HYPERSPECTRAL DIMENSION REDUCTION USING GLOBAL AND LOCAL INFORMATION BASED LINEAR DISCRIMINANT ANALYSIS

Ufuk Sakarya

TÜBİTAK UZAY (The Scientific and Technological Research Council of Turkey, Space Technologies Research Institute), ODTÜ Yerleşkesi, 06531, Ankara, Turkey – ufuk.sakarya@tubitak.gov.tr

Commission VII, WG VII/4

KEY WORDS: *Hyperspectral image processing, dimension reduction, complete global-local linear discriminant analysis, classification.*

ABSTRACT:

Hyperspectral image classification has become an important research topic in remote sensing. Because of high dimensional data, a special attention is needed dealing with spectral data; and thus, one of the research topics in hyperspectral image classification is dimension reduction. In this paper, a dimension reduction approach is presented for classification on hyperspectral images. Advantages of the usage of not only global pattern information, but also local pattern information are examined in hyperspectral image processing. In addition, not only tuning the parameters, but also an experimental analysis of the distribution of the hyperspectral data is demonstrated. Therefore, how global or local pattern variations play an important role in classification is examined. According to the experimental outcomes, the promising results are obtained for classification on hyperspectral images.

1. INTRODUCTION

Hyperspectral image classification has become an important research topic in remote sensing. It has been used in a lot of applications, such as land use / land cover, target detection, anomaly detection, etc. in domain of agriculture, forestry, geology, ecological monitoring and disaster monitoring (Teke, et al., 2013). Integration of spectral and spatial information has become in increasing trend for hyperspectral image classification (Plaza, et al., 2009), (Fauvel, et al., 2013), (Camps-Valls, et al., 2014). Because of high dimensional data, a special attention is needed dealing with spectral data; and thus, one of the research topics in hyperspectral image classification is dimension reduction. In this paper, a dimension reduction approach is introduced for classification on hyperspectral images.

Dimension reduction is an important research topic in pattern recognition. In a classification problem, an essential training data increases exponentially with respect to dimension; therefore, "curse of dimensionality" problem arises (Duda, et al., 2001). In order to overcome the problem, it is needed a technique to reduce dimension size into an acceptable level for classification purposes. A lot of methods have been developed for this purpose. Principal component analysis (PCA) and linear discriminant analysis (LDA) are very-well known two of them. PCA is an unsupervised method that tries to find a projection by using eigenvectors of a scatter matrix of all data (Duda, et al., 2001). Unlike PCA, LDA is a supervised dimension reduction method that tries to find a projection by which different class members are far from each other and same class members are close to each other (Duda, et al., 2001). On the other hand, LDA has drawbacks such as the singularity problem, the distribution assumption, the small sample size problem (Li, et al., 2012).

Much more efforts are made for an improvement of LDA. In order to overcome the problem of a singularity matrix in LDA, Fisherface method (FF), which makes a PCA based projection and a change of matrix size so that the matrix is nonsingular, is proposed (Belhumeur, et al., 1997). Yan et al. (Yan, et al.,

2007) propose Marginal Fisher Analysis (MFA) in which two graphs called by the intrinsic graph and the penalty graph are constructed for modeling intraclass compactness and interclass separability, respectively.

LDA focuses on global geometrical structure information of data points and disregards the local information of data points (Zhang, et al., 2014). Some efforts have been made in order to integrate global and local information of data into LDA. Gao et al. (Gao, et al., 2012) introduce enhanced Fisher discriminant criterion (EFDC) in which a new scatter matrix that models intraclass variability of patterns is added to LDA. Zhang et al. (Zhang, et al., 2014) propose a complete global-local LDA (CGLDA) for dimension reduction. In addition to the two scatter matrixes in LDA, two new scatter matrixes are also calculated in CGLDA. CGLDA uses both global and local information of data. CGLDA uses three types of local information: Local similarity information, local intra-class pattern variation and local inter-class pattern variation.

