Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

CLASSIFICATION OF AURORA SERIES IMAGE BASED ON EIGENVALUE-SCALING KERNEL FISHER DISCRIMINANT ANALYSIS AND KLT

[1]SHAN LU, [1]LICHENG JIAO, [1]JIAJI WU, [1]XIAOZHENG DENG
[1]Key Laboratory of Intelligent Perception and Image Understanding of Ministry of Education of China, Xidian University, Xi'an 710071, P.R. China, lumaomaommbb@126.com
E-mail: [1]lumaomaommbb@126.com

Abstract

Aurora is the typical ionosphere track generated by the interaction of solar wind and magnetosphere. This paper proposed a method based on eigenvalue-scaling kernel fisher discriminant analysis and Karhunen-Loeve Transform (KLT) to take advantage of interband correlation between aurora images to detect the change of aurora in serial time. Conventional classification algorithms are incapable of attaining the desired classification accuracy, and the feature of redundancy of images in a close time sequence is not considered. To solve the problems, we first apply KLT to reduce the spectral redundancies and wavelet transform to extract eigenvalues of the spatial domain. Then we employ eigenvalue-scaling kernel fisher discriminant analysis which is a modified kernel fisher discriminant analysis to realize the desired classification accuracy. Experiments carry out on time series image pointed out the effectiveness of the presented technique, which results in an increase of the classification accuracy with respect to conventional algorithms.

Keywords: *Aurora Series Image, Eigenvalue-Scaling Kernel Fisher Discriminant Analysis, Classification, Decorrelation.*

1. INTRODUCTION

Aurora reflects the interaction between solar wind and magnetosphere, and shape、position and brightness of aurora change rapidly. Aurora series image refers to an aggregate of images of the same scene captured in serial time and captured by the same imaging pattern, and the detection of the trend of auroral change is significant to the study of space weather activity. Aurora series image classification is necessary and important for image analysis and useful information extraction. Similar to multispectral image, because of the featrue of high dimensions, the classification of time series image is different from the classification of ordinary image. On the one hand, the dimensionality of input space strongly affects the performance of many supervised classification methods [1], however, while, on the other hand, due to the large amount of collected data and redundancy, it significantly increases the complexity of the analysis [2]. Nowadays, most researches on time series image classification technique are focused on kernel-based methods (KMs). They have demonstrated excellent performance in

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

hyperspectral data classification in terms of accuracy and robustness [3]–[8].

KMs are based on mapping data from the original input feature space to a kernel feature space of higher dimensionality, and then solving a linear problem in that space. These methods allow us to interpret learning algorithms geometrically in the kernel space (which is nonlinearly related to the input space), thus combining statistics and geometry in an effective way. The properties of kernel methods are well-suited to tackle the problem of time series image classification since they can handle large input spaces efficiently, work with a relatively low number of labeled training samples and deal with noisy samples in a robust way. Both support vector machines (SVMs) [9], [10] and kernel Fisher discriminant analysis (KFDA) [11], [12] can be integrated in the so-called kernel methods framework. An optimization algorithm for solving quadratic problem is given, the data obtained from this algorithm is global optimal solution. SVMs can deal with noisy samples in a robust way and produce sparse solution, and it expresses the model that defines the decision boundary as a function of a subset of training samples. KFDA takes advantage of the same concept of kernel used in SVMs to obtain nonlinear solutions. It can directly estimate the conditional posterior probabilities of classes and the key challenge in KFDA lies in the selection of free parameters such as kernel parameters and regularization parameters. The advantages of KFDA are very effectively for classification. But the geometrical interpretation reveals that KFDA differs from SVMs in that it maximizes the average margin instead of the smallest margin. It forces KFDA to include in the solution all the training samples and thus the important property of sparsity is lost. As we will see, this is a dramatic problem when working with moderate or high number of labeled samples, inducing problems of high computational cost and memory requirements. This is a particularly relevant impairment for serial applications. In addition, the feature of high dimensions and redundancy of the serial image is not sufficiently considered by the common KMs, which degrade the classification performance.

