INSTANCE-SPECIFIC CANONICAL CORRELATION ANALYSIS FOR POSE ALIGNMENT

Deming Zhai[1], Hong Chang[2], Xilin Chen[2], Wen Gao[1,3]

[1]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China, 150001
[2]Key Laboratory of Intelligent Information Processing, Institute of Computing Technology, Chinese Academy of Sciences (CAS), Beijing, China, 100080
[3]School of Electronic Engineering & Computer Science, Peking University, Beijing, China, 100871

ABSTRACT

Canonical correlation analysis (CCA) based methods achieve great success for pose alignment. However, CCA has limitations as a linear and global algorithm. Although some variants have been proposed to overcome the limitations, neither of them achieves locality and nonlinearity at the same time. In this paper, we propose a novel algorithm called Instance-Specific Canonical Correlation Analysis (ISCCA), which approximates the nonlinear data by computing the instance specific projections along the smooth curve of the manifold. Based on the framework of least squares regression, CCA is extended to the instance-specific case which obtains a set of locally-linear smooth but globally-nonlinear transformations. The optimization problem is proved to be convex and could be solved efficiently by alternating optimization. And the globally optimal solutions could be achieved with theoretical guarantee. Experimental results for pose alignment demonstrate the effectiveness of our proposed method.

1. INTRODUCTION

Pose estimation are essential pre-processing step to many applications, such as face recognition, eye gaze estimation, face tracking and human-computer interaction. Thus they are the crucial issues to be solved in computer vision area. The goal of pose estimation is to detect yaw, pitch, and roll angles of the head as pictured in Figure 1. One family of approaches address pose estimation by training independent classifies for each pose. However, it does not scale well due to the requirement of excessive labeled training data. Another emerging line of attack is to do precisely estimation via pose alignment. More specifically, if one person is labeled with all known poses, we can estimate the poses of the other persons just by aligning their facial images to the standard image set. Along this line of direction, canonical correlation analysis (CCA) based methods [1, 2] have provided very promising results and achieved great success for pose estimation.

Canonical Correlation Analysis (CCA) [3] is one of the most popular statistical methods to model the correlations between two views. The goal of CCA is to seek a pair of linear transformations that maximize the correlation between two views after projecting the data to a common lower-dimensional space by applying the transformations. However, a limitation of CCA is that it can only reveal the linear correlation relationship in a global way, making it inadequate for some more complicated applications.

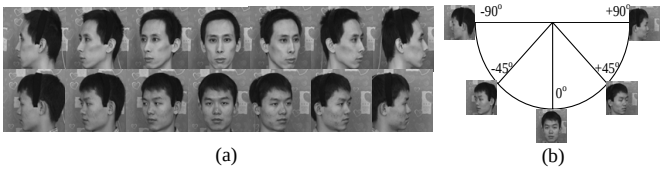


Fig. 1. FACE-10 data set: (a) two sample image sequences; (b) coordinate settings of different poses.

Since in general the poses of one person usually lie in a high-dimensional nonlinear space, Melzer *et al.*[1] first propose to do pose alignment via kernel CCA (KCCA). KCCA is a kernel extension of CCA which essentially applies CCA to data from different views after applying kernel-induced feature mappings. However, like most kernel methods, the nonlinearity in KCCA is applied in a global way in the sense that the nonlinear mapping is uniform anywhere,*i.e*, the induced kernel function with the same parameter is applied to all data pairs. Moreover, to deliver good performance, the choice of both the problem-dependent kernel function and its parameters is still a sticky problem.

Later, locality preserving CCA (LPCCA) [2] was proposed for pose alignment by introducing local manifold structure into CCA. In LPCCA, the globally nonlinear problem is decomposed into a series of locally linear sub-problems whose solutions can be combined to give the mapping vectors. Although local information is considered, the transformation learned in LPCCA is still global due to that the same transformation is applied to all instances from the same view.