Dimension reduction is also well-studied research area in hyperspectral image processing. PCA and LDA are also used in hyperspectral image processing (Cheriyadat & Bruce, 2003). Du and Chang (Du & Chang, 2001) introduce a linear constrained distance-based discriminant analysis (LCDA) method. In this approach, a constraint, in which centers of classes are aligned with predetermined directions, is used besides inter-distance between classes and intra-distance within classes. Modified Fisher's linear discriminant analysis (MFLDA) is introduced by Du (Du, 2007). To overcome the problem of finding training samples, only class signatures are used. In MFLDA as distinct from LDA, the total scatter matrix is used instead of the within-class scatter matrix and class signatures are used as mean values of classes. Kuo and Landgrebe (Kuo & Landgrebe, 2004) propose a nonparametric weighted feature extraction method (NWFE). In this approach, different weights are computed for every sample.

Kozal et al. (Kozal, et al., 2013) make a comparative analysis according to classification performance and computation times

for some linear and non-linear dimension reduction methods. Teke and Sakarya (Teke & Sakarya, 2013) propose a hyperspectral image classification method based on EFDC. Huang et al. (Huang, et al., 2012) introduce semi-supervised marginal Fisher analysis (SSMFA) for hyperspectral image classification. In SSMFA, besides labeled data, an objective function is added for unlabeled training data.

In this paper, a classification method based on CGLDA is proposed for hyperspectral images. The dimension of the hyperspectral data is reduced by using CGLDA. Kuo and Landgrebe emphasize the use of local information to improve in the discriminant analysis for feature extraction (Kuo & Landgrebe, 2004). Advantages of the usage of not only global pattern information, but also local pattern information are examined in hyperspectral image processing. In brief, we propose a novel approach to contribute to solve the problem based on the following points:

- To the best of my knowledge, the proposed method is the first application of CGLDA in hyperspectral application domain. Advantages of the usage of not only global pattern information, but also local pattern information are experimentally demonstrated by comparative analysis.
- In this work, an experimental strategy is proposed in order to tune the parameters of CGLDA in hyperspectral application domain. In addition, not only tuning the parameters, but also an experimental analysis of the distribution of the hyperspectral data is demonstrated. Therefore, how global or local pattern variations play an important role in classification is examined.

The rest of the paper is organized as follows. Section 2 overviews LDA and CGLDA methods. Section 3 introduces the proposed CGLDA based hyperspectral image classification method. Section 4 gives information about the data set, performance evaluation and implementation details. Section 5 introduces the experimental strategy for tuning the parameters. Section 6 demonstrates comparative performance results. Finally, Section 7 presents some concluding remarks and future research directions.

2. LDA AND CGLDA

LDA is a supervised dimension reduction method that tries to find a projection matrix W_{LDA} by which different class members are far from each other and same class members are close to each other (Duda, et al., 2001). Let's x_1, x_2, ..x_i, .. x_n d-dimensional n samples that belong to c number of classes, that is a matrix X in dimension of d x n. Two scatter matrixes are calculated for LDA: S_W (within-class scatter matrix) and S_B (between-class scatter matrix).

$$S_W = \sum_{i=1}^{c}\sum_{j=1}^{N_i}(x_j^i - \mu^i)(x_j^i - \mu^i)^T \qquad (1)$$

$$S_B = \sum_{i=1}^{c} N_i(\mu^i - \mu)(\mu^i - \mu)^T \qquad (2)$$

where N_i is the number of samples in class i, μ_i is the mean vector of class i, μ is the mean vector of all samples. Objective function of LDA is as follows:

$$W_{LDA} = \arg\max_{W}\left|\frac{W^T S_B W}{W^T S_W W}\right| \qquad (3)$$

If S_W is nonsingular, then Eq. (3) can be solved by Eq. (4). In order to overcome the problem of singularity of S_W, one of the proposed approaches is FF (Belhumeur, et al., 1997). FF firstly makes a PCA based projection and a change of matrix size so that S_W is nonsingular.

