UDC 004.65, DOI: 10.2298/CSIS1001127Z

Effective Semi-supervised Nonlinear Dimensionality Reduction for Wood Defects Recognition

Zhao Zhang[1] and Ning Ye[1,2]

[1] School of Information Technology, Nanjing Forestry University,
210037 Nanjing, China
zzhang618@gmail.com
[2] School of Computer Science and Technology, Shandong University,
250100 Jinan, China
ye.ning@yahoo.com.cn

Abstract. Dimensionality reduction is an important preprocessing step in high-dimensional data analysis without losing intrinsic information. The problem of semi-supervised nonlinear dimensionality reduction called KNDR is considered for wood defects recognition. In this setting, domain knowledge in forms of pairs constraints are used to specify whether pairs of instances belong to the same class or different classes. KNDR can project the data onto a set of 'useful' features and preserve the structure of labeled and unlabeled data as well as the constraints defined in the embedding space, under which the projections of the original data can be effectively partitioned from each other. We demonstrate the practical usefulness of KNDR for data visualization and wood defects recognition through extensive experiments. Experimental results show it achieves similar or even higher performances than some existing methods.

Key words: semi-supervised learningm, dimensionality reduction, wood defects recognition, (dis-)similar constraints.

1. Introduction

Many research and application areas need to deal with high-dimensional data, which leads to a hot of studying the methods of dimensionality reduction, whose aim is to find a meaningful low dimensional manifold from the original data. In many real applications, data lying in high-dimensional ambient space can be modeled by a low-dimensional nonlinear manifold.

Principal component analysis (PCA) [8], kernel principal component analysis (KPCA) [9] and kernel Fisher discriminate analysis (KFD) [12][14] are widely used in pattern recognition[15][17] [19]. Utilizing domain knowledge has been an important issue in data mining tasks [1]. In general, domain knowledge can be expressed in diverse forms, such as class labels and pairs constraints [2]. Semi-supervised dimensionality reduction is a new issue in semi-supervised learning, which learns from a combination of both labeled and

unlabeled data. In many practical applications, unlabeled samples are readily available but labeled ones are fairly expensive to obtain, so semi-supervised dimensionality reduction has attracted much attention.

Currently, automatic wood defects recognition is one of hot issues in the mechanical wood industry. Matti Niskanen [3] and P. Meinlschmidt [6] used different technologies to study wood based defect detection problems. Here, we also present a solution the wood defects recognition and evaluate the performance of the proposed system by giving quantitative experiments. The block diagram of SVM classifier based detection system is shown in Fig.1.

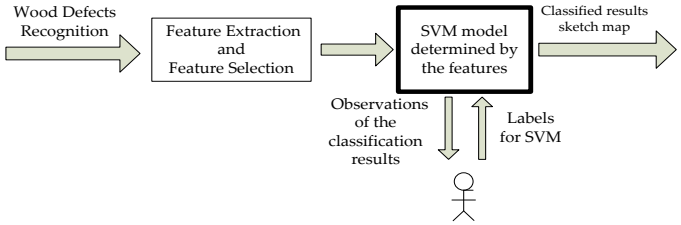

Fig.1. The diagram of SVM based recognition system.

Considering the pairs constraints can be derived from labeled data and can be automatically obtained without human intervention [1]. Here, we propose a semi-supervised nonlinear dimensionality reduction method (KNDR) with the pairs constraints for wood defects recognition. The rest of this paper is organized as follows. In section 2, we present a sensitivity analysis of KNDR and numerically evaluate it for wood defects recognition in section 3. We conclude this paper and raise some issues for future research in section 4.

2. Related Work

2.1. Linear dimensionality reduction method

For a set of wood samples together with some pairs of similar constraints and dissimilar constraints, we denote the domain knowledge containing similar and dissimilar pairs by *S* and *D*. For a set of data $X \in R^n$ and $X = (x_1, x_2, ..., x_m)$, a simpler way of defining a criterion for the desired metric is to demand: if x_i and x_j are similar, pairs $\{x_i, x_j\}$ belong to set *S*, and *D* otherwise. To improve the tightness among similar pairs and separate dissimilar ones better, we consider shrinking distances between similar pairs, i.e. $(x_i, x_j) \in S$, by minimizing $\| x_i - x_j \|^2$, while expanding distances between dissimilar pairs, i.e. $(x_i, x_j) \in D$, by maximizing $\| x_i - x_j \|^2$ .Given a set of multivariate data X with respect to the

pairs constraints, we aim to find a set of vectors $\omega = (\omega_1, \omega_2, ..., \omega_d)$, such that the transformed low- dimensional representations $y_i = \omega^T x_i$ of x_i can preserve the structure of labeled and unlabeled samples as well as the pairs constraints, i.e. instances involved by *S* should be close while instances involved by *D* should be far as soon as possible.