As pointed out by Bottou and Vapnik [4], it is usually not easy to find a single function which holds good predictability for the entire data space, but it is much easier to seek some functions that are capable of producing good predictions on some specified regions. For many computer vision applications in particular, it has been demonstrated that the idea of local learning is very useful, e.g., [5, 6, 7]. To achieve

locality and nonlinearity simultaneously, we propose in this paper a novel algorithm called Instance-Specific Canonical Correlation Analysis (ISCCA), which approximates the nonlinear data by computing the instance specific projections along the smooth curve of the manifold. Based on the equivalent solution of least square regression with CCA, we extend CCA to the instance-specific case. Specially, each instance has its own specific transformation and all transformations are locally-linear smooth but globally-nonlinear. Hence the learned model is more robust and flexible. Our method can be formulated as a convex optimization problem and solved efficiently via alternating optimization procedure. Irrespective of initial values, the globally optimal solutions could be achieved with theoretical guarantee.

The rest of this paper is organized as follows. Section 2 introduces some preliminary work. In Section 3, we detail the proposed instance-specific CCA method. Experimental results are given in Section 4. Section 5 concludes the paper.

2. PRELIMINARY WORK

Let us represent two data sets $\mathcal{X}$ and $\mathcal{Y}$ with their matrix forms $\mathbf{X} \in \mathbb{R}^{d_x \times l}$ and $\mathbf{Y} \in \mathbb{R}^{d_y \times l}$, where l is the number of training data and each column vector $\mathbf{x}_i$ ($\mathbf{y}_j$) in $\mathbf{X}$ ($\mathbf{Y}$) denotes an instance with d_x (d_y) dimensions. For the case of reduction to d dimension, CCA computes two projection matrices $\mathbf{w}_x \in \mathbb{R}^{d_x \times d}$ and $\mathbf{w}_y \in \mathbb{R}^{d_y \times d}$ to maximize the correlation coefficient ρ between the two data sets:

$$\rho = \max_{\mathbf{W}_x, \mathbf{W}_y} \frac{\mathrm{tr}\left(\mathbf{W}_x^T \mathbf{C}_{xy} \mathbf{W}_y\right)}{\sqrt{\mathrm{tr}\left(\mathbf{W}_x^T \mathbf{C}_{xx} \mathbf{W}_x \mathbf{W}_y^T \mathbf{C}_{yy} \mathbf{W}_y\right)}}, \tag{1}$$

where $\mathrm{tr}(\cdot)$ is the trace operator, $\mathbf{C}_{xx}$ and $\mathbf{C}_{yy}$ denote the sample covariance matrices of each view and $\mathbf{C}_{xy}$ the sample covariance between the two views. $\mathbf{W}_x$ and $\mathbf{W}_y$ can be finally obtained by solving a generalized eigenvalue problem: $\mathbf{C}_{xy}\mathbf{C}_{yy}^{-1}\mathbf{C}_{yx}\mathbf{W}_x = \mathbf{C}_{xx}\mathbf{W}_x\mathbf{\Lambda}^2$,where $\mathbf{\Lambda}$ is a diagonal matrix containing different eigenvalues as diagonal entries.

Maximizing the canonical correlation in Eq. (1) is also equivalent to minimizing the distance between the two views:

$$\begin{aligned} &\min_{\mathbf{W}_x, \mathbf{W}_y} \sum_{i=1}^{l} \left\| \mathbf{W}_x^T \mathbf{x}_i - \mathbf{W}_y^T \mathbf{y}_i \right\|_2^2 = \left\| \mathbf{W}_x^T \mathbf{X} - \mathbf{W}_y^T \mathbf{Y} \right\|_F^2 \\ &\text{s.t. } \mathbf{W}_x^T \mathbf{X}\mathbf{X}^T \mathbf{W}_x = \mathbf{I}, \mathbf{W}_y^T \mathbf{Y}\mathbf{Y}^T \mathbf{W}_y = \mathbf{I}, \end{aligned} \tag{2}$$

where $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the ℓ_2 vector norm and Frobenius matrix norm, respectively.