According to the previous observations, in this paper we present a novel classification algorithm based on eigenvalue-scaling kernel Fisher discriminant analysis (ES-KFDA) and klt. In particular, the novel technique properly integrates the following: 1) ES-KFDA retains the advantage of KFDA which are very effectively for classification; 2) In ES-KFDA, each feature has its own scaling factor, if some feature is insignificant or irrelevant for classification, the associated scaling factor will be set smaller and share the same scaling factors. So it improves the performance of KFDA in the presence of many irrelevant features and obtains improved sparsity; 3) In ES-KFDA, all the free parameters are analytically chosen, so the learning process is fully automatic; 4) KLT is used to reduce redundancy, so it can restrain the negative influence of the correlation and improve the classification accuracy.

This letter is organized as follows. Section II introduces the decorrelation process realized by KLT. Section III briefly reviews the formulation of KFDA classifiers and presents the formulation of ES-KFDA classifiers. Section IV presents the specific application of ES-KFDA to time series image analysis and the experimental results. In Section V, we conclude this paper with a discussion.

2. KARHUNEN-LOEVE TRANSFORM

Karhunen-Loeve Transform (KLT) belongs

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

to multivariate statistical analysis. From the characteristics of statistics of image, we apply a group of parameters which are not related to denote the continuous signals and realize orthogonal transform, under the rule of mean square error, and it is the transform of the smallest distortion, so it is called optimal transform. After transform, greater variances only exist in few coefficients, under allowed distortion; it is possible to compress the image data to the minimum. Transform process is as follows:

1) The image can be expressed as

$X=[X_1,X_2,\ldots X_j],1\le j\le n$, Where

$$X_j=[x_1,x_2,\ldots x_i]^T,1\le i\le m \quad (1)$$

Then standard evaluation matrix can be calculated as

$$Y=[y_{ij}],\text{ where } y_{ij}=(x_{ij}-\overline{x_j})/s_j \quad (2)$$

mean value is denoted by $\overline{x_j}$ and standard deviation is denoted by s_j.

2) The covariance matrix can be calculated as

$S=\frac{1}{n}\left[Y-\overline{Y}l\right]\left[Y-\overline{Y}l\right]^T$, where

$$l=[1,1,\ldots 1]_{1\times n},\quad \overline{Y}=[\overline{y_1},\overline{y_2},\ldots \overline{y_m}]^T, \quad (3)$$

$$\overline{y_i}=\frac{1}{n}\sum_{k=1}^{n}y_{ik} \quad (4)$$

3) We can easily find the following characteristic equation

$$(\lambda I-S)U=0, \quad (5)$$

From which eigenvalue λ and eigenvector u can be found and transform matrix T can be constructed.

4) A new set of image is denoted by X_{new}, it can be expressed as follows:

$$X_{new}=TX \quad (6)$$

3. FEATURE-SCALING KERNEL FISHER DISCRIMINANT ANALYSIS

In this section, we briefly review KFDA and discuss the detailed formulation of ES-KFDA.

3.1 Kernel Fisher Discriminant Analysis

The idea of the KFDA is to solve the well-known problem of Fisher's linear discriminant in a kernel feature space, which produces a nonlinear discriminant classifier in the input space. In KFDA, between-class scatter matrix and within-class scatter matrix are defined by $sm_b^F=(n_1^F-n_2^F)(n_1^F-n_2^F)^T$ and $sm_\omega^F=\sum_{i=1}^{2}\sum_{j=1}^{l_i}(\phi(x_j^i)-n_i^F)(\phi(x_j^i)-n_i^F)^T$, where the mean n_i^F of ith class is $n_i^F=\frac{1}{l_i}\sum_{j=1}^{l_i}\phi(x_j^i)$. An optimal transformation Q is given by maximizing the between-class scatter while simultaneously minimizing the within-class scatter:

$$\max_Q\left(\frac{Q^T sm_b^F Q}{Q^T sm_\omega^F Q}\right), \quad (7)$$

Q can be formulated as $Q=\sum_{j=1}^{l}\alpha_j\phi(x_j)$. So we can calculate α by

$$\max_\alpha\left(\frac{\alpha^T \overline{sm_b}^F\alpha}{\alpha^T \overline{sm_\omega}^F\alpha}\right), \quad (8)$$