Noticing that $\omega^T X$ means matrix ω basically projects the data on to a set of 'useful' features in embedding spaces [18]. Ideally, the set should be small that means the small rank for $\omega^T \omega$ or ω, as rank $(\omega^T \omega)$ = rank (ω) is desired. Our aim is to minimize it by finding an eigen-decomposition $E(\omega^T \omega) = V \Lambda V^T$, then rank $(\omega^T \omega)$ = rank $(\Lambda) = \| \Lambda \|_0$, but a direct minimization of this zero norm is very difficult and here we approximate it by the L2-norm $\| \Lambda \|_2 = \| \omega^T \omega \|_2$ in the model [18]. Then the n × d KNDR transformation matrix T_{KNDR} is defined in Eq.1, where $\| . \|$ denotes the L2-norm.

$$T_{KNDR} = \arg\min_{\omega} \frac{1}{2} \| \omega^T \omega \|^2 + \frac{C_u}{2 | N_u |} \sum_{(x_i, x_j)} \| \omega^T x_i - \omega^T x_j \|^2 + \frac{C_S}{2 | N_S |} \sum_{(x_i, x_j) \in S} \| \omega^T x_i - \omega^T x_j \|^2 - \frac{C_D}{2 | N_D |} \sum_{(x_i, x_j) \in D} \| \omega^T x_i - \omega^T x_j \|^2 \quad (1)$$

The two terms of Eq.1 expresses the average squared distance between all unlabeled samples in the embedding space and N_u is the number of unlabeled samples. $y_i \in R^d (1 \le d \le n)$ are the embedded data where *d* is the dimension of reduced space. N_S and N_D are respectively the numbers of samples under the similar and dissimilar constraints. The intuition behind Eq.1 is to let the average distances in the embedding space between instances involved by *S* be as small as possible, while distances between instances involved by *D* should be as large as possible. Since metrics between instances in the same class is typically smaller than those in different classes, here we add a scaling parameter C_S to balance the contribution of two terms and C_D for three terms in Eq.1. Intuitively, distances of samples involved in *S* should typically be close to the expected metric [20], so we empirically set $C_u = 1, C_S \ge 1$ and $C_D > 1$.

2.2. Kernelization

Implicitly in the kernel Hilbert space *H* connected to the kernel function *K* used. According to [4], a kernel is a function in the input space and at the same time is the inner product in the feature space through the kernel-induced nonlinear mapping. Since for each kernel there exist a mapping ϕ corresponds to a

scalar product and maps input patterns x_i to $\phi(x_i)$,here we define a mapping $\phi: R^n \to H^p (p > n)$ and choose *RBF* kernel for data projections:

$$K(x, x^T) = \exp\wedge\left(-\| x - x^T \|^2 / 2\sigma^2\right) \tag{2}$$

Mika et al proved that every solution $\omega \in H$ (Kernel space) can be written as an expansion in terms of the mapped samples data [1]. i.e. the vectors ω^ϕ in the high-dimensional kernel space can be rewritten as

$$\omega^\phi = \sum_{i=1}^{m} \alpha_i \phi(x_i) = \phi(X)\alpha = K(X, \alpha) \tag{3}$$

where the input patterns in *H* are denoted by $\phi(X) = (\phi(x_1), \phi(x_2), ..., \phi(x_n))$ and $\alpha = (\alpha_1, ..., \alpha_d)$ are the transformation matrix and $x_i, i = 1, 2..., m$ form a vector space. Especially, for all functions with the form of Eq.3, we get

$$\langle K(., X), K(., X^T) \rangle_R = \langle \phi(X), \phi(X^T) \rangle_H = K(X, X^T) \tag{4}$$

