International Journal of Scientific & Engineering Research, Volume 5, Issue 4, April-2014
ISSN 2229-5518

Feature Extraction and Classification of Hyperspectral Images using Novel Support Vector Machine based Algorithms

P.Elaveni, N.Venkateswaran.

Abstract— In this paper, Hyperspectral image feature extraction and classification using two algorithms KPCA-SVM and ICA-SVM is proposed. Hyperspectral images are high dimensional, with a large amount of spectral and spatial information. Spatial information describes the physical features such as the texture, color etc. of the materials present in the hyperspectral image of interest, while the spectral information comprises of the reflectance values of these materials across different wavelengths. In this paper, principal components and Independent Components are chosen as the feature of interest. They are extracted by using the Kernel Principal Component Analysis and Independent Component Analysis algorithms respectively. These features are then used for training the Kernel based Support Vector Machine (SVM) to perform the classification process. Simulations are carried out to verify the efficacy of KPCA vs. ICA methods.

Index Terms— Classification, Feature reduction, hyperspectral image, Independent Component Analysis (ICA), Kernel Principal Component Analysis (KPCA), remote sensing, support vector machines (SVM).

1 Introduction

HYPERSPECTRAL images [HSI] are defined as a collection of high resolution pixels with very high spectral and spatial detail from the instantaneous field of view of the pixels across a large wavelength region. The major defining feature of a HSI is that the spectral information is available across more than 100 wavelength bands. A conventional RGB image has abundant spatial information but the spectral information is very limited. Hence to classify materials that have identical spatial features, the multispectral images are used. The multispectral images use only limited number of wavelengths. The spectral signature of a particular pixel can be plotted across the bands over which the multispectral image is obtained. This spectral signature can then be used as the feature for classification. There are cases where two different materials may have almost similar spectral signatures. In such cases, the number of wavelength bands over which the image needs to be captured has to be increased. In such cases, the HSI are used as they have relatively more spectral information.

The very high spectral resolution of the hyperspectral image extends its role in all the applications that require a high discrimination capability in the spectral domain such as target detection and material quantification. [1]The automatic analysis of hyperspectral images is beset with several deterrents such as

1. The redundancy of the huge amounts of spectral data to be processed.
2. The atmospheric effects.
3. The Hughes phenomenon, also known as the curse of dimensionality.

- *P.Elaveni is currently pursuing masters degree program in Communication Systems in SSN College of engineering affiliated to Anna University, Indian, E-mail: elavenipalanivel@gmail.com*
- *N.Venkateswaran name is currently working as Professer in SSN College of engineering affiliated to Anna University, Indian, E-mail: venkateswarann@gmail.com*

Since the bands of wavelengths used to capture a hyperspectral image are contiguous in nature, the spectral data obtained is usually characterized with high correlation. Removing this redundancy helps reduce the complexity of the classification algorithm. In case of supervised classification, the small ratio between the number of available training samples and the number of features is a main hurdle. As a result, the computation of class conditional hyper-dimensional probability density functions becomes impossible. This results in the Hughes phenomenon where the classification accuracy greatly reduces if the number of training samples available are not proportional to the number of features involved in the process. [2]

Several techniques have been discussed in the literature for overcoming this major obstacle namely

1. Regularization of the sample covariance matrix [3].
2. Estimation of the adaptive statistics by exploiting the semi labeled samples [4].
3. Feature selection/transformation pre-processing techniques to reduce or transform the original feature space to a lower dimensional one [5].
4. Analysis of the spectral signatures for modelling the classes defined under the classification process.[6]

For certain wavelengths, the interaction with sunlight reduces the amount of reflected energy. The transmittance of the atmosphere also gets reduced because of absorption or scattering by the molecules of different gases present in the atmosphere at certain wavelengths. Information can't be retrieved from such corrupted bands. Therefore, those bands are removed.

This paper is organized into the following sections. Section 2 involves the detailed discussion of the feature extraction techniques used KPCA and ICA. Section 3 describes the basics of SVM and the One vs all algorithm used for multiclass problem. Section 4 gives details about the hyperspectral datasets used and the result of the experimental results and finally Section 5 presents the conclusions.

International Journal of Scientific & Engineering Research Volume 5, Issue 4, April-2014
ISSN 2229-5518

2 Feature Extraction

Most of the algorithms used for the processing of hyperspectral data have a complexity that can directly be linked to its dimension. Hence, it's of importance to evolve techniques that help reduce this dimension but retain the maximum representation in the process. Feature extraction is a solution. The term feature can refer to the spectral bands comprising of the hyperspectral image or a transformation of these bands. The major requirement for feature extraction technique especially in hyperspectral image processing is to reduce the redundancy of the spectral information thereby reducing the complexity of further processing steps. In this paper the Kernel Principal Component Analysis and Independent Component Analysis feature extraction techniques are discussed.

