Research Article

Hyperspectral Image Classification Using Kernel Fukunaga-Koontz Transform

Semih Dinç¹ and Abdullah Bal²

¹ Department of Control and Automation Engineering, Yildiz Technical University, Esenler, 34220 Istanbul, Turkey

² Electronics and Communications Engineering, Yildiz Technical University, Esenler, 34220 Istanbul, Turkey

Correspondence should be addressed to Abdullah Bal; bal@yildiz.edu.tr

Received 6 May 2013; Accepted 18 September 2013

Academic Editor: Yudong Zhang

Copyright © 2013 S. Dinç and A. Bal. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper presents a novel approach for the hyperspectral imagery (HSI) classification problem, using Kernel Fukunaga-Koontz Transform (K-FKT). The Kernel based Fukunaga-Koontz Transform offers higher performance for classification problems due to its ability to solve nonlinear data distributions. K-FKT is realized in two stages: training and testing. In the training stage, unlike classical FKT, samples are relocated to the higher dimensional kernel space to obtain a transformation from non-linear distributed data to linear form. This provides a more efficient solution to hyperspectral data classification. The second stage, testing, is accomplished by employing the Fukunaga-Koontz Transformation operator to find out the classes of the real world hyperspectral images. In experiment section, the improved performance of HSI classification technique, K-FKT, has been tested comparing other methods such as the classical FKT and three types of support vector machines (SVMs).

1. Introduction

In the last decade, hyperspectral remote sensing technology has been included in popular study issues. Many articles have been proposed regarding hyperspectral images and spectral analysis since it offers new insight to various application areas such as agriculture [1], medical diagnose [2], illegal drug field detection [3], face recognition [4], and military target detection [5].

The idea behind the remote sensing technology relies on the relationship between photons and surface materials. Hyperspectral images are captured by the spectral sensors which are sensitive to larger portion of the electromagnetic spectrum than the traditional color cameras. While a digital color camera can capture only 3 bands (Red, Green, and Blue) in the range of 400 nm to 700 nm spectral wavelength, a typical hyperspectral sensor captures more than 200 bands within the range of 400 nm to 2500 nm. This means that HSI offers 200 or more features for an image pixel, instead of 3 values. HSI contains diverse information from a wide range of wavelengths. This characteristic yields more effective classification power for the application areas mentioned above.

Different type of materials can be represented by a set of bands which is called “spectral signature” that simplifies the separation of these materials.

There are also several challenges of the HSI to be solved. For instance, water absorption and some other environmental effects may induce some spectral bands to be noisy. These specific bands are called “noisy bands” and are required to be removed from the dataset. Another problem is the size of the data. Even in small scenes, hyperspectral images may have much larger size than traditional gray scale and color images, which means that the processing time is also longer than usual images. Detection of the redundant bands and removing them is crucial to reduce the number of features and total processing time. For this purpose, we refer to two papers [6, 7] to select best informative bands of our dataset.

In the literature, there are various classification techniques proposed for hyperspectral image classification problem, including neural networks, support vector machines, and Bayesian classifiers. In 2005, Benediktsson et al. [8] proposed a solution based on extended morphological models and neural networks. In 2007, Borges et al. [9] published their studies which is based on discriminative class learning

FIGURE 1: (a) RGB image of dataset, band numbers R = 24, G = 14, and B = 8. (b) Ground truth data.

FIGURE 2: Ground truth data with selected classes.

using a new Bayesian based HSI segmentation method. In 2008, Alam et al. [10] proposed a Gaussian filter and post processing method for HSI target detection. Samiappan et al. [11] introduced a SVM based HSI classification approach which uses the same dataset used in this paper.

In this study, we present Kernel based Fukunaga-Koontz Transform method which is a novel solution to hyperspectral image classification. Classical FKT is a powerful method to solve two-pattern classification problems. However, when data is more complicated, classical FKT cannot produce satisfactory results. In this case, kernel transformations help FKT to increase separability of the data. In order to evaluate the K-FKT algorithm performance on HSI classifications, we select AVIRIS hyperspectral dataset which is a benchmark problem of this area. We have also used some other HSI dataset in our earlier studies and obtained high accuracy results which are presented in [12].