By substituting Eq.3 into Eq.1, we can update Eq.1 by

$$\begin{aligned} \min_\alpha J(\alpha) = & \frac{1}{2} \alpha^T \left(\phi(X)\right)^T \phi(X) \alpha \alpha^T \left(\phi(X)\right)^T \phi(X) \alpha + \\ & \frac{1}{2} \sum_{i,j} \| \alpha^T \left(\phi(X_i)\right)^T \phi(X_i) - \alpha^T \left(\phi(X_j)\right)^T \phi(X_j) \|^2 R_{ij} \end{aligned} \tag{5}$$

where weights R_{ij} defined in Eq.5 satisfy the following formulations:

$$R_{ij} = \begin{cases} \frac{C_u}{|N_u|} + \frac{C_S}{|N_S|} & if \ \ pairs(x_i, x_j) \in S \\ \frac{C_u}{|N_u|} - \frac{C_D}{|N_D|} & else\, if \ \ pairs(x_i, x_j) \in D \\ \frac{C_u}{|N_u|} & otherwise \end{cases} \tag{6}$$

By substituting Eq.6 into Eq.5, we can rewrite Eq.5 as

$$\begin{aligned} J(\alpha) &= \alpha^T \left(\phi(X)\right)^T \phi(X) \alpha \alpha^T \left(\phi(X)\right)^T \phi(X) \alpha \\ &\quad + \frac{1}{2} \sum_{i,j} \| \alpha^T \left(\phi(X_i)\right)^T \phi(X_i) - \alpha^T \left(\phi(X_j)\right)^T \phi(X_j) \|^2 R_{ij} \\ &= \alpha^T \left(\phi(X)\right)^T \phi(X) I \left(\phi(X)\right)^T \phi(X) \alpha + \sum_i \alpha^T \left(\phi(X)\right)^T \phi(x_i) W_{ii} \left(\phi(x_i)\right)^T \phi(X) \alpha \\ &\quad - \alpha^T \left(\phi(X)\right)^T \phi(X) R \left(\phi(X)\right)^T \phi(X) \alpha \\ &= \alpha^T \left(\phi(X)\right)^T \phi(X) (I + W - R) \left(\phi(X)\right)^T \phi(X) \alpha \\ &= \alpha^T M \alpha \end{aligned} \tag{7}$$

where $M = KQK$, $K = \left(\phi(X)\right)^T \phi(X)$ and *W* is a diagonal matrix whose entries are column (or row) sums of the matrix R , i.e. $W_{ii} = \sum_j R_{ij}$. $Q = I + W - R$ is called Laplacian matrix. The coefficient for the identity matrix is omitted here. Thus Eq.5 or Eq.7 can be simplified as

$$\text{minimize } J(\alpha) = \alpha^T M \alpha \qquad subject\ to\ \ \alpha^T \alpha = I \tag{8}$$

Clearly, the problem in Eq.8 is a standard eigenvalue problem. Forming the Lagrangian of Eq.8 with the multiplier λ, i.e. $L(J,\lambda)=\alpha^T M\alpha-\lambda(\alpha^T\alpha-1)$, by taking partial derivatives with respect to the variables α and zeroing it by $\nabla_\alpha L(J,\alpha)=2M\alpha-2\lambda\alpha=0$, from which α can be easily solved by computing the eigenvectors of the matrix *M* corresponding to the *d* smallest eigenvalues.

3. Experiments and analysis

In the experiments, we investigate the performance of KNDR method for data visualization and wood defects recognition. We first take the Iris and Soybean datasets [16] for visualization. Then, we perform experiments on the wood image database [7,13] for wood defects recognition. Local Binary Pattern (LBP) [5,10] is used to extract the features from the wood images. In the experiments, support vector machine (SVM) classifier is used for classification. The process of semi-supervised learning can be described as follows: we first use the labeled samples to train a decision-making function, and then use the function to label the unlabeled samples. The tunable parameters C_u, C_S and C_D are always set to 1,1 and 20 respectively if without extra explanations.