$$S_W^{-1}S_B W = \lambda W \qquad (4)$$

Zhang et al. (Zhang, et al., 2014) proposed CGLDA for dimension reduction. In addition to the two scatter matrixes in LDA, two new scatter matrixes are also calculated in CGLDA. These are S_{TL} called by the total local pattern variation scatter matrix and S_{LW} called by the local within-class scatter matrix. Objective function of CGLDA is as follows:

$$W_{CGLDA} = \arg\max_{W}\left|\frac{W^T[\alpha S_B + (1-\alpha)S_{TL}]W}{W^T[\varepsilon S_W + (1-\varepsilon)S_{LW}]W}\right| \qquad (5)$$

where ε is a parameter that handles the tradeoff between global and local similarity, α is a parameter that handle the tradeoff between global discriminant and local discriminant information. The both of them are in the range of 0 to 1. S_{LW} is calculated as follows:

$$S_{LW} = X(D - S)X^T \qquad (6)$$

where D is a diagonal matrix that is calculated as follows:

$$D_{ii} = \sum_{j} S_{ij} \qquad (7)$$

S is a similarity matrix that models the local similarity information and is defined as follows:

$$S_{ij} = \begin{cases} \exp\left(-\|x_i - x_j\|^2 / t\right), & If \ x_i \in N_k(x_j) or \\ & \quad x_j \in N_k(x_i) and \ C_i = C_j, \\ 0, & otherwise. \end{cases} \qquad (8)$$

where t is a parameter that controls how a similarity between two samples decreases against the distance. $N_k(x_i)$ gives a set of k nearest neighbors of x_i. C_i implies the class label of x_i. S_{TL} is calculated as follows:

$$S_{TL} = X(P - L)X^T \qquad (9)$$

where P is a diagonal matrix that is calculated as follows:

$$P_{ii} = \sum_{j} L_{ij} \qquad (10)$$

L is a weight matrix that models the local inter-class pattern variations and local intra-class pattern variations and is defined as follows:

$$L_{ij} = \begin{cases} \exp\left(-t / \|x_i - x_j\|^2\right)\left(1 + \exp\left(-\|x_i - x_j\|^2 / t\right)\right), \\ \quad If \ x_i \in N_k(x_j) \, or \, x_j \in N_k(x_i) \, and \, C_i = C_j \\ \exp\left(-t / \|x_i - x_j\|^2\right)\left(1 - \exp\left(-\|x_i - x_j\|^2 / t\right)\right), \\ \quad If \ x_i \in N_k(x_j) \, or \, x_j \in N_k(x_i) \, and \, C_i \neq C_j \\ 0, \quad otherwise. \end{cases} \quad (11)$$

If $\left[\varepsilon S_W + (1-\varepsilon)S_{LW}\right]$ is nonsingular, then Eq. (5) can be solved by similar way in LDA. In order to overcome the problem of the singularity, FF can be used. In this work, FF method is used for implementation.

In brief, CGLDA uses both global and local information of data. CGLDA uses three types of local information: Local similarity information, local intra-class pattern variation and local inter-class pattern variation. Zhang et al. (Zhang, et al., 2014) also proposed enhanced within-class LDA (EWLDA) for dimension reduction. EWLDA is a special case of CGLDA when α is set to 1.

3. PROPOSED METHOD

In this paper, a classification method based on a supervised dimension reduction, i.e. CGLDA, is proposed for hyperspectral images. The proposed method system architecture can be seen in Figure 1. In the proposed method, there are two inputs: Hyperspectral data and training samples. Training samples are used for two training processes. One of them is finding parameters of CGLDA and dimension reduction projection matrix W_{CGLDA}. The other training is classification training. The dimension of the hyperspectral data is reduced by using W_{CGLDA}. The classification process is applied on dimension reduced data and then classification map (output of the proposed method) is obtained.
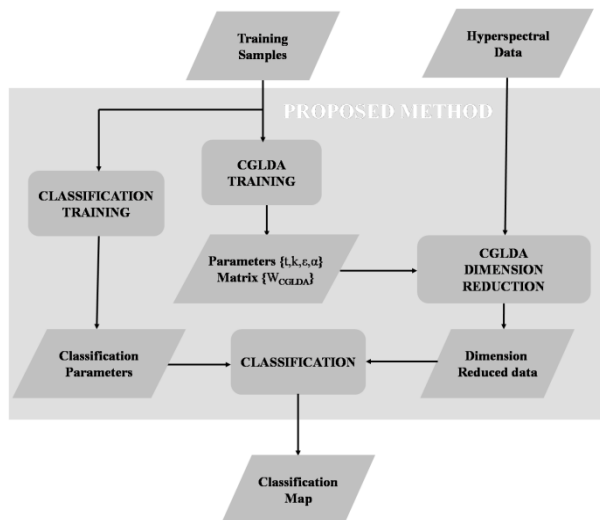