Where $\overline{sm_b}^F=(\overline{n}_1^F-\overline{n}_2^F)(\overline{n}_1^F-\overline{n}_2^F)^T$ and

$\overline{sm_\omega}^F=\sum_{i=1}^{2}\sum_{j=1}^{l_i}(\delta_j^i-\overline{n}_i^F)(\delta_j^i-\overline{n}_i^F)^T$

with $\delta_j^i=[k(x_1,x_j^i),\ldots,k(x_l,x_j^i)]^T$ and

Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

$$\bar{n}_i^F = \frac{1}{l_i}[\sum_{j=1}^{l_i} k(x_1, x_j^i), \ldots, \sum_{j=1}^{l_i} k(x_l, x_j^i)]^T .$$

For a new sample x, we can predict its label by

$$g(x) = \mathrm{sgn}((Q \cdot \phi(x)) + b) = \mathrm{sgn}(\sum_{j=1}^{l} \alpha_j k(x_j, x) + b),$$

Where $b = -\alpha^T \frac{l_1 \bar{n}_1^F + l_2 \bar{n}_2^F}{l}$. (9)

3.2 Eigenvalue-Scaling Kernel Fisher Discriminant (Es-Kfd) Analysis

Motivated by a closed form of the leave-one-out error of KFD which was given by Cawley and Talbot (2003) and feature scaling for kernel fisher discriminant Analysis which was presented by Liefeng Bo (2005), ES-KFD is constructed in three steps: replacing the kernel function in the leave-one-out error of KFD with a feature-scaling kernels function, then optimizing the resulting leave-one-out error via a gradient-descent algorithm and making the neighboring features share the same scaling factors, finally, predicting the label of a new sample via the decision function, it is expressed as

$$g(x) = \mathrm{sgn}(\sum_{j=1}^{l} \alpha_j k_\theta(x_j, x) + b), \quad (10)$$

where the coefficients are determined by minimizing the following loss function:

$$f(\bar{\alpha}) = \bar{\alpha}^T (W^T W + \lambda U)\bar{\alpha} - 2\bar{\alpha}^T W^T Y + Y^T Y, \quad (11)$$

where $\bar{\alpha} = \begin{bmatrix} \alpha \\ b \end{bmatrix}$, $W = [K \quad 1]$, $U = [\begin{matrix} I & 0 \\ 0^T & 0 \end{matrix}]$,

and I denotes the unit matrix. The coefficients are determined by minimizing the leave-one-out error, it is expressed as

$$lot(\theta, \lambda) = \frac{1}{l}\sum_{i=1}^{l}(\frac{1 - y_i \tanh(\gamma(y_i - r_i))}{2}), \quad (12)$$

where $r = (I - H)y \circ (1 - D(H))$,

$$H = C(C^T C + \lambda U)^{-1} C^T,$$

$$\tanh(\gamma t) = \frac{\exp(\gamma t) - \exp(-\gamma t)}{\exp(\gamma t) + \exp(-\gamma t)}.$$

We use a gradient-descent method to minimize this estimate.

According to the chain rule, the derivative of $lot(\theta, \lambda)$ is formulated as $\frac{\partial(lot(\theta,\lambda))}{\partial \theta_k} = \frac{\partial(lot(\theta,\lambda))}{\partial r^T}\frac{\partial r}{\partial \theta_k}$, we need only to calculate $\frac{\partial(lot(\theta,\lambda))}{\partial r^T}$ and $\frac{\partial r}{\partial \theta_k}$ respectively. Finally, we can compute the derivative of the leave-one-out error with respect to θ_k.

Two of the most popular feature-scaling kernels are polynomial kernel and Gaussian kernel, as given below:

$$K_\theta(x_i, x_j) = (1 + \sum_{k=1}^{d} \theta_k x_i^{(k)} x_j^{(k)})^r, \quad (13)$$

$$K_\theta(x_i, x_j) = \exp(-\sum_{k=1}^{d} \theta_k \| x_i^{(k)} - x_j^{(k)} \|^2), \quad (14)$$

In a feature-scaling kernel, each feature has its own scaling factor. If some feature is insignificant or irrelevant for classification, the associated scaling factor will be set smaller, otherwise, it will be set larger.