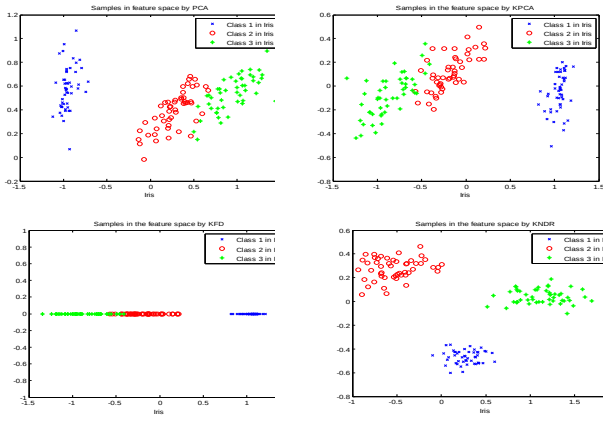

Fig. 2. Distributions of the features extracted by PCA, KPCA, KFD and KNDR on the Iris dataset (D = 4, N = 150, T = 3, d = 2)

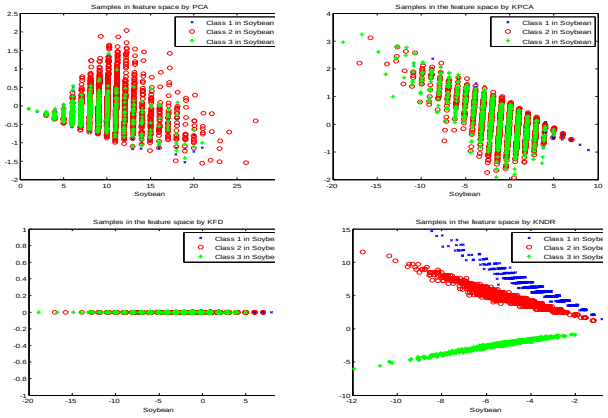

Fig. 3. Distributions of the features extracted by PCA, KPCA, KFD and KNDR on the Soybean dataset (D = 8, N = 4177, T = 3, d = 2)

3.1. Data visualization

In this subsection, we apply KNDR and some existing dimensionality reduction methods (i.e., PCA, KPCA and KFD) to the Iris and Soybean data sets with three classes and investigate how they behave in data visualization tasks. Figs.2 and 3 respectively depict the Iris and Soybean data embedded in the two-dimensional embedding space discovered by each method, where D is the dimension of the dataset, N is the number of instances, T is the number of classes and d is the number of selected features. The iris data set, popularly used for testing clustering and classification algorithms. The embedding spaces by PCA and KPCA are better, but both inferior to that of KNDR. The '×'-class can be clearly observed by the four methods, however, for other two classes, PCA, KPCA and KFD tend to mix the data, but KNDR can separate them well. For the Soybean data set, the three classes are completely mixed in the original input space. PCA, KPCA and KFD can not work on the data set, while KNDR can keep in-class sample pairs close and between-class sample pairs apart effectively. Based on above experimental results, KNDR is found to be more appropriate for embedding the samples data than PCA, KPCA and KFD, implying that our primal goal has been successfully achieved.

3.2. Experimental results on wood defects recognition

There are many kinds of defects on the wood surfaces, such as discoloration, decadent, knots, etc. While knots can be seen more often and directly affect the quality of boards. Several varieties of defects are shown in Fig.4.

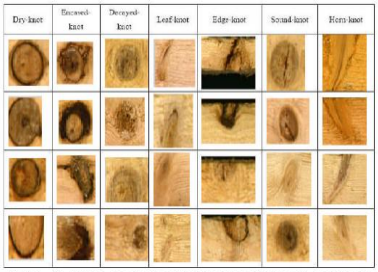

Fig. 4. The common kinds of knot defects.

Table 1. Sample class distribution.

Types	Number of samples		Total
	Training set	Test set	
Dry	69	27	96
Encased	19	4	43
Decayed	14	8	22
Leaf	37	3	70
Edge	62	1	63
Sound	126	10	186
Horn	32	5	37
Total	362	63	425

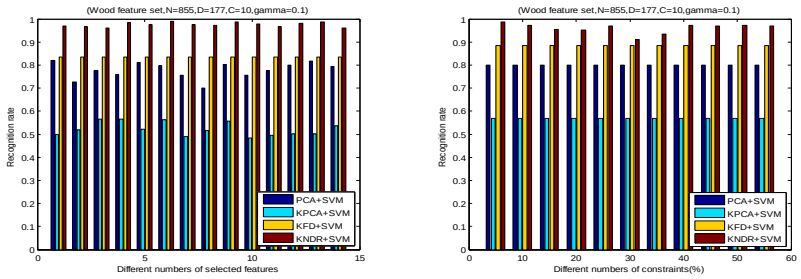

Fig. 5. Accuracy vs. different numbers of selected features and constraints.