Figure 1. The proposed method system architecture

Classification process is also needed a special attention. K nearest neighbors (KNN) (Duda, et al., 2001), support vector machine (SVM) (Melgani & Bruzzone, 2004), (Mountrakis, et al., 2011), relevance vector machine (RVM) (Demir & Erturk, 2007) or any other classification method can be used for classification. The important point in use of hyperspectral data

reduced by CGLDA is that classification method should take into consideration global and local pattern variations.

Due to advantages of the usage of not only global pattern information, but also local pattern information, it can be expected that the proposed method gives good results and this statement is examined experimentally in Section 6. In addition, an experimental strategy is proposed in order to tune the parameters of CGLDA in hyperspectral application domain and it is demonstrated in Section 5. It must be also noticed that not only tuning the parameters, but also an experimental analysis of the distribution of the hyperspectral data is demonstrated. Therefore, how global or local pattern variations play an important role in classification is examined.

4. DATA SET, PERFORMANCE EVALUATION AND IMPLEMENTATION DETAILS

4.1 Data Set and Ground-truth

In this work, a well-known the AVIRIS Indian Pine data set is used (AVIRIS Indian Pine data set). 200 bands of 220 bands are used in the experiments. The low SNR bands [104-108, 150-163, 220] are removed (Huang, et al., 2012). The nine biggest classes are selected and the total number of samples is 9345.

4.2 Performance Evaluation

The classification map is compared with the ground-truth map and the number of correct classification pixel is divided by the total number of the used pixel ("percentage correct" in (Foody, 2002)). It measures the quality of the classification and ranges from 0 to1. Value of 1 confirms a perfect result.

4.3 Implementation Details

The proposed method is implemented using R (R Core Team, 2013). Package "raster" (Hijmans, 2014) is used for hyperspectral data read/write processes. Package "FNN" (Beygelzimer, et al., 2013) is used in the implementation of a function $N_k(x_i)$. Package "class" (Venables and Ripley, 2002) is used for KNN. Package "kernlab" (Karatzoglou, et al., 2004) is used for SVM.

5. SELECTION OF PARAMETERS

The proposed method is based on CGLDA and CGLDA has four parameters. It must be noticed that these four parameters can be affected from each other and it has a big complexity to tune these parameters in all together. There may be several approaches to tune process in an easy way. In this work, the following experimental strategy is proposed. At first, parameter t is examined while keeping ε, α and k remarkable fixed values. Hence, t is very important to see effects of other parameters. Next, parameter k is determined. Parameters α and ε are all together examined to tune process.

Randomly selected 1800 samples (200 for each class, ~19% of all data) are used in the parameter tuning processes. In each experiment, 900 samples (100 for each class) are randomly selected from these 1800 samples and tested on these 1800 samples. Experiments are randomly repeated 10 times and the averages of them are used. Euclidean distance calculation and KNN classifier is used for simplicity. KNN parameter is set to 7. Destination dimension size is set to 8.

Parameter t is a constant that controls how a similarity between two elements decreases against the distance. It is experimentally

determined while keeping ε=0.8, α=0.8 and k=10. Table 1 demonstrates KNN performance results of the proposed method with respect to different values of t. According to experiments, t is selected as 1.00E+07.