In the literature, the relationship between CCA and least squares regression (LSR) has been investigated [8]. Suppose we are given a training set $\{(\mathbf{x}_i, \mathbf{h}_{x_i})\}_{i=1}^{l}$ where $\mathbf{x}_i$ is an input and $\mathbf{h}_{x_i}$ is the corresponding target, LSR can be expressed as:

$$\min_{\mathbf{W}_x} \sum_{i=1}^{l} \frac{1}{2} \left\| \mathbf{W}_x^T \mathbf{x}_i - \mathbf{h}_{x_i} \right\|_2^2 = \frac{1}{2} \left\| \mathbf{W}_x^T \mathbf{X} - \mathbf{H}_x \right\|_F^2,$$

where $\mathbf{H}_x = \{\mathbf{h}_{x_1}, \cdots, \mathbf{h}_{x_l}\}$. As shown in Sun *et al.*[8], if the target matrix $\mathbf{H}_x$ for CCA is defined as $\mathbf{H}_x = \mathbf{C}_{yy}^{-\frac{1}{2}}\mathbf{Y}$, LSR has a equivalence solution with that of CCA.

3. INSTANCE-SPECIFIC CANONICAL CORRELATION ANALYSIS

3.1. Basic Formulation

Based on the relationship between CCA and LSR, we propose ISCCA in the framework of regression to compute instance-specific projection, which essentially induce different local projection functions at different locations of each view space. In devising the learning algorithm, two important criteria are taken into consideration: (1) The overall projection matrices of ISCCA should be globally consistent with that of CCA such that the correlation between data from two views is maximized; (2) The instance-specific transformations should not differ significantly in a local region and should vary smoothly over the entire input space of each view.

Inspired by graph-based propagation technique [7], we define a regularization term Ω to enforce the smoothness of instance-specific projection matrix $\mathbf{W}_i$:

$$\Omega(\widetilde{\mathbf{W}}) = \sum_{i,j=1}^{l} \vartheta_{ij} \|\mathbf{W}_i - \mathbf{W}_j\|_F^2 = \mathrm{tr}(\widetilde{\mathbf{W}}\mathbf{L}\widetilde{\mathbf{W}}^T),$$

where $\widetilde{\mathbf{W}} = [\mathrm{vec}(\mathbf{W}_1), \cdots, \mathrm{vec}(\mathbf{W}_l)]$ and $\mathrm{vec}(\cdot)$ denotes the operator to convert a matrix into a vector in a columnwise manner. $\mathbf{L}$ is the graph Laplacian matrix [9] of the similarity graph. Here, ϑ_{ij} is the pairwise weight to reflect the similarity between instances $\mathbf{x}_i$ and $\mathbf{x}_j$. In our experiments, we use the local scaling method [10] to define the similarity graph as:

$$\vartheta_{ij} = \begin{cases} \exp(\frac{-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2}{\sigma_i \sigma_j}) & \text{if } \mathbf{x}_i \in \mathcal{N}_k(\mathbf{x}_j) \text{ or } \mathbf{x}_j \in \mathcal{N}_k(\mathbf{x}_i) \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

where $\mathcal{N}_k(\mathbf{x}_i)$ denotes the set of k nearest neighbors of $\mathbf{x}_i$, σ_i (σ_j) is the distance from $\mathbf{x}_i$ ($\mathbf{x}_j$) to its m-th nearest neighbor, and m is a local scale factor. By exploiting the regularization term, the instance-specific transformations given by the solution will vary smoothly over the data graph.

