

Hybrid Subspace Fusion for Microcalcification Clusters Detection^{*}

Xinsheng Zhang^{*}, Hongyan He, Naining Cao, Zhengshan Luo

School of Management, Xi'an University of Architecture and Technology, Xi'an 710055, China

Abstract

Early detection of breast cancer, a significant public health problem in the world, is the key for improving breast cancer early prognosis. Mammography is considered the most reliable and widely used diagnostic technique for early detection of breast cancer. However, it is difficult for radiologists to perform both accurate and uniform evaluation for the enormous mammograms with widespread screening. Microcalcification clusters is one of the most important clue of the breast cancer, and their automated detection is very helpful for early breast cancer diagnosis. Because of the poor quality of the mammographic images and the small size of the microcalcifications, it is a very difficult task to perform detecting the early breast cancer. In this paper, we propose a novel approach based on hybrid subspace fusion for detection microcalcification clusters, and successfully apply it to detection task in digital mammograms. In such a system, subspace learning algorithms will be selectively fused according to the ability of preserving the classification information. Experimental results show that the proposed method improved the performance and stability of microcalcification cluster detection and could be adapt to the noise environments better. The proposed methods could get satisfactory results on sensitivity and reduce false positive rate, which provide some new ideas and methods for the research and development of computer-aided detection system in the breast cancer detection community.

Keywords: Subspace Learning; Data Fusion; Micalcification Cluster; Support Vector Machine; Digital Mammograms

1 Introduction

In digital mammograms, an important sign of the early breast cancer is the existence of Microcalcification Clusters (MCs). One of the key techniques for early diagnosis of the breast cancer is to detect MCs and to judge whether they are malignant or not in mammograms. However, there is only about 3% information in mammograms, which can be seen with the naked eyes. Due to the most details in mammograms cannot be perceived by human eyes, it is even very

^{*}Project supported by the National Natural Science Foundation (No. 61271278) and the Key Discipline of Shaanxi Province (No. E08001).

^{*}Corresponding author.

Email address: xinsheng.zh@gmail.com (Xinsheng Zhang).

difficult for a skillful radiologist to find the sign of early breast cancer, i.e., microcalcification clusters, as a result missing the best time for treatment. To detect early sign of this disease and we need to do research on how to build a computer-aided diagnosis system to aid doctors to diagnose breast cancer in early stage. Machine learning techniques have been successfully applied to it [1–5].

Most of these methods learn hypotheses from a large amount of diagnosed mammogram samples, i.e., the data collected from a number of necessary medical examinations with the corresponding diagnoses made by the doctors, to aid the other doctors to do a diagnosis in the future. Because of the big dimensionality of the image samples, there always are an overfitting problem caused the curse of dimensionality in those machine learning methods [6].

This curse of dimensionality is compounded by extra complexity introduced by data set heterogeneity. Heterogeneous data sets usually require more sample data points than homogeneous data sets for machine learning performance. Combining heterogeneous data types for classification is a difficult machine learning problem, but one that has shown promise in bio-informatics applications. Data fusion is the study of optimal information processing in heterogeneous data environments through intelligent integration of heterogeneous data. The development of data fusion was motivated chiefly by the increasing demand for more accurate information, more practical and robust procedures to manage data efficiently, and improved system reliability and performance.

Decision fusion has the ability to capture general trends in data while remaining robust to the effects of noise. As the noise and complexity of a data set increase, it becomes easier for a classifier model to overfit the data rather than capture the underlying data trends. Currently, researchers have proposed a lot of microcalcification detection methods based on ensemble learning. As Li *et al.* [7] proposed a Co-Forest-based ensemble learning method to detect microcalcifications when labeled samples is limited. Jesneck *et al.* [8] proposed an optimized breast cancer detection method based on heterogeneous multi-source information fusion to improve diagnostic performance, which can combine heterogeneous information, such as from different modalities, feature categories, or institutions. Experimental results showed that the decision fusion directly optimized clinically significant performance measures, such as AUC and pAUC, and sometimes outperformed two well known machine-learning techniques. But a single type of image features was not always a good choice to improve the performance of detecting microcalcification clusters. To improve the performance of detection system, Massimo D. S. *et al.* [9] proposed a microcalcification cluster detection algorithm based on information fusion with a multiple expert system. In this system, there are several experts, some of which are devoted to classify the single microcalcifications and others are aimed to classify the cluster considered as a whole. Then the knowledge of multiple experts are fused together to make a final overall decision.