It can be easily seen from Eq. (8) and Eq. (11), t must be a comparable value with squares of nearest neighbour elements' distances. In order to check up any relation between t and the statistics of the classes, all nine classes' members are analyzed. Means and variances of dimensions of the classes are calculated. The sums of the dimension variances are seen in the range of between 1.15E+07 and 3.30E+07 and the weighted mean of them is 2.45E+07. It can be seen that the values of them and the experimentally calculated t are comparable values. Therefore, it can be told that before making a parameter tuning of t, sums of dimension variances of classes can be calculated, and then; these values can be used an initial estimate of t.

t	Performance
1.00E+05	0.8759
5.00E+06	0.9026
1.00E+07	0.9061
5.00E+07	0.9050
1.00E+08	0.9055

Table 1. Parameter tuning for parameter t

Parameter k is a constant that controls how many neighbours are used in a similarity calculation. It is experimentally determined while keeping ε=0.8, α=0.8 and t=1.00E+07.Table 2 demonstrates KNN performance results of the proposed method with respect to different values of k. According to experiments, k is selected as 10.

k	Performance
5	0.9013
10	0.9043
25	0.9031
40	0.8891
50	0.8785

Table 2. Parameter tuning for parameter k

Parameters ε and α are constants that control weights of effects for local and global information. In the experiments, they are changed in a range of 0.5 to 0.9 in order to keep away from over-fitting to the training data; and thus, a conservative approach for global information is preferred. It is experimentally determined while keeping t=1.00E+07 and k=10. Table 3 demonstrates KNN performance results of the proposed method with respect to different values of ε and α. According to experiments, ε is selected as 0.5 and α is selected as 0.8.

It can be easily seen from the value of ε that a conservative approach of local pattern variation within classes increases classification performances for this data set. This situation, of course, can change according to data sets; however, the tuning of parameter ε can model the importance of global or local information for classification purposes. It can be easily seen from the value of α that a conservative approach of global information among classes increases classification performances for this data set. This situation, of course, can change according to data sets; however, the tuning of parameter α can model importance of global or local information for classification purposes.

ε	α	Performance
0.5	0.5	0.9066
0.5	0.6	0.9108
0.5	0.7	0.9116
0.5	0.8	0.9139
0.5	0.9	0.9132
0.6	0.5	0.9053
0.6	0.6	0.9092
0.6	0.7	0.9094
0.6	0.8	0.9124
0.6	0.9	0.9126
0.7	0.5	0.9025
0.7	0.6	0.9057
0.7	0.7	0.9071
0.7	0.8	0.9097
0.7	0.9	0.9097
0.8	0.5	0.8987
0.8	0.6	0.9019
0.8	0.7	0.9023
0.8	0.8	0.9071
0.8	0.9	0.9074
0.9	0.5	0.8927
0.9	0.6	0.8941
0.9	0.7	0.8949
0.9	0.8	0.9017
0.9	0.9	0.9006

Table 3. Parameter tuning for parameters ε and α

6. RESULTS

For comparative analysis, five methods are compared each other: PCA, FF, EFDC, EWLDA and CGLDA. Parameter tuning is not made for EFDC and EWLDA. Because of similarities with CGLDA, the tuned parameters in CGLDA are used as follows: a=0.8, t=1.00E+07, k=10 for EFDC, ε=0.5, t=1.00E+07, k=10 for EWLDA. Randomly selected N, which is in the range of 25 to 400 with the step size of 25, samples for each class are used in the training processes. Tests are realized on all nine classes' data, i.e. total number of samples is 9345. Experiments are randomly repeated 10 times and the averages of them are used. Euclidean distance calculation and KNN classifier are used for simplicity. KNN parameter is set to 7. Destination dimension size is set to 8. The experimental results can be seen in Figure 2. According to experimental results, EWLDA and CGLDA show no significant change; thus, only CGLDA is drawn in the figure. Moreover, FF and EFDC results also show no significant change; thus, only FF is drawn in the figure. It can be easily seen from Figure 2 that the performances of CGLDA are the best except for the case of 25 sample size.

It can be seen form Figure 2 that CGLDA has the significant better performances than the other when the training sample size is increased. CGLDA and FF are similar method based on LDA except for CGLDA in which local information is used. The performance difference between CGLDA and FF is significant and the same experimental tools are used for fair comparison; and thus, it can be told that this difference is the most probable come from use of local information.