For dimensionality reduction, we first transform the color wood images into three cues of R, G and B respectively and divide the images into many small blocks, and then extract the features from each block. Finally, we obtain 59 features from each cue, totally 177 dimensional features. In the experiments, the training set used in the experiments is very important. If the training set is chose improperly, the test results will be affected directly and make the system null. Here, we aggregately select 855 samples, including 417 positive samples (labeled 1) and 438 negative samples (labeled -1). We mainly take the knot and wood image database as basis, in which the sample class distribution is described in Table 1 and the number of different kinds of knot defects is also listed. The results reported here are based on a set of hundreds of wood images with over 200 labeled defects in the wood database.

Classification. Here, some experiments on the feature set are performed. The pairs constraints are obtained by randomly selecting pairs of instances from the whole data set, and creating similar or dissimilar constraints depending on whether the underlying classes of the two instances are the same or not. After obtaining the constraints, data without constraints in the whole data set are used as unlabeled data. Fig.5 displays the accuracy on the wood feature set under different numbers of selected features and constraints, where C is for SVM classifier and gamma is chose as the kernel parameters. From Fig.5, we find KDNR can almost always receive the highest recognition rate over 95% by comparing with the popular KFD, KPCA and PCA methods. The performance of KPCA is relatively poor in this feature set. With the increasing of the numbers of selected features, KNDR can keep stable for a

wide range and the performance of KFD method can also be thought better here.

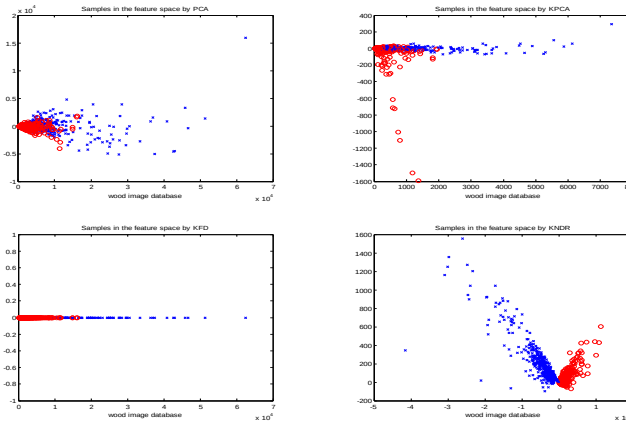

Fig.6. Distributions of the features extracted by PCA, KPCA, KFD and KNDR on the wood image database (D = 177, N = 855, T = 2, d = 10).

Fig.6 shows the distributions of the selected features extracted by PCA, KFD, KPCA and KNDR from the feature set. Intuitively, the embedded positive and negative samples are easier to be partitioned from each other in the feature spaces discovered by KNDR. Similar to section 3.1, PCA, KPCA and KFD tend to mix the projections of data of different classes.

Table 2. Results of the averaged accuracy and runtime.

Methods	Averaged accuracy (%)	Total runtime (s)
PCA+SVM	78.65	8.7
KFD+SVM	83.47	16.5
KPCA+SVM	55.80	20.6
KNDR+SVM	97.43	11.2

The above experiments have evaluated the learning ability of the proposed method. Next, the runtime performance of KNDR will be discussed. Table 2 gives the averaged accuracy and runtime of wood defects recognition under different number of selected features according to the left panel of Fig.5. From Table 2, the averaged accuracy of KNDR reaches 97.43% and the runtime performance is better than that of KPCA and KFD, but slightly worse than PCA, and we believe this is because PCA does use the linear projection.

Recognition results. There are no woods with same properties in color and surfaces show many varieties of texture characteristics, such as rough, fuzzy, etc. Even for the same species, the defects might greatly vary in shape and colors [11]. Furthermore, with large varieties of defects and the involvement of human factors caused the current detection methods are vulnerable to the

interferences on wood surfaces. Here, we barely emphasize particularly the category, but classify them as defects.