θ_k and λ are automatic chosen, but the optimal value for γ is difficult, we try several different values for γ in order to choose the one leading to the smallest leave-one-out error. As γ was changed from 0 to 400, an ROC curve could be estimated. The resulting ROCs with leave-one-out (lot) error and value of γ are shown in Fig.1. The optimal value can be obtained from Fig.1, the minimum of error corresponds to $\gamma = 10$.

Journal of Theoretical and Applied Information Technology
10th April 2013. Vol. 50 No.1

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

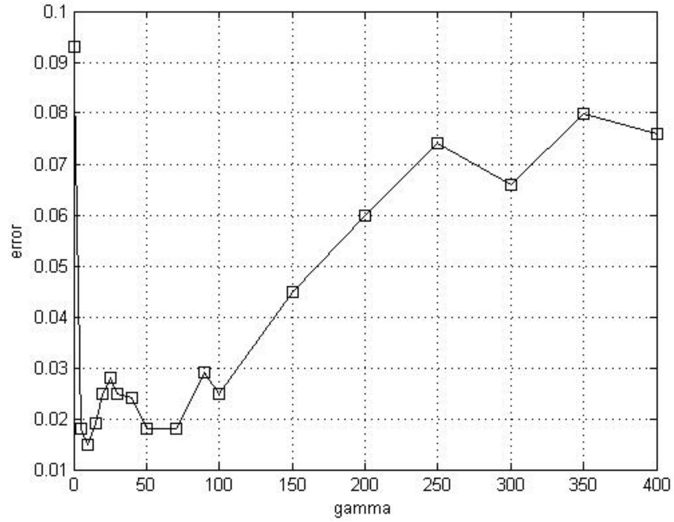
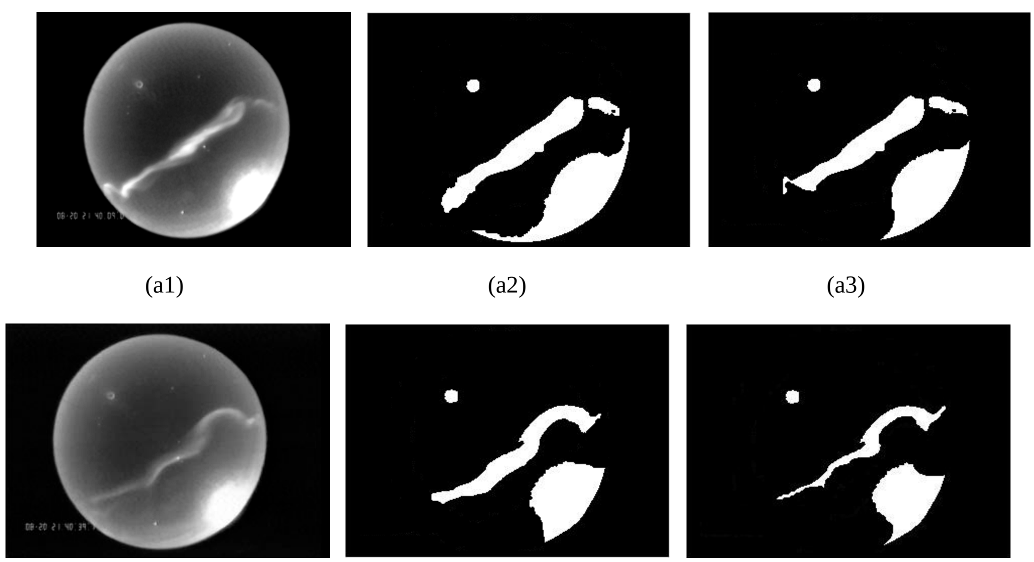


Fig.1 Rocs With Leave-One-Out (Lot) Error And Value Of γ

According to the result, in our following experiment, we set $\gamma = 10$.