Methods discussed above have achieved some good results by fusing different learning methods, but most of these methods are carried out the classification task based on the certain fixed characteristics or domain knowledge, without considering how to get these features or dynamic domain knowledge. To take advantage of decision fusions strengths and to keep more information, based on results of previous research, subspace learning and ensemble learning of the latest research study, we pursued a data fusion approach that was a hybrid of the feature fusion and decision fusion techniques to perform the MCs detection task. We first extracted features from the raw data (carefully selected sample database). We applied these features to a local classifier and thresholds to these classifiers outputs, creating local binary decisions. We then combined these

decisions with decision fusion theory. And finally we applied the method to detect microcalcification clusters, mainly aimed to solve: 1) How to select multiple hybrid subspace features; 2) How to select the level of fusing (fusion feature subspace level or fusion decision subspace level); 3) How to fusion the subspace after careful selection.

2 Data Fusion

Since the 90's, ensemble learning or data fusion has become to important research in the field of machine learning theory, and has become a powerful tool in dealing with problems of pattern recognition community. Ensemble methods use multiple learning algorithms to obtain better predictive performance than could be obtained from any of the constituent learning algorithms. Unlike a statistical ensemble in statistical mechanics, which is usually infinite, a machine learning ensemble refers only to a concrete finite set of alternative models, but typically allows for much more flexible structure to exist between those alternatives.

Existing studies about data fusion or ensemble learning include a variety of algorithms, such as Bagging and Boosting *et al.* In general, we can get a better prediction classifier with a fusion method than with any single classifier. Furthermore, data fusion or ensemble learning can avoid overfitting phenomena when learning. One of the fundamental problems in machine learning is to improve the generalization ability of learning systems, that is, how to use existing data to train a learner as well as possible so that it can easily handle the new data.

Traditionally, researchers divide the level of fusion techniques into three levels [10]:

- 1) Data-level fusion: to fuse/ensemble the raw input source together;
- 2) Feature-level fusion: to fuse/ensemble the extracted features from the original input source;
- 3) Decision-level fusion: to fuse/ensemble results of every single decision or classifiers together.

As discussed above, there are three levels of fusion architecture of for the final decision. For the data-level fusion, it will fuse different mortality original data or signal together. And in the feature-level fusion, it will fuse features from different feature space together. While in the decision-level fusion, each single learner or classifier will make a decision according the original image data or extracted features and then get the final overall decision by fusing the single obtained decisions.

In order to improve the overall performance of the hybrid fusing system, we will focus on how to fuse different subspace (hybrid subspace features) and decision together. Based on this, we propose a new fusing method using hybrid selective feature subspaces to make classification decision.

For a learning task, we can train a group of learners. Theoretically, it is better for us to choose from a group of base learners when available to fuse than to select all base learners to fuse or ensemble [11, 12]. This is the meaning of “Many Could Be Better Than All”, which provides the theoretical basis for the selective fusion. The main idea of selective fusion is to select some of base learners according to some policy to get better results. Predictive learning task can be divided into regression with an output of continuous value and classification with an output of discrete value. In this paper, we will choose base learners with continuous output, and employ the weighted average method in the final result fusion.

3 Selective Fusion Based on Hybrid Subspace Learning

3.1 Idea of the Algorithm

At present, there are many methods for subspace learning. Subspace learning is a kind of statistical analysis method. Currently, subspace learning generally are divided into linear and nonlinear methods. The traditional linear subspace learning methods are mainly included PCA (Principal Components Analysis) [13], LDA (Linear Discriminate Analysis) [14], factor analysis [15], projection pursuit (Projection Pursuit) [16], etc. Nonlinear subspace learning methods can be divided into two categories, one is based on the global nonlinear subspace learning methods, including MDS (Multidimensional Scaling) [17], SPE (Stochastic Proximity Embedding) [18], Isomap [19], FastMVU (Fast Maximum Variance Unfolding) [20], LPP (Linearity Preserving Projection) [21], NPE (Neighborhood Preserving Embedding) [22], LLTSA (Linear LTSA) [23], etc; methods for global alignment of the linear model based mainly LLC (Locally Linear Coordination) [24], Manifold charting [25], CFA (Coordinated Factor Analysis) [26] and so on.