Let $A_i = [\psi_{i1}, \psi_{i2}, ..., \psi_{in}]$, $A_j = [\psi_{j1}, \psi_{j2}, ..., \psi_{jn}]$ represent two wood feature matrices, then the distance metric between them can be defined as

$$Dist(A_i, A_j) = \sum_{t=1}^{n} \left\| \psi_{it} - \psi_{jt} \right\|^2 \quad (9)$$

Supposing matrices A_1 and A_2 are use for storing the positive and negative samples, and are labeled by 1 and -1 respectively. For any new data point A, if $Dist(A, A_1) = \min Dist(A, A_j)$ and A_1 belongs to the negative samples with defects, then we judge A as a defect region, and normal region otherwise.

During the experiments, choosing an appropriate feature set is also very important. In the experiments, all the feature sets are selected by pixels 35*35 with good generalization ability. The number of constraints (60%) is randomly selected from the training set. Fig.7 displays the recognition results of the proposed method and the rectangular boxes are all located and drawn by the computers automatically.

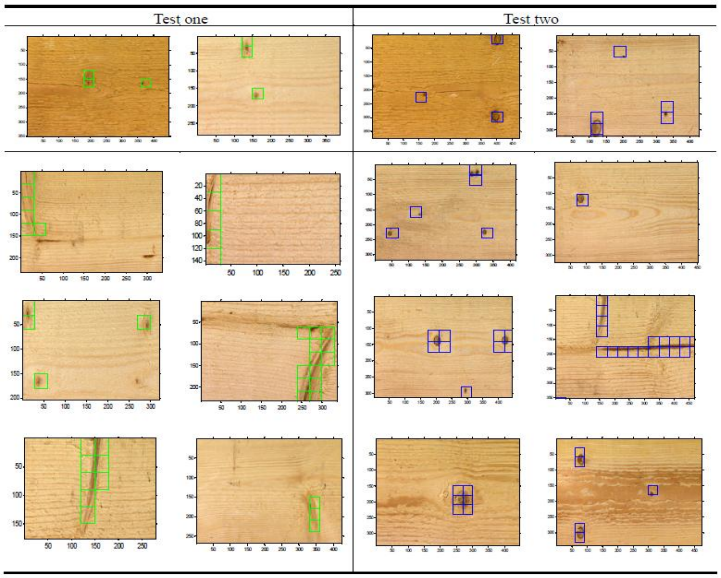

Fig.7. Experimental results of the wood defects recognition.

4. Conclusions and further works

The aim of this paper is to present a semi-supervised dimensionality reduction method called KNDR, focusing on domain knowledge in the form of the pairs constraints together with unlabeled data samples for defects recognition, not recognition results, however, the experimental results presented here are good or considerably better. We consider shrinking distances between similar pairs, while expanding distances between dissimilar ones in the embedding space. KNDR algorithm is interesting from a number of advantages: (1) KNDR is a standard eigenvalue problem and can be efficiently computed; (2) KNDR can preserve the structure of labeled and unlabeled samples and the constraints defined in embedding spaces; (3) For data visualization, the projections of the data in different classes can be effectively partitioned from each other; (4) In most cases, KNDR performs better than the classical PCA, KPCA and KFD methods. Furthermore, the runtime performance of KNDR is better than those of KPCA and KFD, but slightly worse than linear PCA.

Next, we will investigate how to choose the proper kernel parameters for the nonlinear model. Moreover, investigating whether KNDR can preserve the local structure of the original data is also an interesting future work.

5. Acknowledgments

This research was supported by the Innovation Foundation 2008 of Jiangsu Province of China under Grant No.164070265, the Innovation Program Foundation 2009 of Nanjing Forestry University, the Scientific Research Foundation 2009 of Jiangsu Province of China under Grant No.CX09S_013Z and the Natural Science Foundation of Jiangsu Province under Grant. BK2009393.

6. References

1. A.BAR-HILLEL, T. HERTZ, N. SHENTAL, AND D. WEINSHALL, Learning a mahalanobis metric from equivalence constraints, Journal of Machine Learning Research, 6, 937-965. (2005)
2. Xing, E.P., Ng, A.Y., Jordan, M.I., Russell, S.: Distance metric learning with application to clustering with side-information. In S. Becker, S.T., Obermayer, K., eds. In Proceeding of NIPS. MIT Press, Cambridge, MA, 15, 505-512. (2003)
3. Olli Silv'en,Matti Niskanen, Hannu Kauppinen. Wood inspection with non-supervised clustering. Machine Vision and Applications, 275-285. (2003)
4. Zhang, D.Q., Chen, S.C.: Clustering incomplete data using kernel-based fuzzy c-means algorithm. Neural Processing Letters, 18, 3, 155-162. (2003)
5. Ojala T, Pietikäinen M & Mäenpää T. Multi -resolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24, 7, 971-987. (2002)