4. EXPERIMENTAL ANALYSIS AND RESULTS

Simulations of classification performance have been carried out on the aurora serial image to test the performance of the proposed method. Results are reported for two images. Two different classes available in the original images, we divide the images into two classes: aurora and background. The classes were used to generate a set of 562 training samples (for the learning phase of the classifiers) and 530 testsamples (for validating their performance). See Table.1 for details. In experimental analysis, we consider SVMs and KFD.

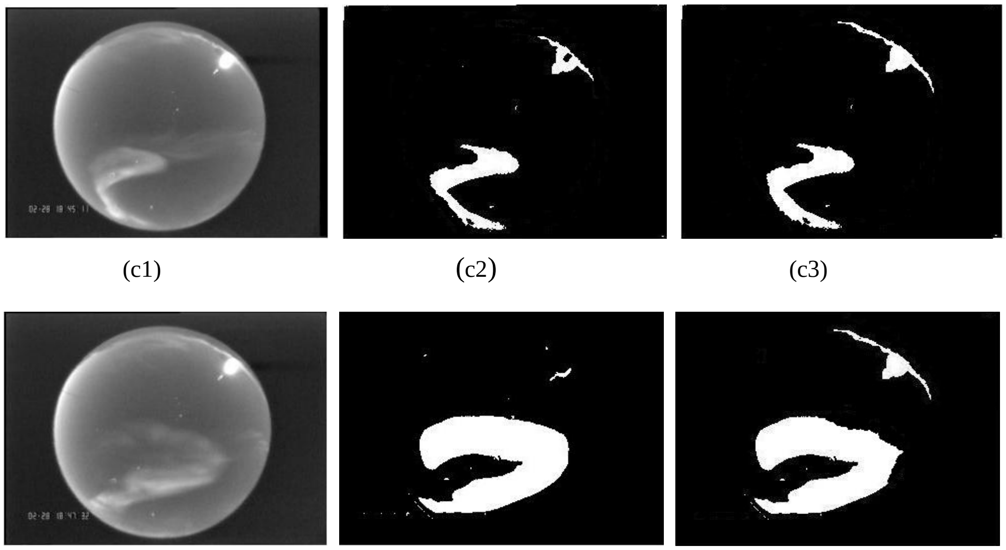


(a1) (a2) (a3)

(b1) (b2) (b3)

Fig.2 (A1-A3) Reference Map Of Bands 60 Of Scene1, Examples Of Classification Without Inter-Frame Correlation, Examples Of Classification With Inter-Frame Correlation. (B1-B3) Reference Map Of Bands 68 Of Scene1, Examples Of classification without inter-frame correlation, Examples of classification with inter-frame correlation.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195



(c1) (c2) (c3)

(d1) (d2) (d3)

Fig.3 (C1-C3) Reference Map Of Bands 60 Of Scene2, Examples Of Classification Without Inter-Frame Correlation, Examples Of Classification With Inter-Frame Correlation. (D1-D3) Reference Map Of Bands 68 Of Scene2, Examples Of Classification Without Inter-Frame Correlation, Examples Of Classification With Inter-Frame Correlation.

Table1 shows the classification accuracy of scene 1 and scene 2, that we take into account spectral information by computing KLT matrix and spatial information by simply computing the eigenvalue in a 7 × 7 window based on wavelet transform. Table2 shows the classification accuracy of scene 1 and scene 2 without considering the redundancy of spectral information. In the two situations, we used the Gaussian kernel (RBF) in the case of SVMs and KFDA. Furthemore, we tested the polynomial kernel, Line kernel and Sigmoid kernel for SVMs.