On the same issue different feature extraction methods using subspace learning may get different subspace features. For classification problems, there is no method with subspace learning can achieve good effect on any kind of data, because there are often noises in the real data, and subspace learning algorithms are always sensitive to noises, which leads difficult to set learning parameters, and thus will greatly affect the accuracy of subspace learning. Hybrid subspace learning aims at achieving higher classification accuracy rate than any single feature subspace method by ensemble or fusing hybrid subspace features to overcome the limitations of single feature subspace and then achieve better adaptability. At present, most of subspace learning algorithms fusion feature vectors from different subspace together to improve the classification ability of the feature vectors. When combining all of them as a vector, because of the correlation, discriminant information may be lost, although every single feature vector in its corresponding subspace has a good classification ability. Therefore, the method of direct feature fusion in multi-subspace, to a certain extent, may add computing complexity to the classification algorithm but good classification ability.

In fact, a pattern in one feature subspace of are easily to recognize, but it would be difficult to recognize after the pattern is projected onto another subspace. That is the original easily recognized pattern may be difficult to recognize, and that difficult to recognize may turn out to be identifiable. Therefore, based on the characteristic of the subspace learning algorithm and the ability to discriminate information in original subspace, we could selectively fuse multiple feature subspaces together to improve the discriminant performance of the learning system.

3.2 Description of the Algorithm

In order to take full advantage of the selective advantages of and subspace learning algorithm, we designed a hybrid subspace selective fusion algorithm (Hybrid Subspace Selective Fusion, HSSF) to perform MCs detection task (shown in Fig. 1).

As shown in Fig. 1, there are two different levels of selective fusion in Stage 3. Given different levels of fusion, we could get different algorithms. In this paper, we only consider the feature subspace fusion level, and thus design a new selective fusion method with subspace learning (shown in Algorithm 1). In each of the corresponding subspace, we can train these corresponding base

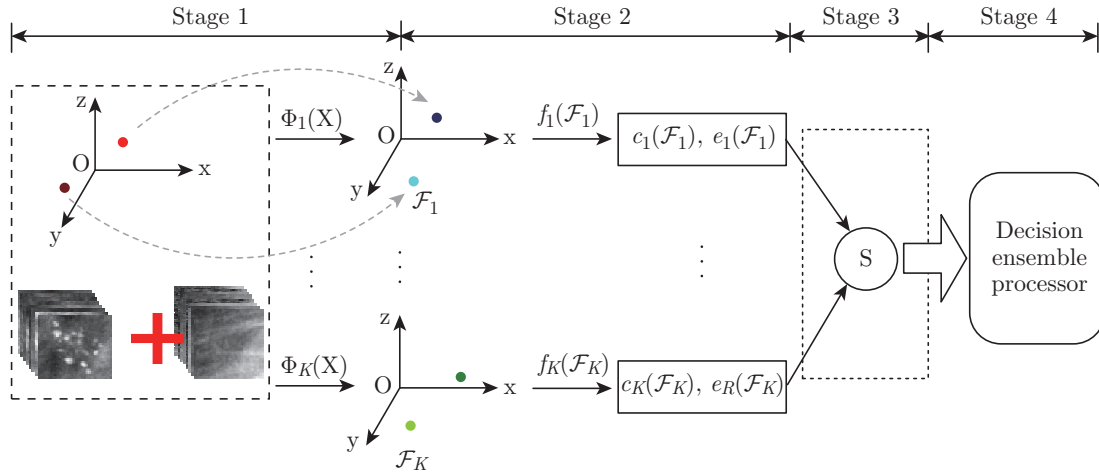


Fig. 1: Framework of selective hybrid subspace fusion method

learners $f_k(\cdot)$ and then fuse them with these single learners to get a better performance classifier. Each baser learner is set a weight τ'_i when selective fusion each subspace. Since every subspace is first obtained by randomly selecting the training sample and training base classifier with randomly selected training sample, we believe that the resulting base classifier is different from each other (also known as diversity), and independent of others. Therefore, based on this premise, well-behaved subspace learning methods can be selectively integrated into the final corresponding fusion classifier model.

4 Experimental Results and Analysis

4.1 Experimental Data

Among the algorithm performance evaluation, in order to avoid fitting problems, we carry our experiments by the procedure of training, validation and retesting. Sample sets used in the experiments are divided into training, verification and test samples, where the test samples are not all involved in the training process. After getting the trained fusion model, a test will be performed on this fusion model to test its efficiency and performance with the test samples. This algorithm uses the methods described in [3] to generate data sets. As described in Table 1, two training sets were randomly selected from the sample database in the experiment, one (S^1) used to train learning algorithm for subspace and the other (S^2) used to train a classifier (for training and validation). In order to test the fusion algorithm performance, we select another sample set (S^3) as a test set to test the final fusion model.