The remainder of this paper is organized as follows. The following section gives information about the contents of the AVIRIS dataset. A detailed description of the Kernel

Fukunaga-Koontz Transform is presented in Section 3 with training and testing stages. Classification results are given in Section 4, including the comparison with other methods. In the last section, we conclude our paper.

2. AVIRIS Dataset

This section includes detailed information about the AVIRIS Hyperspectral Image dataset which is called “Indian Pines” [13]. The dataset contains several different areas. Among these, there are mostly agricultural crop fields. Remaining parts have forests, highways, a rail line, and some low density housing. A convenient RGB colored view of the image can be seen in Figure 1(a).

Indian Pines dataset contains 16 different classes of crop fields. The ground truth data is shown in Figure 1(b). Table 1 shows the names of the classes and total number of samples for each class.

Basically, our dataset is $145 \times 145 \times 220$ matrix that corresponds to 220 different bands of images having size of 145×145 . In order to have more convenient form, this 3D matrix is transformed into 2D form as 21025×220 matrix which indicates 21025 samples, and each sample has 220 numbers of features.

Before the classification processing, we removed regions that do not correspond to any class (dark blue areas in Figure 1(b)) from the dataset. Almost half of the samples do not belong to one of 16 classes. Once we eliminate these redundant bands, only 10336 samples are kept in the dataset.

3. Kernel Fukunaga-Koontz Transform

Traditional Fukunaga-Koontz Transform is a statistical transformation method which is a well-known approach [14–17], for two class classification problems. Basically, it operates by transforming data into a new subspace where both classes share the same eigenvalues and eigenvectors. While a subset of these eigenvalues can best represent ROI class, the

TABLE 1: Class names and number of samples.

Class number	Class	Samples
1	Alfalfa	54
2	Corn-notill	1434
3	Corn-mintill	834
4	Corn	234
5	Grass-pasture	497
6	Grass-trees	747
7	Grass-pasture-mowed	26
8	Hay-windrowed	489
9	Oats	20
10	Soybean-notill	968
11	Soybean-mintill	2468
12	Soybean-clean	614
13	Wheat	212
14	Woods	1294
15	Buildings-Grass-Trees-Drives	380
16	Stone-Steel-Towers	95
Total		10366

remaining eigenvalues represent the clutter class. With this characteristic, FKT differs from other methods.

Traditional FKT has been proposed to solve linear classification problems. When the data is nonlinearly distributed, the classical approach is not the best solution. Like other linear classifiers such as linear discriminant analysis (LDA), independent component analysis (ICA), and support vector machines (SVM), classical FKT suffers from nonlinearly distributed data. Therefore, in this paper, we used an improved version of classical FKT which basically changes the data distribution with a Kernel transformation to classify nonlinearly distributed data in a linear classification fashion. We will call it “K-FKT” in the rest of the paper. K-FKT algorithm consists of two stages: training and testing.

3.1. Training Stage. Since Fukunaga-Koontz Transform is a binary classification method, the training dataset is divided into two main classes. The region of interest (ROI) class is the first one to be classified. And the clutter class contains all other classes in the dataset except ROI. The algorithm is initiated by collecting an equal number (N) of training samples for ROI and clutter classes that are represented as X and Y , respectively. Similarly, x_i and y_i are the training samples (or training signatures) of ROI and clutter classes as follows:

$$\begin{aligned} X &= [x_1, x_2, \dots, x_N], \\ Y &= [y_1, y_2, \dots, y_N]. \end{aligned} \quad (1)$$

The training sets X and Y are first normalized to avoid unexpected transformation results. Then they are mapped into higher dimensional Kernel space via the kernel transform. In simple terms, we assume that there is a mapping function $\phi()$ to map all training samples to the Kernel space.