In the second comparative experiment, four methods are compared each other: PCA, EFDC, EWLDA and CGLDA. Randomly selected 100 samples for each class are used in the training processes. Tests are realized on all nine classes' data, i.e. total number of samples is 9345. Experiments are randomly repeated 10 times and the averages of them are used. Euclidean distance calculation and KNN classifier are used for simplicity. KNN parameter is set to 7. Destination dimension size is set to

between 8 and 50. The experimental results can be seen in Figure 3. According to experimental results, the performances of CGLDA are the best. It can be noted that EFDC also uses local information (Gao, et al., 2012) and it is also a recent used work in hyperspectral dimension reduction (Teke & Sakarya, 2013). On the other hand, CGLDA gives significantly better performance than EFDC. Moreover, another important note is CGLDA gives better results than EWLDA when the dimension size is increased. It can be inferred from these results that three types of local information modeling in CGLDA let an advantage for CGLDA against to the competitive methods in the experiments.
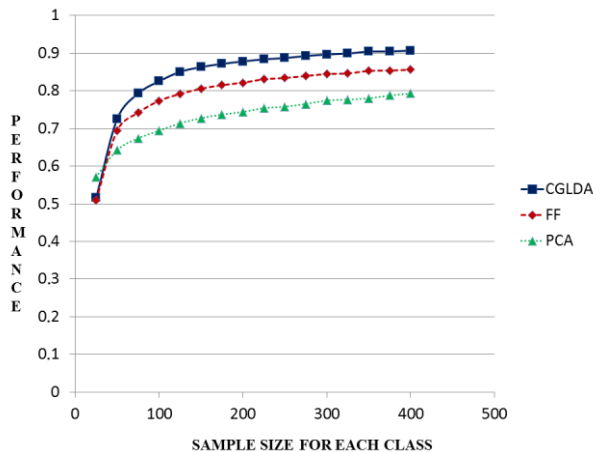

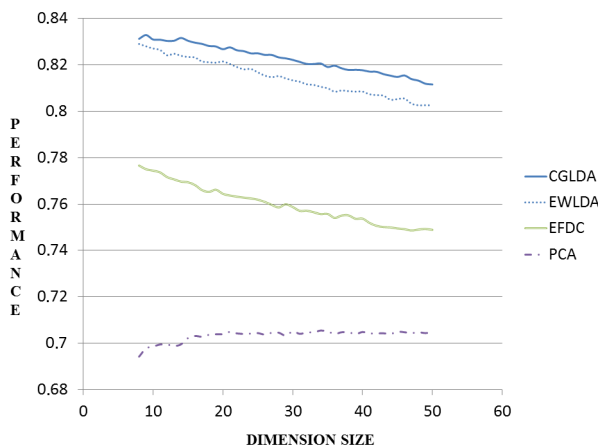Figure 2. Comparative performance analysis. Training samples versus performance.

Figure 3. Comparative performance analysis. Dimension size versus performance.

In the third experiment, KNN and SVM performance results of CGLDA are compared. According to Figure 3, KNN gives the best result in the case that dimension is set to 9. Therefore, firstly SVM parameters are tuned when the dimension is set to 9. Randomly selected 900 samples (100 for each class) are used in the parameter tuning processes. Tests are realized on all nine classes' data, i.e. total number of samples is 9345. Similar to experiment given in (Kozal, et al., 2013), SVM kernel is selected as radial bases function and the parameters "C" and "$sigma$" are examined from 2^{-10} to 2^{30}, from 2^{-60} to 2^{15} respectively. According to one experiment, C is selected as 512 and $sigma$ is selected as 0.00390625. In order to compare classification type scores, 900 samples (100 for each class) are randomly selected in the each experiment, and tested on all nine