Table 1: Dataset distribution used in experiments

	Training dataset 1 (S^1)	Training dataset 2 (S^2)	Test dataset 3 (S^3)
Positive samples	1000	500	500
Negative samples	1000	500	500

Algorithm 1 Hybrid subspace selective fusion method (HSSF)

Input: Training data set with N samples: S_t^1 ;
 Training data set with M samples: S_t^2 ;
 Subspace learning algorithm set: $\Sigma = T_1, \dots, T_K$;
 Trained projection function set: $\Xi = \Phi_1(\cdot), \dots, \Phi_K(\cdot)$;
 Base learners set: $f(\cdot)$;
 Accuracy threshold: T .

Output: Model of trained fusion classifier ENS_{HSSF}

- 1: **for** $k = 1$ to K **do**
- 2: Train subspace learning algorithm T_k with dataset S_t^1 ;
- 3: Get the projection function $\Phi_k(\cdot)$ and its parameters;
- 4: **end for** // End stage 1
- 5: $\Psi \leftarrow \{\}$; // Set of subspace learning algorithms, corresponding base learners and related parameters
- 6: **for** $k = 1$ to K **do**
- 7: Project S_t^2 to get $S_{t,k}^{2'}$ and reconstruction error ζ_k using projection function $\Phi_k(\cdot)$;
- 8: Train base learner $f(\cdot)$ using 50% of samples from $S_{t,k}^{2'}$ to obtain the trained base learner $f_k(\cdot)$;
- 9: Get the generalization error ξ_k of $f_k(\cdot)$ using samples from the other 50% of $S_{t,k}^{2'}$;
- 10: $\mathfrak{S} \leftarrow \{f_k(\cdot), \xi_k, \Phi_k(\cdot), \zeta_k\}$;
- 11: Add \mathfrak{S} to Ψ ;
- 12: **end for** // End stage 2
- 13: $\Psi' \leftarrow \{\}$; // Set of selected subspace learning algorithms, corresponding base learners and related parameters
- 14: **for** $k = 1$ to K **do**
- 15: $\tau_k \leftarrow \log(1/(e^{-\xi_k} + e^{-\zeta_k}))$;
- 16: $\mathfrak{S}' \leftarrow \{f_k(\cdot), \xi_k, \Phi_k(\cdot), \zeta_k, \tau_k\}$;
- 17: **if** $\tau_k \geq T$ **then**
- 18: Add \mathfrak{S}' to Ψ' ;
- 19: **end if**
- 20: **end for** // End stage 3 (selective decision-level fusion)
- 21: $ENS_{HSSF} \leftarrow \Psi'$;
- 22: $L \leftarrow |\Psi'|$;
- 23: Given any sample X and ENS_{HSSF} ; // / Stage 4: test the final result of trained fusion classifier
- 24: **for** $J = 1$ to L **do**
- 25: Project X onto subspace using $\mathfrak{S}_j(j = 1, \dots, L)$ from Φ' and get $F_j(j = 1, \dots, L)$;
- 26: Feed F_j to the trained base learner $f_j(j = 1, \dots, L)$ and obtain the prediction $h_j(j = 1, \dots, L)$;
- 27: Compute the final overall prediction $h^* = \sum_{j=1}^L \tau_j' h_j$, where τ_j' is the normalized result of τ_j .
- 28: **end for**

4.2 Experimental Parameter Configuration

In order to verify the effectiveness of the algorithm in this paper, 25 subspace learning algorithms are selected as a set of basic hybrid subspace learning algorithms. In the training procedure

of subspace learning algorithm, parameter settings for each algorithm are settled according to empirical values and samples from training dataset 1 S^1 (shown in Table 1) are used to train and test subspace learning algorithm. Support Vector Machines (SVMs) with RBF kernel are employed to train and valid the learned algorithms with training dataset 2 S^2 (shown in Table 1), where $\sigma = 15, C = 1000$. When we test the fusion classifier, the same base classifier and parameter settings will be selected to perform the testing procedure with Test dataset 3 (S^3) as shown in Table 1.