For finding the projection matrices $\{\mathbf{W}_{x_i}\}_{i=1}^{l}$, we formulate the optimization procedure of ISCCA as the following minimization problem:

$$\min_{\{\mathbf{W}_{x_i}\}} \sum_{i=1}^{l} \left(\frac{1}{2} \left\| \mathbf{W}_{x_i}^T \mathbf{x}_i - \mathbf{h}_i \right\|_2^2 + \frac{\mu}{2} \left\| \mathbf{W}_{x_i} \right\|_F^2 \right) + \frac{\rho}{2} \mathrm{tr}(\widetilde{\mathbf{W}}_x \mathbf{L}_x \widetilde{\mathbf{W}}_x^T), \tag{4}$$

where $\mathbf{W}_{x_i}$ is the projection matrix for the ith data point $\mathbf{x}_i$ in $\mathcal{X}$-view, μ and ρ are the regularization parameters, $\widetilde{\mathbf{W}}_x = [\mathrm{vec}(\mathbf{W}_{x_1}), \cdots, \mathrm{vec}(\mathbf{W}_{x_l})]$, and $\mathbf{L}_x$ is the graph Laplacian matrix defined from the perspective of the $\mathcal{X}$-view. The first term in Eq (4) measures the empirical loss from the least

squares regression perspective for CCA, the second term penalizes the complexity of each projection matrix, and the last term serves to preserve the smoothness of the projections.

After obtaining $\{\mathbf{W}_{x_i}\}$ for $\mathcal{X}$-view, we can compute the projections matrices $\{\mathbf{W}_{y_i}\}_{i=1}^{l}$ based on Eq. (2):

$$\min_{\{\mathbf{W}_{y_i}\}} \sum_{i=1}^{l} \left(\frac{1}{2} \left\| \mathbf{W}_{y_i}^T \mathbf{y}_i - \mathbf{W}_{x_i}^T \mathbf{x}_i \right\|_2^2 + \frac{\mu}{2} \left\| \mathbf{W}_{y_i} \right\|_F^2 \right) + \frac{\rho}{2} \mathrm{tr}(\widetilde{\mathbf{W}}_y \mathbf{L}_y \widetilde{\mathbf{W}}_y^T),$$

where the target vectors of the regression are defined as $\mathbf{H}_y = \mathbf{W}_{x_i}^T \mathbf{x}_i$, according to the optimal solutions of $\{\mathbf{W}_{x_i}\}$.

Since ISCCA learns the projection matrices in a local manner, error estimation is also local which make the model more robust and flexible. Moreover, graph Laplacian regularization term makes the projections vary smoothly and hence the local geometric structure can also be well preserved after dimension reduction. Therefore, the locally linear smooth learning essentially induce globally nonlinear property.

3.2. Optimization Problem

Without loss of generality, we will take $\mathcal{X}$-view as an example to illustrate how to derive the optimal solutions for ISCCA. The derivations for $\mathcal{Y}$-view can be obtained in a similar way.

Theorem 1 *The objective function in Eq. (4) is jointly convex with respect to all variables.*

Due to space limitation, the proof of Theorem 1 is omitted.

Even though the objective function in Eq. (4) is jointly convex, optimizing with respect to all $\mathbf{W}_{x_i}$'s simultaneously is computationally challenging. Hence we employ an alternating procedure to solve the problem more effectively. Specifically, we sequentially solve a subproblem for $\mathbf{W}_{x_i}$ by keeping all other $\mathbf{W}_{x_j}$'s ($j \neq i$) fixed. This procedure is repeated until convergence or after a maximum number of iterations T has been reached. Since the optimization problem is convex with respect to all variables, the solution found by this alternating procedure is guaranteed to be globally optimal [11]. In what follows, we will present the subproblem and its optimal solution.