2.1 Kernel Principal Component Analysis

The most commonly used feature transformation technique is the Principal Component Analysis (PCA). It is an orthogonal linear transformation technique that uses eigen analysis [7] to extract the Principal Components which can then be used as the feature for further classification. The Principal Components are preferred because of their following unique properties.

1. Principal Components are the linear combinations of the original data.
2. They maximize the variance of the input data.

The properties mentioned above are illustrated in the simulation shown in figure 1.

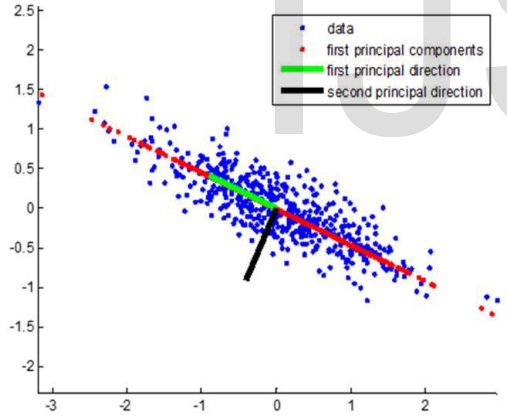

Fig. 1. Principal components of sample

KPCA ALGORITHM :

1. Convert the multidimensional Hyperspectral data into a two dimensional data.
2. Subtract the mean of the data F
3. Calculate the kernel matrix.
4. Calculate the Covariance matrix for the mean subtracted data.

$$C = F \cdot F^T \quad (1)$$

5. Calculate the eigenvalues λj and eigenvectors of C .
6. These are sorted in descending order of the eigenvalues.
7. The eigenvector with the largest eigen value is the principal component

The non-linear relationships which may sometimes contribute to better classification results are lost. Introducing the concept of Kernels can help overcome this deterrent. A kernel function can be defined as a function that maps the non-linear data to a higher dimensional feature space where the data becomes seperable. As a result even the non-linear relationships can be taken into consideration by the linear algorithm. Hence, the Kernel approach has become a sought after approach in the field of machine learning as it helps in the non-linear generalization of any linear algorithm [8],[9],[10]

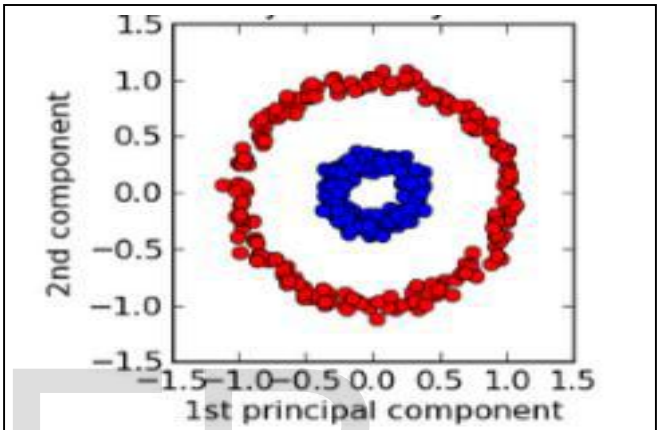

Fig. 2. Projection of sample data using PCA

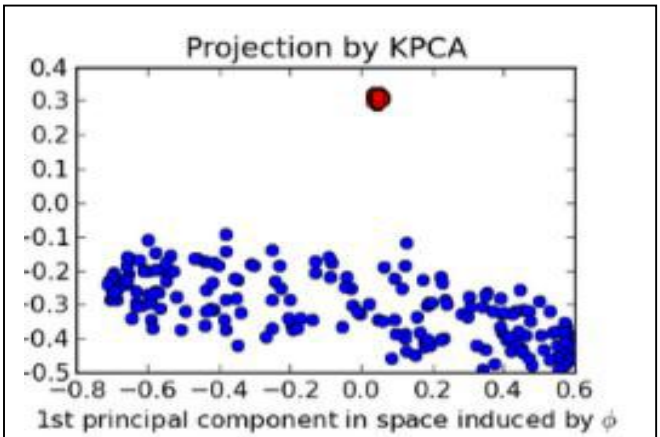

Fig. 3. Projection of sample data using KPCA

As seen by comparing figure 2 and figure 3 the Kernel Principal Component Analysis is an improved version of the Principal Component Analysis, where the data is transformed into a form where classification becomes simpler. The transformed features using KPCA inherit all the properties of the features obtained using PCA but enable easy classification of the non-linear data unlike PCA. The KPCA uses the techniques of the

International Journal of Scientific & Engineering Research Volume 5, Issue 4, April-2014
ISSN 2229-5518

kernel methods to carry out the non-linear mapping of the original features on the reproducing kernel Hilbert space. The Kernel trick transforms the linear PCA algorithm into a non-linear form. Thus the KPCA can manipulate the hyperspectral data better than the PCA.