4.3 Experimental Results

We first test the performance of the proposed algorithm and select those subspaces where τ_k is greater than a threshold as the candidate subspaces of the next step according to the discriminant ability (measured by the SVM classifier) after mapping onto them. As shown in Table 2, four indicators to evaluate and choose based on a single integrated algorithm subspace subspace-based hybrid performance: accuracy (Accuracy), sensitivity (Sensitivity / True Positive Rate), specificity (Specificity / 1- The area under the FPR), ROC curve (Area Under ROC Curve, Az). From Table 2, the selective integration of learning subspace method based on hybrid four indicators are better than a single subspace-based learning approach.

Table 2: Performance evaluation of MCs detection with the selected single subspace learning methods and the ensemble learning algorithm

Selected Methods	Accuracy	Sensitivity	Specificity	Az
PCA	0.8786	0.8786	0.8786	0.9355
Prob PCA	0.8393	0.8369	0.8393	0.8907
GDA	0.7589	0.8112	0.7664	0.8269
DiffusionMaps	0.6982	0.6637	0.7074	0.7931
LPP	0.8214	0.8169	0.8215	0.8884
NPE	0.6821	0.6449	0.695	0.797
LLTSA	0.6839	0.702	0.6854	0.8022
HSSF	0.9286	0.9444	0.9291	0.9674

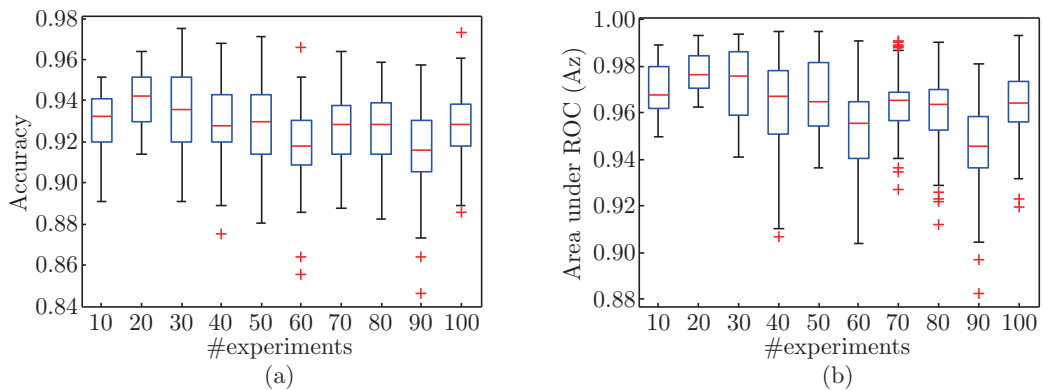


Fig. 2: Boxplot of the evaluation result with algorithm 1 (HSSF): (a) Accuracy; (b) Area Under ROC Curve (Az)

Furthermore, in order to test the stability of the proposed algorithm, we train and test this algorithm respectively, and in each round the training procedure will be performed $N = 10 \times i$ ($i = 1, \dots, 10$) times. Let $S = S^1 \cup S^2$, in each run 50% of samples were randomly selected as the training sample set, and the remaining 50% as the test sample set. The experimental results are shown in Fig. 2, which gives a statistical analysis of the experimental results box plots. The figure shows that the algorithm in terms of stability or in the detection performance have achieved good results.

5 Conclusion

In digital mammograms, an important sign of the early breast cancer is the existence of Microcalcification Clusters (MCs). One of the key techniques for early diagnosis of the breast cancer is to detect MCs and to judge whether they are malignant or not in mammograms. In our paper, we first briefly summarizes the status of subspace learning methods, and then introduces the principle of integrated learning levels and selective integration of learning and design a hybrid subspace selective fusion approach to detect MCs. In our algorithm, the subspace learning algorithms will be selectively fused according to the ability of preserving the classification information. Experimental results show that the proposed method can improve the performance and stability of MCs detection task and could get satisfactory results on sensitivity and reduce false positive rate, which can achieve better results than a single detection subspace learning-based approach, and stability of the method is better, to better adapt to the noise environment..