Optimizing with respect to $\mathbf{W}_{x_i}$ when $\mathbf{W}_{x_j}$'s ($j \neq i$) are fixed. Let $\mathbf{W}_{x_{-i}} \overset{\text{def}}{=} [\mathbf{W}_{x_1}, \cdots, \mathbf{W}_{x_{i-1}}, \mathbf{W}_{x_{i+1}}, \cdots, \mathbf{W}_{x_l}]$, which denotes a constant parameter matrix here. First we decompose $\widetilde{\mathbf{W}}_x$ and $\mathbf{L}_x$ as $\widetilde{\mathbf{W}}_x = (vec(\mathbf{W}_{x_i}), \widetilde{\mathbf{W}}_{x_{-i}})$ and $\mathbf{L}_x = \begin{bmatrix} l_{ii} & \mathbf{l}_i^T \\ \mathbf{l}_i & \Lambda_{-i} \end{bmatrix}$, where l_{ij} is the (i,j)th element of $\mathbf{L}$, $\mathbf{l}_i$ is a column vector consisting of the elements in the ith column of $\mathbf{L}$ except l_{ii}, and Λ_{-i} is a submatrix deleting the ith row and the ith column. Then the third term in the objective function can be written as:

$$\begin{aligned} &\mathrm{tr}(\widetilde{\mathbf{W}}_x \mathbf{L}_x \widetilde{\mathbf{W}}_x^T) \\ &= \mathrm{tr}\left((vec(\mathbf{W}_{x_i}), \widetilde{\mathbf{W}}_{x_{-i}}) \begin{bmatrix} l_{ii} & \mathbf{l}_i^T \\ \mathbf{l}_i & \Lambda_{-i} \end{bmatrix} \begin{bmatrix} vec(\mathbf{W}_{x_i})^T \\ \widetilde{\mathbf{W}}_{x_{-i}}^T \end{bmatrix} \right) \\ &= l_{ii} \left\| \mathbf{W}_{x_i} \right\|_F^2 + 2\mathrm{tr}(\mathbf{M}_i \mathbf{W}_{x_i}) + \mathrm{tr}(\widetilde{\mathbf{W}}_{x_{-i}} \Lambda_{-i} \widetilde{\mathbf{W}}_{x_{-i}}^T), \end{aligned}$$

where $\mathbf{M}_i$ is a matrix satisfing $\mathrm{vec}(\mathbf{M}_i) = \mathbf{W}_{x_{-i}} \mathbf{l}_i$. Note that the last term in the above equation is unconcerned with $\mathbf{W}_{x_i}$. Hence the optimization problem *w.r.t.* $\mathbf{W}_{x_i}$ becomes:

$$\begin{aligned} \min_{\{\mathbf{W}_{x_i}\}} \sum_{i=1}^{l} \frac{1}{2} \left\| \mathbf{W}_{x_i}^T \mathbf{x}_i - \mathbf{h}_i \right\|_2^2 + \sum_{i=1}^{n} \frac{\mu}{2} \left\| \mathbf{W}_{x_i} \right\|_F^2 \\ + \frac{\rho}{2} \left(l_{ii} \left\| \mathbf{W}_{x_i} \right\|_F^2 + 2\mathrm{tr}(\mathbf{M}_i \mathbf{W}_{x_i}) \right). \end{aligned} \tag{5}$$

Then we compute the gradient of the objective function J_x with respect to $\mathbf{W}_{x_i}$ and set it to zero, the optimal $\mathbf{W}_{x_i}$ in current subproblem could be obtained in an analytical form:

$$\mathbf{W}_{x_i} = \left((\mu + \rho l_{ii}) \mathbf{I} + \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \left(\mathbf{x}_i \mathbf{h}_i^T - \rho \mathbf{M}_i \right).$$

3.3. Out-of-Sample Extension

The learning described above is only for training data. In the following, we extend the proposed algorithm to handle a new test sample $\mathbf{x}_t$ by optimizing the following objective function:

$$J_{x_t} = \frac{1}{2} \sum_{i=1}^{l} \theta_{it} \left\| \mathbf{W}_{x_t}^T \mathbf{x}_i - \mathbf{h}_i \right\|_2^2 + \frac{\mu}{2} \left\| \mathbf{W}_{x_t} \right\|_F^2, \tag{6}$$

where θ_{it} is the similarity between instances $\mathbf{x}_i$ and $\mathbf{x}_t$ as defined in Eq. (3). The first term in the objective function is moving least squares [12], which has been demonstrated to work very well for interpolation problems in that the learned transformations vary smoothly. As a result, the optimal projection $\mathbf{W}_{x_t}$ can be found by projecting itself onto the training data and minimizing the weighted reconstruction error.