In this manner, we would obtain new training sets $\tilde{X}$ and $\tilde{Y}$ in which $\phi(x_i)$ and $\phi(y_i)$ denote the training samples in Kernel space. Equation (2) shows the mapping process. The symbol tilde “ $\sim$ ” indicates that corresponding variable is a kernel variable which has been transformed into the Kernel space as follows:

$$\begin{aligned} \tilde{X} &= [\phi(x_1), \phi(x_2), \dots, \phi(x_N)] = [\tilde{x}_1, \tilde{x}_2, \dots, \tilde{x}_N], \\ \tilde{Y} &= [\phi(y_1), \phi(y_2), \dots, \phi(y_N)] = [\tilde{y}_1, \tilde{y}_2, \dots, \tilde{y}_N]. \end{aligned} \quad (2)$$

Unfortunately, such a mapping function is not available for many cases. Even if it was available, complexity of this operation would be very high since all training samples must be mapped to higher dimensional space separately. In order to overcome this problem, we may bypass the mapping function $\phi()$ and get the same results in a faster way by using an approach called the “Kernel Trick” [16, 18]. According to the kernel trick, a kernel function is employed instead of mapping function. The following equation shows a generalized form of the kernel function:

$$K(x_i, x_j) = \langle \phi(x_i), \phi(x_j) \rangle, \quad (3)$$

where K is the kernel function with the parameters x_i and x_j which represent i th and j th training samples, respectively. In this paper we have examined two well-known kernel functions, Gaussian and Polynomial kernel. Gaussian kernel (4) relocates the samples in accordance to Gaussian distribution and employs “ σ ” parameter to calibrate sensitivity as follows:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right). \quad (4)$$

Polynomial Kernel is shown in (5). This function requires “ d ” parameter to change the degree of the polynomial function and calibrate the sensitivity as follows:

$$K(x_i, x_j) = (x_i x_j + 1)^d. \quad (5)$$

In traditional FKT, computation of the covariance matrices of X and Y would be the next step. If we would apply the same operation to matrices $\tilde{X}$ and $\tilde{Y}$, the results would be as follows:

$$\begin{aligned} T_0 &= \tilde{X} \tilde{X}^T, \\ C_0 &= \tilde{Y} \tilde{Y}^T, \end{aligned} \quad (6)$$

where T_0 and C_0 are the kernel matrices of the ROI and clutter classes, respectively. At this step, we are able to exploit the covariance properties to realize the Kernel Trick. As shown in (7), one of the kernel functions may be employed to complete the kernel transformation without requiring the mapping operation [19] as follows:

$$\begin{aligned} T_0(i, j) &= \tilde{x}_i \tilde{x}_j^T = K_T(i, j), \\ C_0(i, j) &= \tilde{y}_i \tilde{y}_j^T = K_C(i, j). \end{aligned} \quad (7)$$

FIGURE 3: (a) Classification comparison for class number 8. (b), (c), and (d) Classification results of K-FKT, Classical FKT and Radial Based SVM, respectively.

After the kernel operations, the summation matrix of T_0 and C_0 is computed. Then it is decomposed into eigenvalues and eigenvectors as follows:

$$T_0 + C_0 = VDV^T, \quad (8)$$

where the symbols V and D represent eigenvector matrix and eigenvalue matrix, respectively. The diagonal elements of D are eigenvalues of the summation matrix. By using V and D , transformation operator P can be constructed as follows:

$$P = VD^{-1/2}. \quad (9)$$

After the multiplication by P , matrices T_0 and C_0 are transformed into eigenspace where both ROI and clutter

classes share the same eigenvalues and eigenvectors as follows:

$$\begin{aligned} T &= PT_0P^T, \\ C &= PC_0P^T, \end{aligned} \quad (10)$$

where T and C are the transformed matrices, respectively. Since they are transformed into the same eigenspace, the sum of matrices is equal to identity matrix as follows:

$$T + C = I. \quad (11)$$

Equation (11) implies that if θ_i is an eigenvector of T and its corresponding eigenvalue is λ_i , then $1 - \lambda_i$ is the eigenvalue