References

- [1] E. A. Alanís-Reyes, J. L. Hernández-Cruz, J. S. Cepeda, C. Castro, H. Terashima-Marín, S. E. Conant-Pablos, Analysis of machine learning techniques applied to the classification of masses and microcalcification clusters in breast cancer computer-aided detection [J], *Journal of Cancer Therapy*, 3(1), 2012, 1020-1026
- [2] X. Zhang, A new approach for clustered MCs classification with sparse features learning and TWSVM [J], *The Scientific World Journal*, 2014 (2014), Article ID 970287, 8
- [3] X. Zhang, X. Gao, Twin support vector machines and subspace learning methods for microcalcification clusters detection [J], *Engineering Applications of Artificial Intelligence*, 25(5), 2012, 1062-1072
- [4] J. Ren, Ann vs. svm: Which one performs better in classification of McCs in mammogram imaging [J], *Knowledge-Based Systems*, 26(1), 2012, 144-153
- [5] R. Sivakumar, M. Karnan, A novel approach to detect microcalcification in mammogram image using evolutionary algorithm [J], *Artificial Intelligent Systems and Machine Learning*, 4(3), 2012, 135-140
- [6] L. Van der Maaten, G. Hinton, Visualizing data using T-SNE [J], *Journal of Machine Learning Research*, 9(85), 2008, 2579-2605
- [7] M. Li, Z. H. Zhou, Improve computer-aided diagnosis with machine learning techniques using undiagnosed samples [J], *IEEE Transactions on Systems, Man and Cybernetics, Part A: Systems and Humans*, 37(6), 2007, 1088-1098

- [8] J. L. Jesneck, L. W. Nolte, J. A. Baker, C. E. Floyd, J. Y. Lo, Optimized approach to decision fusion of heterogeneous data for breast cancer diagnosis [J], *Medical Physics*, 33(8), 2006, 2945-2954
- [9] M. De Santo, M. Molinara, F. Tortorella, M. Vento, Automatic classification of clustered microcalcifications by a multiple expert system [J], *Pattern Recognition*, 36(7), 2003, 1467-1477
- [10] J. Llinas, D. L. Hall, M. E. Liggins, *Handbook of Multisensor Data Fusion: Theory and Practice*, CRC Press, 2009
- [11] Z. H. Zhou, J. Wu, W. Tang, Ensembling neural networks: Many could be better than All [J], *Artificial Intelligence*, 137(1), 2002, 239-263
- [12] M. P. Perrone, L. N. Cooper, When networks disagree: Ensemble methods for hybrid neural networks [J], DTIC Document, Tech. Rep., 1992
- [13] H. Abdi, L. J. Williams, *Principal component analysis* [J], *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(4), 2010, 433-459
- [14] R. A. Fisher, The use of multiple measurements in taxonomic problems [J], *Annals of Eugenics*, 7(2), 1936, 179-188
- [15] H. H. Harman, *Modern Factor Analysis*, University of Chicago Press, 1976
- [16] H. Caussinus, A. Ruiz-Gazen, Projection pursuit exploratory data analysis [J], *Computational Statistics & Data Analysis*, 20(6), 2009, 67-92
- [17] T. F. Cox, M. A. Cox, *Multidimensional Scaling*, CRC Press, 2010
- [18] D. K. Agrafiotis, Stochastic proximity embedding [J], *Journal of Computational Chemistry*, 24(10), 2003, 1215-1221
- [19] J. B. Tenenbaum, Mapping a manifold of perceptual observations [J], *Advances in Neural Information Processing Systems*, Vol. 10, 1998, 682-688
- [20] K. Q. Weinberger, L. K. Saul, An introduction to nonlinear dimensionality reduction by maximum variance unfolding [J], *AAAI*, Vol. 6, 2006, 1683-1686
- [21] X. Niyogi, Locality preserving projections [J], *Neural Information Processing Systems*, Vol. 16, 2004, 153-162
- [22] X. He, D. Cai, S. Yan, H. J. Zhang, Neighborhood preserving embedding [J], *Tenth IEEE International Conference on Computer Vision (ICCV 2005)*, Vol. 2, 1208-1213
- [23] T. Zhang, J. Yang, D. Zhao, X. Ge, Linear local tangent space alignment and application to face recognition [J], *Neurocomputing*, 70(7), 2007, 1547-1553
- [24] B. Schölkopf, A. Smola, K. R. Müller, Nonlinear component analysis as a kernel eigenvalue problem [J], *Neural Computation*, 10(5), 1998, 1299-1319
- [25] M. Brand, Charting a manifold [J], *Advances in Neural Information Processing Systems*, Vol. 14, 2002, 961-968
- [26] J. Verbeek, Learning nonlinear image manifolds by global alignment of local linear models [J], *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(8), 2006, 1236-1250