 2005.

[17] T. K. Kim, S. F. Wong, B. Stenger, J. Kittler and R. Cipolia. Incremental linear discriminant analysis using sufficient spanning set approximations. In CVPR, 2007.

[18] R. S. Lin, D. Ross, J. Lim and M. H. Yang. Adaptive discriminative generative model and its applications. In NIPS, 2004.

[19] T. K. Kim, J. Kitter and R. Cipolla. Discriminative learning and recognition of image set classes using canonical correlations. In IEEE Transactions on Pattern Analysis and Machine Intellegence, vol.29, no.6, June, 2007.

[20] P. Hall, D. Marshall and R. Martin. Merging and splitting eigenspace models. In IEEE Transactions on Pattern Analysis and Machine Learning. vol.22, no.9, pp.1042-1049, Sep.2000.

[21] R. O. Duda, P. E. Hart and D. G. Stork. Pattern classification, seconded. John Wily and Sons, 2000.

[22] T.K. Kim, J. Kittler and R. Cipolla. Incremental Learning of Locally Orthogonal Subspaces for Set-based Object Recognition. In BMVC, 2006.

[23] A. Bissacco, M.H. Yang amd S. Soatto. Fast human pose estimation using appearance and motion via multi-dimensional boosting regression. In CVPR, 2007.