FIGURE 4: (a), (b), (c), and (d) Classification comparisons for classes number 5, 6, 10, and 13, respectively.

of C with the same eigenvector θ_i . This relation can be represented as follows:

$$\begin{aligned}\lambda_i \theta_i &= T \theta_i, \\ \lambda_i \theta_i &= (1 - C) \theta_i, \\ (1 - \lambda_i) \theta_i &= C \theta_i.\end{aligned}\quad (12)$$

The above equations state that the more information an eigenvector contains about ROI class, the less information it has about clutter class. This characteristic evolves from the classical FKT algorithm.

3.2. Testing Stage. Testing stage starts with normalization of the test sample as it is done in the training stage. Similarly, the test sample must be mapped into Kernel space, but it is

not applicable due to the reasons explained in training stage. So we shall use kernel trick operation once more as follows:

$$Z = [K(x_1, z), K(x_2, z), \dots, K(x_N, z)]. \quad (13)$$

Equation (13) shows the kernel transformation of the test sample z . In the equation, x_i represents the ROI training samples and Z represents the kernel matrix of the corresponding test sample. Matrix Z is employed to calculate feature vector F in (16). Other factor required to calculate F is the normalized T_0 matrix which is obtained by (14) as follows:

$$\hat{T} = T_0 - I_{1/N} T_0 - T_0 I_{1/N} + I_{1/N} T_0 I_{1/N}, \quad (14)$$

$$I_{1/N} = \frac{I_{N \times N}}{N}, \quad (15)$$

TABLE 2: Overall classification results.

Class	Precision	Recall	Accuracy
Alfalfa	1	1	1
Corn-notill	0.65	1	0.75
Corn-mintill	0.82	0.85	0.8
Corn	0.7	1	0.9
Grass-pasture	0.9	0.81	0.83
Grass-trees	1	0.85	0.84
Grass-pasture-mowed	1	1	1
Hay-windrowed	0.86	0.84	0.80
Oats	1	1	1
Soybean-notill	0.76	0.99	0.87
Soybean-mintill	0.79	0.94	0.82
Soybean-clean	0.8	1	0.77
Wheat	0.84	0.82	0.89
Woods	0.81	1	0.88
Buildings-Grass-Trees-Drives	0.84	0.95	0.88
Stone-Steel-Towers	1	0.94	0.91
Overall			0.86

where $\hat{T}$ represents the normalized form. Once we have the matrices Z and $\hat{T}$, we are able to calculate the feature vector F as follows:

$$F_j = \frac{1}{\sqrt{\lambda_j}} \phi_j^T Z \quad j = 1, 2, \dots, N, \quad (16)$$

where ϕ_j and λ_j denote the eigenvectors and eigenvalues of the $\hat{T}$ matrix, respectively. The final step is the multiplication of the feature vector F by the transpose of eigenvector matrix of T as follows:

$$R = \Phi^T F, \quad (17)$$

where Φ is the eigenvector matrix of T and R represents the result vector of test sample z . The test sample is estimated in ROI class if the L_2 norm of R is a large value; otherwise it is estimated in clutter class. In order to summarize the proposed method and give a brief representation, steps of our algorithm are described below.

(1) Training stage

- (i) Select N number of training samples for ROI and Clutter class.
- (ii) Map training samples into Kernel space using “Kernel Trick” approach.
- (iii) Calculate the transformation P using eigenvalues and eigenvectors.
- (iv) Transform the the matrices T_0 and C_0 into the eigenspace via P operator.

(2) Testing stage

- (i) Map test sample z into Kernel space to obtain kernel matrix Z .
- (ii) Use Z and normalized ROI matrix $\hat{T}$ to calculate feature vector F .
- (iii) Use F and eigenvalues of T to reach result value.
- (iv) Make the final decision by thresholding the result value.