. Table.1 Mean Results In The Dataset (168 Input Bands) With KLT

		scene1(OA[%])	scene2(OA[%])
with KLT	SVM+Line	85.23	85.05
	SVM+Poly	93.95	93.07
	SVM+Sigmoid	89.56	88.89
	SVM+RBF	93.23	92.93
	KFDA	94.21	93.86
	ES-KFDA	**96.23**	**95.36**

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table.2 Mean Results In The Dataset (168 Input Bands) Without KLT

		scene1(OA[%])	scene2(OA[%])
without KLT	SVM+Line	82.11	81.97
	SVM+Poly	92.18	91.76
	SVM+Sigmoid	87.21	86.93
	SVM+RBF	91.86	91.67
	KFDA	91.45	91.04
	ES-KFDA	**93.10**	**92.62**

The improved algorithm can get better results, because that unlike traditional method try to reduce the dimension of input space, the improved algorithm increases dimension of input space to ensure it is linearly separable in high dimension space, rising dimension only changes inner product operation and the complexity of algorithm does not increase, so the Hughes phenomenon in series analysis is avoided, classification accuracy can retain at high level.

The results confirm the effectiveness of the proposed ES-KFDA corporated with KLT in addressing classification of series images. On the one hand, in general, kernel-based methods are suitable for series images classification. On the other hand, it comes out that, besides the employment of kernel composition, the performance of classification benefit from the improvement of eigenvalue and eigenvector.

5. CONCLUSION

In this paper, we defined a novel supervised kernel classifier based on kernel fisher discriminant analysis (KFDA) designed for addressing classification of series images. This novel method is developed to tune the scaling factors for the eigenvalue-scaling kernel and reduce the spectral redundancies. It focuses on the eigenvalue-scaling kernel where each feature of eigenvector individually associates with a scaling factor. In this way, we can handle some problems of KFDA.

Experiments on the aurora series data set show that the proposed method has the advantages of low computational cost and good classification performance.

Future work will concentrate on extending the experimental analysis to more different methods to describe the classification accuracy, in order to confirm the effectiveness of the proposed method.

This work was supported by the National Natural Science Foundation of China (No. 61001202 and 61003199), the National Research Foundation for the Doctoral Program of Higher Education of China (No. 200807010003, 20090203120016 and 20100203120008).

REFERENCES

[1] Yuhua Feng, "Analysis on Algorithm and Application of Cluster in Data Mining," *Journal of Theoretical and Applied Information Technology*, Vol. 46, No. 1, 2012, pp. 416-419.

[2] K. Fukunaga, "Effects of Sample Size in Classifier Design," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 8, 1989, pp. 873-885.

[3] Mohamed Ettaouil, Mohamed Lazaar and Youssef Ghanou, "Architecture Optimization Model for the Multilayer Perceptron and Clustering", *Journal of Theoretical and Applied Information Technology*, Vol: 47, No. 1, 2013, pp. 64-72.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[4] K. In Kim, K. Jung and H. Joon Kim, "Face Recognition Using Kernel Principal Component Analysis", *IEEE Signal Processing Letters*, Vol. 9, No. 2, 2002.
[5] S. Mika, Jason Weston, Klaus-Robert Müller, "Fisher Discriminant Analysis With Kernels", *Neural Networks for Signal Processing IX*, 1999.
[6] Christopher J C Burges, "A Tutorial on Support Vector Machines for Pattern Recognition", *Bell Laboratories, Lucent Technologies*, 1998.
[7] B. Scholkpof, A. Smola, and K. R. Muller, "Nonlinear Component Analysis as a kernel Eigenvalue Problem", *Tubingen, Germany, Tech. Report,* No 44, 1996.
[8] T. Joachims, "Making large-Scale S VM L earning Practical. Advances in Kernel Methods - Support Vector Learning", *C. Burges and A. Smola (ed.), MIT-Press,* 1999.
[9] C. W. Hsu and C. J. Lin, "A Comparison of Methods for Multiclass Support Vector Machines", *IEEE Transactions on Neural Networks,* VOL. 13, NO. 2, 2002.
[10] C.J.C. Burges, "A Tutorial on Support Vector Machines for Pattern Recognition," *Data Mining and Knowledge Discovery,* 1998.
[11] S. Mika, G. Ratsch, J. Weston, "Fisher Discriminant Analysis with Kernels", *Neural Networks for Signal Processing IX,* 1999, pp. 41-48.
[12] S.A. Billings, K.L. Lee, "Nonlinear Fisher discriminant analysis using a minimum squared error cost function and the orthogonal least squares algorithm", *Neural Networks,* Vol. 15, No. 2, 2002, pp. 263–270.