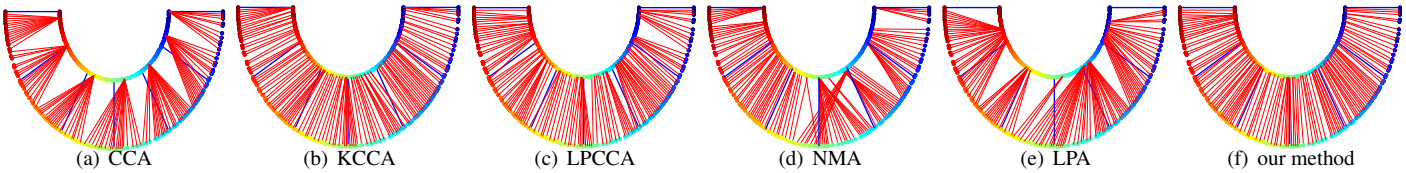
Taking the derivative of J_{x_t} with respect to $\mathbf{W}_{x_t}$ and setting it to zero, $\mathbf{W}_{x_t}$ can be computed in closed-form as:

$$\mathbf{W}_{x_t} = \Big(\sum_{i=1}^{l} \theta_{it} \mathbf{x}_i \mathbf{x}_i^T + \mu \mathbf{I} \Big)^{-1} \Big(\sum_{i=1}^{l} \theta_{it} \mathbf{x}_i \mathbf{h}_i^T \Big).$$

4. EXPERIMENTAL STUDY

In this section, we conduct empirical studies on a typical computer vision application: pose alignment on FACE-10 dataset. Our algorithm is compared with existing pose alignment methods: CCA [3], KCCA [1], LPCCA [2], Nonlinear Manifold Alignment (NMA) [13], and Linear Procrustes Analysis (LPA) [14]. In experiments, we attempt to align the images with the same pose from different persons for pose estimation. FACE-10 dataset consists of 1011 images of 10 persons taken under normal indoor lighting conditions and fixed background with a Sony EVI-D31 camera. Two sample image sequences are shown in Figure 1(a). The poses are almost continuous, from -90° to $+90^\circ$ as shown in Figure 1(b).Each image contains 32×32 pixels with the image intensity values as features. In our experiments, $m = 2, k = 6$ and the regularization parameters are determined by cross-validation.

To illustrate the alignment results in an intuitionists way, we inlay two aligned sequences on two half concentric circles. As indicated in Figure 2, the 'blue bold' lines connect the



(a) CCA (b) KCCA (c) LPCCA (d) NMA (e) LPA (f) our method

Fig. 2. The results of aligning 'person 1' and 'person 2' inlayed on two half concentric circles when labeled the data number is 7 and evenly distributed. (a) CCA, (b) KCCA, (c) LPCCA, (d)NMA, (e) LPA and (f) our method.

labeled correspondence pairs, while the 'red' lines denote the connections for aligned test points. In perfect alignment, the connections should be along the direction of radius. Under this criterion, it can be seen that our method has no line cross and the lines connecting test data are more regular compared with other methods.

Moreover, the quantitative comparison of the pose estimation is summarized in Table 1. We can find that CCA give poor results as a linear and global method. KCCA, as a nonlinear method, significantly decreases the error rates compared to CCA. LPCCA also works better than CCA by further considering local geometric information. NMA gives relatively good results to handle new test data. However, since no explicit mapping is learned, NMA need retraining for out-of-sample which is time consuming. And due to the the restriction in LPA, it cannot give satisfactory results in this situation since the relationship between the two manifolds is beyond the affine transformation. Benefiting from the jointly modeling ability of locality and nonlinearity, ISCCA consistently give the lowest error rates.