4. Classification Results

In this section, classification results are presented. For each case, we selected a class among 16 classes and marked it as ROI class. Since it is not feasible to present the graphical results of all classes, we labeled some of the classes as in Figure 2. Particularly, we have shown the classified images of the class number 8 (Hay-windrowed). ROC curves are presented for the rest of labeled classes. Finally, we present Table 2 which includes the recall, precision, and accuracy results for all classes.

In first experiment “Hay-windrowed” class (labeled as number 8 in the figure) is selected as the ROI class. ROC curve results of three methods are shown in Figure 3(a). The results indicate that the kernel transformation remarkably improves the classification results. Also, our method shows higher accuracy than SVM at the same false acceptance rate (FAR) levels. For a better view of classification, result images are shown in Figures 3(b), 3(c), and 3(d) which are the results of K-FKT, Radial based SVM, and classical FKT, respectively. As shown in the figures, while classical FKT cannot classify the area, K-FKT and SVM classify the area with high accuracy.

Figures 4(a), 4(b), 4(c), and 4(d) show the ROC curves for other 4 classes, which are labeled as 5 (Grass-pasture), 6 (Grass-trees), 10 (Soybean-notill), and 13 (Wheat), respectively. According to the results, K-FKT presents higher accuracy than other classification methods. The results indicate that the ROC curve may vary for different classes since classes have different distributions.

Table 2 shows classification results of all classes. Precision, Recall, and Accuracy results are presented in each column. The results show that K-FKT offers promising classification capability for the hyperspectral image classification problem.

Experiments show that the overall accuracy of some specific classes such as “Corn-notill” and “Soybean-clean” is not higher than 80%. To clarify the ambiguity, we investigated the samples of these complicated classes and we realized that some classes are not in a “well-separable” condition. Our correlation analyses show that they are highly correlated with each other. It is the main reason behind the lower accuracy. In this study, we have only studied spectral features of the dataset, but employing also spatial features (e.g., neighbourhood information) may improve the results.

It is usually not an easy task to show a fair comparison of two studies due to unknown experimental parameters. However, by its experimental similarities, [11] can be considered as a comparison paper to our study. Samiappan et al. classify the

same dataset using SVMs with nonuniform feature selection. Their results present 75% of overall accuracy. Our results point out a remarkable contribution and exceed 75% accuracy by reaching 86% overall accuracy.

5. Conclusions

In this paper, we have presented a solution for hyperspectral image classification problems using supervised classification method called Kernel Fukunaga-Koontz Transform which is improved version of classical FKT. Since the classical FKT gives low performance for classification of nonlinearly distributed data, we have mapped the HSI data to higher dimensional Kernel space by kernel transformation. In that Kernel space, each region can be separated by classical FKT with higher performance. The experimental results verify that Kernel Fukunaga-Koontz Transform has higher classification performance than classical FKT, Linear, Polynomial, and Radial based SVM methods. Under these considerations, we can conclude that the Kernel FKT is a robust classification method for hyperspectral image classification. Our next goal is to use different kinds of kernel functions and investigate their effects to the classification results. We have an ongoing study that compares performances of different kernels.

Acknowledgment

This research was supported by a Grant from The Scientific and Technological Research Council of Turkey (TUBITAK-112E207).