6. P. Meinlschmidt, Wilhelm-Klauditz-Institut(WKI), Fraunhofer-Institute for wood research, Braunsch weig. Thermo-graphic detection of defects in wood and wood -based materials. In Proceeding of the 14th international Symposium of nondestructive testing of wood, Hannover, Germany: (2005)
7. Sommardahl O, Usenius A. Wood samples image data. VTT Building Technology. (1999)
8. Hastie, T., Tibshirani, R., & Friedman, J. (Eds.). The elements of statistical learning: data mining, inference, and prediction. New York: Springer. (2001)
9. B. Schölkopf, S. Mika, A.J. Smola, G. Rätsch, and K. -R. Müller. Kernel PCA pattern reconstruction via approximate pre-images. In L. Niklasson, M. Bodén, and T. Ziemke, editors. In Proceeding of the 8th International Conference on Artificial Neural Networks, Berlin, Springer-Verlag, 147-152. (1998a)\
10. Mäenpää T & Pietikäinen M.Texture analysis with local binary patterns. In: Chen CH & Wang PSP (eds) Handbook of Pattern Recognition & Computer Vision, 3rd ed, World Scientific, Singapore, in press (invited chapter).(2004)
11. Moring, I. and Silven, O. Automatic Visual Lumber Inspection: Experiments, Finnish Artificial Intelligence Symposium STeP-86, 2, 290-294. (1986)
12. Mika, S., Ratsch, G.,Weston, J. et al..Fisher discriminant analysis with kernels. In Y.H. Hu, J. Larsen, E. Wilson, and S. Douglas, editors, Neural Networks for Signal Processing, IX,41-48. (1999)
13. Olli Silv'en, Matti Niskanen, Hannu Kauppinen.Wood inspection with non-supervised clustering. Machine Vision and Applications, 275-285. (2003)
14. Mika., S., Ratsch, G., Müller, K. R., A mathematical programming approach to the Kernel Fisher algorithm. In: Leen T.K., Dietterich T.G.,Tresp V. eds, Advances in Neural Information Processing Systems, Cambridge, MA:MIT Press, 13, 591-597. (2001)
15. J. Yang, A.F. Frangi, J.Y. Yang, D. Zhang, and Z. Jin. KPCA plus LDA: a complete kernel Fisher discriminant framework for feature extraction and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 27, 2, 230-244. (2005)
16. C. Blake, E. Keogh, and C.J. Merz. UCI repository of machine learning databases. Department of Information and Computer Science, University of California, Irvine, 1998.[http://www.ics.uci.edu/~mlearn /ML Repository.html].
17. J.W. Lu, K.N. Plataniotis, and A.N. Venetsanopoulos. Face recognition using kernel direct discriminant analysis algorithms. IEEE Transactions on Neural Networks, 12, 117-126. (2003)
18. Ivor W. Tsang, James T. Kwok. Distance Metric Learning with Kernels. In Proceeding of International Conference on Artificial Neural Networks. (2003)
19. T. Xiong, J.P. Ye, Q. Li, V. Cherkassky, and R. Janardan. Efficient kerne discriminant analysis via QR decomposition. In Advances in Neural Information Processing Systems, 17. (2005)
20. Daoqiang Zhang, Zhi-Hua Zhou, and Songcan Chen. Semi-supervised dimensionality reduction. In Proceedings of the 7th SIAM International Conference on Data Mining (SDM'07), Minneapolis, MN, 629-634. (2007)

Zhao Zhang has graduated at the Faculty of Information Science and Technology in Nanjing Forestry University, in 2008. So far, he has authored several research papers. He is currently a master candidate for Computer Applied Science and Technology in Nanjing Forestry University. His research interests are machine learning and pattern recognition.

Ning Ye has graduated from Master's Degree from the Faculty of Test Measurement Technology and Instruments in Nanjing University of Aeronautics and Astronautics, in 1996. He got his Ph.D. from the Faculty of Computer Science and Technology in Southeast University, in 2006. He is currently a Professor of Computer Science in Nanjing Forestry University.

Received: May 05, 2009; Accepted: September 26, 2009.