classes' data. Experiments are randomly repeated 10 times and the averages of them are used. Euclidean distance calculation is used. KNN parameter is set to 7. Destination dimension size is set to 9. The parameters of CGLDA are set as follows: α =0.8, ε=0.5, t=1.00E+07, k=10. The average performance result of KNN is obtained as 0.8304 and the average performance result of SVM is obtained as 0.7954. It must be noted that the experiment is made according to the tuned parameters in KNN. On the other hand, CGLDA can also be examined for different dimensions and parameters in order to get better performances in SVM. This work could be a future research issue. One of the performance results of the proposed method using KNN (dimension = 9) can be seen in Figure 4.
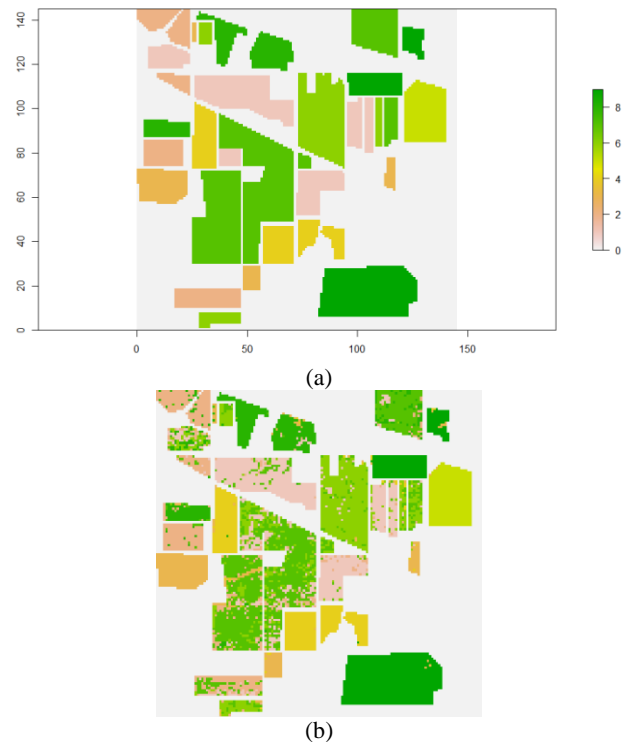

(a)

(b)

Figure 4. (a) Ground-truth (b) One of the performance results of the proposed method (the percentage correct is 0.8275).

7. CONCLUSIONS

In this paper, a classification method based on CGLDA is proposed for hyperspectral images. Advantages of the usage of not only global pattern information, but also local pattern information are experimentally demonstrated by comparative analysis. In addition, not only tuning the parameters, but also an experimental analysis of the distribution of the hyperspectral data is demonstrated. Therefore, how global or local pattern variations play an important role in classification is examined. It can be concluded from the experimental results that three types of local information modeling in CGLDA let an advantage for CGLDA against to the competitive methods in the experiments.

For future research directions, some topics can be thought about. The proposed framework can also be examined for different types of hyperspectral data such as thermal hyperspectral. In addition, it must be noticed that CGLDA is a linear method so new investigations are still researched at non-linear adaptation of CGLDA for some non-linear problems (Zhang, et al., 2014).

REFERENCES

AVIRIS Indian Pine data set, [Online] Available at: ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/92AV3C.tif.zip (data), ftp://ftp.ecn.purdue.edu/biehl/MultiSpec/ThyFiles.zip (ground truth), [Accessed 07 May 2012].

Belhumeur, P., Hespanha, J. and Kriegman, D., 1997. Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Jul, 19(7), pp. 711-720.

Beygelzimer, A., Kakadet, S., Langford, J., Arya, S., Mount, D. and Li, S., 2013. FNN: Fast nearest neighbor search algorithms and applications. R package version 1.1. http://CRAN.R-project.org/package=FNN

Camps-Valls, G., Tuia, D., Bruzzone, L. and Atli Benediktsson, J., 2014. Advances in hyperspectral image classification: Earth monitoring with statistical learning methods. *IEEE Signal Processing Magazine,* 31(1), pp. 45-54.

Cheriyadat, A. and Bruce, L., 2003. Why principal component analysis is not an appropriate feature extraction method for hyperspectral data. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS '03*, pp. 3420-3422, vol.6.

Demir, B. and Erturk, S., 2007. Hyperspectral image classification using relevance vector machines. *IEEE Geoscience and Remote Sensing Letters,* 4(4), pp. 586-590.

Duda, R., Hart, P. and Stork, D., 2001. *Pattern Classification 2nd ed..* A Wiley-Interscience Publication, John Wiley & Sons, Inc..