References

- [1] P. S. Thenkabail, J. G. Lyon, and A. Huete, *Hyperspectral Remote Sensing of Vegetation*, CRC Press, New York, NY, USA, 2012.
- [2] L. Zhi, D. Zhang, J.-Q. Yan, Q.-L. Li, and Q.-L. Tang, "Classification of hyperspectral medical tongue images for tongue diagnosis," *Computerized Medical Imaging and Graphics*, vol. 31, no. 8, pp. 672–678, 2007.
- [3] M. Kalacska and M. Bouchard, "Using police seizure data and hyperspectral imagery to estimate the size of an outdoor cannabis industry," *Police Practice and Research*, vol. 12, no. 5, pp. 424–434, 2011.
- [4] Z. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face recognition in hyperspectral images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 12, pp. 1552–1560, 2003.
- [5] D. Manolakis, D. Marden, and G. A. Shaw, "Hyperspectral image processing for automatic target detection applications," *Lincoln Laboratory Journal*, vol. 14, no. 1, pp. 79–116, 2003.
- [6] I. Faulconbridge, M. Pickering, and M. Ryan, "Unsupervised band removal leading to improved classification accuracy of hyperspectral images," in *Proceedings of the 29th Australasian Computer Science Conference (ACSC '06)*, V. Estivill-Castro and G. Dobbie, Eds., vol. 48 of *CRPIT*, pp. 43–48, Hobart, Australia, 2006.
- [7] O. Rajadell, P. Garca-Sevilla, and F. Pla, "Textural features for hyperspectral pixel classification," in *Pattern Recognition and Image Analysis*, H. Araujo, A. Mendona, A. Pinho, and M. Torres, Eds., vol. 5524 of *Lecture Notes in Computer Science*, pp. 208–216, Springer, Berlin, Germany, 2009.
- [8] J. A. Benediktsson, J. A. Palmason, and J. R. Sveinsson, "Classification of hyperspectral data from urban areas based on extended morphological profiles," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 3, pp. 480–491, 2005.
- [9] J. Borges, J. Bioucas-Dias, and A. Maral, "Bayesian hyperspectral image segmentation with discriminative class learning," in *Pattern Recognition and Image Analysis*, J. Mart, J. Bened, A. Mendona, and J. Serrat, Eds., vol. 4477 of *Lecture Notes in Computer Science*, pp. 22–29, Springer, Berlin, Germany, 2007.
- [10] M. S. Alam, M. N. Islam, A. Bal, and M. A. Karim, "Hyperspectral target detection using Gaussian filter and post-processing," *Optics and Lasers in Engineering*, vol. 46, no. 11, pp. 817–822, 2008.
- [11] S. Samiappan, S. Prasad, and L. M. Bruce, "Automated hyperspectral imagery analysis via support vector machines based multi-classifier system with non-uniform random feature selection," in *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium (IGARSS '11)*, pp. 3915–3918, July 2011.
- [12] S. Dinc, *Hyperspectral target recognition using multiclass kernel fukunagakoontztransform [M.S. thesis]*, Graduate School of Natural and Applied sciences, Yildiz Technical University, İstanbul, Turkey, 2011.
- [13] Aviris hyperspectral data, 2013, <http://aviris.jpl.nasa.gov/html/data.html>.
- [14] K. Fukunaga and W. L. G. Koontz, "Application of the Karhunen-Loève expansion to feature selection and ordering," *IEEE Transactions on Computers*, vol. 19, no. 4, pp. 311–318, 1970.
- [15] S. Ochilov, M. S. Alam, and A. Bal, "Fukunaga-koontz transform based dimensionality reduction for hyperspectral imagery," in *Algorithms and Technologies for Multispectral, Hyperspectral, and Ultraspectral Imagery XII*, vol. 6233 of *Proceedings of SPIE*, April 2006.
- [16] Y.-H. Li and M. Savvides, "Kernel fukunaga-koontz transform subspaces for enhanced face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '07)*, pp. 1–8, June 2007.
- [17] W. Zheng and Z. Lin, "A new discriminant subspace analysis approach for multi-class problems," *Pattern Recognition*, vol. 45, no. 4, pp. 1426–1435, 2012.
- [18] R. Liu and H. Zhi, "Infrared point target detection with fisher linear discriminant and kernel fisher linear discriminant," *Journal of Infrared, Millimeter, and Terahertz Waves*, vol. 31, no. 12, pp. 1491–1502, 2010.
- [19] D. Tuia and G. Camps-Valls, "Urban image classification with semisupervised multiscale cluster kernels," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 1, pp. 65–74, 2011.