Table 1. Pose estimation errors (unit: degree) on FACE-10 dataset (mean±std-dev)

Method	♯ labels = 4	♯ labels = 7	♯ labels = 10	♯ labels = 13
CCA	38.67±19.88	32.36±16.73	27.44±14.74	27.16±15.08
KCCA	10.61±4.47	5.87±1.27	2.90±0.55	2.60±0.88
LPCCA	11.92±8.45	5.59±2.56	2.94±1.66	2.88±1.92
NMA	12.11±0.78	9.08±0.41	5.14±0.76	3.37±0.23
LPA	28.12±12.10	28.56±13.80	28.41±14.77	28.55±15.43
ISCCA	**6.34±2.99**	**2.01±0.39**	**1.70±0.22**	**1.22±0.16**

We further evaluate the iterative convergence for alternative optimization procedure. As depicted in Figure 3, we can observe that: the residual error value rapidly decreases at the first few iterations; and the residual error value becomes stable in less than 20 times. These confirm the efficiency of alternating procedure to solve the proposed optimization problem.

5. CONCLUSION

We presented a novel method to model the properties of locality and nonlinearity simultaneously for multi-view statistical correlation learning. Our method extends CCA to instance-specific case, which obtains a set of locally-linear smooth but globally-nonlinear transformations. Experimental results for pose alignment verifies the effectiveness of proposed method.

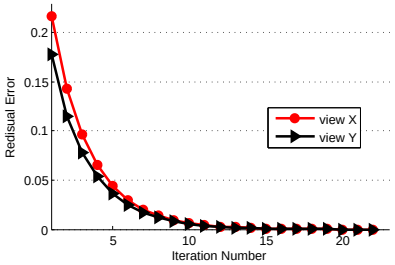


Fig. 3. Iterative convergence rate on FACE-10 dataset

Acknowledgement: This work is partially supported by the National Science Foundation of China under contract No. 61272319, the National Basic Research Program of China (973 Program) under contract 2009CB320900, Beijing Natural Science Foundation (New Technologies and Methods in Intelligent Video Surveillance for Public Security) under contract No. 4111003, and the new Ph.D researcher award of Chinese Ministry of Education.

6. REFERENCES

[1] T. Melzer, M. Reiter, and H. Bischof, Appearance models based on kernel canonical correlation analysis, Pattern Recognition, 2003.
[2] T. Sun and S. Chen, Locality preserving CCA with applications to data visualization and pose estimation, Image and Vision Computing, 2007.
[3] H. Hotelling, Relations between two sets of variates, Biometrics.
[4] L. Bottou and V. Vapnik, Local learning algorithms, Neural Computation, vol. 4, no. 6, 1992.
[5] M.Wu and B. Scholkopf, A local learning approach for clustering, in Advances in Neural Information Processing Systems, 2006.
[6] M. Wu and B. Scholkopf, Transductive classification via local learning regularization, in Artificial Intelligence and Statistics, 2007.
[7] D. C. Zhan, M. Li, Y.-F. Li, and Z.H. Zhou, Learning instance specific distances using metric propagation, in ICML, 2009.
[8] L. Sun, S. Ji, and J. Ye, A least squares formulation for canonical correlation analysis, in ICML, 2008.
[9] F. R. K. Chung, Spectral Graph Theory, American Mathematical Society, Rhode Island, 1997.
[10] L. Zelnik-Manor and P. Perona, Self-tuning spectral clustering, in Advances in Neural Information Processing Systems 17. 2005.
[11] D. Bertsekas, Nonlinear programming, Athena Scientific, 1999.
[12] D. Levin, The approximation power of moving least squares, Mathematics of Computation, vol. 67, no. 224, 1998.
[13] J. Ham, D. Lee, and L. Saul, Semisupervised alignment of manifolds, in Artificial Intelligence and Statistics, 2005.
[14] C. Wang and S. Mahadevan, Manifold alignment using procrustes analysis, in ICML, 2008.