Du, Q., 2007. Modified Fisher's linear discriminant analysis for hyperspectral imagery. *IEEE Geoscience and Remote Sensing Letters,* 4(4), pp. 503-507.

Du, Q. and Chang, C.-I., 2001. A linear constrained distance-based discriminant analysis for hyperspectral image classification. *Pattern Recognition ,* 34(2), pp. 361-373.

Fauvel, M., Tarabalka, Y., Benediktsson, J., Chanussot, J. and Tilton, J., 2013. Advances in spectral-spatial classification of hyperspectral images. *Proceedings of the IEEE,* 101(3), pp. 652-675.

Foody, G. M., 2002. Status of land cover classification accuracy assessment. *Remote Sensing of Environment ,* 80(1), pp. 185-201.

Gao, Q., Liu, J., Zhang, H., Hou, J. and Yang, X., 2012. Enhanced Fisher discriminant criterion for image recognition. *Pattern Recognition,* 45(10), pp. 3717-3724.

Hijmans, R.J., 2014. raster: raster: Geographic data analysis and modeling. R package version 2.2-5. http://CRAN.R-project.org/package=raster

Huang, H., Liu, J. and Pan, Y., 2012. Semi-supervised marginal Fisher analysis for hyperspectral image classification. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences,* Volume I-3, 2012, XXII ISPRS Congress, 25 August – 01 September 2012, Melbourne, Australia.

Karatzoglou, A., Smola, A., Hornik, K. and Zeileis, A., 2004. kernlab - An S4 package for kernel methods in R. *Journal of Statistical Software* 11 (9), pp. 1-20. URL http://www.jstatsoft.org/v11/i09/

Kozal, A., Teke, M. and Ilgin, H., 2013. Comparative analysis of hyperspectral dimension reduction methods. *IEEE 21st Signal Processing, Communication and Applications Conference,* Girne, KKTC, April 24-26, 2013.

Kuo, B.-C. and Landgrebe, D., 2004. Nonparametric weighted feature extraction for classification. *IEEE Transactions on Geoscience and Remote Sensing,* May, 42(5), pp. 1096-1105.

Li, D.-L., Prasad, M., Hsu, S.-C., Hong, C.-T. and Lin, C., 2012. Face recognition using nonparametric-weighted Fisherfaces. *EURASIP Journal on Advances in Signal Processing.*

Melgani, F. and Bruzzone, L., 2004. Classification of hyperspectral remote sensing images with support vector machines. *IEEE Transactions on Geoscience and Remote Sensing,* Aug, 42(8), pp. 1778-1790.

Mountrakis, G., Im, J. and Ogole, C., 2011. Support vector machines in remote sensing: A review. *ISPRS Journal of Photogrammetry and Remote Sensing,* 66(3), pp. 247-259.

Plaza, A., Benediktsson, J.A., Boardman, J.W., Brazile, J., Bruzzone, L., Camps-Valls, G., Chanussot, J., Fauvel, M., Gamba, P., Gualtieri, A., Marconcini, M., Tilton, J.C. and Trianni, G., 2009. Recent advances in techniques for hyperspectral image processing. *Remote Sensing of Environment,* 113, Supplement 1(0), pp. S110 - S122.

R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL http://www.R-project.org/.

Teke, M., Deveci, H., Haliloglu, O., Gurbuz, S. and Sakarya, U., 2013. A short survey of hyperspectral remote sensing applications in agriculture. *6th International Conference on Recent Advances in Space Technologies (RAST 2013),* pp. 171-176.

Teke, M. and Sakarya, U., 2013. Classification on hyperspectral images using enhanced Fisher discriminant criterion. *IEEE 21st Signal Processing, Communication and Applications Conference,* Girne, KKTC, April 24-26, 2013.

Venables, W.N. and Ripley, B.D., 2002. *Modern Applied Statistics with S.* Fourth Edition. Springer, New York. ISBN 0-387-95457-0

Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q. and Lin, S., 2007. Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* 29(4), pp. 40-51.

Zhang, D., He, J., Zhao, Y., Luo, Z. and Du, M., 2014. Global plus local: A complete framework for feature extraction and recognition. *Pattern Recognition ,* 47(3), pp. 1433-1442.