The Kernel is initially used to transform the original data of interest. Let us consider a feature space Q related to the original input domain RN . If φ is the kernel then,

$$\varphi: x \to Q, x \to \varphi(x) \quad (2)$$

The feature space Q may have infinite dimensions. Thus the Kernel Matrix is computed as follows.

$$K_{ij} = K(x_i, x_j) \quad (3)$$

In order to ensure that the mean of the dataset involved has a zero mean, the data is centered by using the equation given below

$$K_c = K - 1_N K - K 1_N + 1_N K 1_N \quad (4)$$

1_N is a N square matrix for which $(I_N)_{ij} = 1/N$ for all (i, j). K is then diagonalised and the corresponding eigen vectors are normalized.

$$\lambda_k(\alpha^k \alpha^K) = 1 \quad (5)$$

where λ_k are the non-zero eigen values and α^K the corresponding non zero eigen vectors. The first K principal components are finally extracted.

$$\varphi(x)^k_{k_{pc}} = \sum_{i=1}^{N} \alpha_i^k \, [\varphi(x_i).\varphi(x)] \quad (6)$$

If a linear Kernel is used then the KPCA will reduce to a normal PCA. The other kernels that can be used are polynomial or Gaussian. The performance of KPCA on using the different kernels are compared in this paper to determine their corresponding efficiency.

2.2 Independent Component Analysis

Independent Component Analysis (ICA) is a multivariate data analysis method that, given a linear mixture of statistical independent sources, recovers these components by producing an unmixing matrix. It's a special case of blind Source separation (BSS) technique.

The system model is given by the equation

$$x = As \quad (7)$$

Where x is the n dimensional observation vector,A the mixing matrix of size n x m and s , the m dimensional random vector. The ultimate goal is to determine (s ,A) with the knowledge of only x . If s is statistically independent, then the technique used to extract them is called the Independent Component Analysis.

The solution of ICA can be modelled as a problem defined by,

$$y = Wx \quad (8)$$

In the above equation, W is the unmixing matrix, y the m dimensional vector that is observed. With respect to the hyperspectral image, this refers to the final output where the classes are separated out. The x is the n dimensional observation vector which in the case of a hyperspectral image is the spectral band where the pixels belonging to different classes are mixed together. The key to ICA estimation is non-gaussianity. Inorder to obtain a unique solution, the input data should be non-Gaussian. The number of sources should be smaller than or equal to the number of observations. From equation (8), y can be estimated as

$$y = W^T x = \sum_{i=1}^{N} W_i x_i \quad (9)$$

When W is one of the rows of A –1, then y is one of the independent components. Here W is a vector that maximizes the non-gaussianity of the data. Hence maximizing the non-gaussianity of the data provides the independent components [11]. The measure of non-gaussianity can be achieved by using several values such as kurtosis, negentropy or mutual information. Negentropy is chosen as the calculation of this value is simplified by the approximation developed by Hyvarinen in 1998 called the maximum entropy principle. According to this principle, the negentropy is proportionally related to the expectations of some contrast function. The corresponding equation is given by,

$$J(y) = [E\{G_i(y)\} - E\{G_i(v)^2\}]^2 \quad (10)$$

Where $G\,i\,(y\,)$ is a contrast function. $-\exp(-y^2/2)$ .Thus, by using the negentropy as the measure of non-gaussianity, the independent components can be calculated. The independent components extracted can then be used as the features using which further classification can be achieved using SVM.[12].

ICA ALGORITHM

1. Choose an initial vector W.
2. Let $W^+ = E\{xG(W^T x)\} - E\{G'(W^T x)\}W$where $G(u) = u \exp(-\frac{u^2}{2})$
3. W=W_+/$||W^2||$
4. If not converged repeat steps from 2.
5. Converged if ($W_{new} - W_{old}$)>ε

3 Support Vector Machines

Support Vector Machine is a pattern recognition technique that adopts the structural risk minimization (SRM) criterion rather than the concept of empirical risk minimization (ERM) [13] i.e. the classification strategy adopted in SVM depends on the geometrical criterion rather than the statistical criterion. There is no requirement for the estimation of any statistical distributions characterizing the classes in case of SVMs. The entire classification process depends only on the support vectors. As a result, the accuracy of classification, in the case of SVM doesn't depend on the amount of training data available. Thus the Hughes phenomenon doesn't impact the results when SVM is adopted. Generally, the overall accuracy of SVM is accepted to be greater than the other commonly used classification techniques such as maximum likelihood and multilayer perceptron neural network classifiers.

In spite of several merits listed above, the complexity of calculation may unduly be increased by the high spectral correlation of the hyperspectral data. Hence some pre-processing steps are indeed necessary. The non-linearity of the training data can be exploited by incorporating the Kernel trick in SVMs. Another problem faced while working with SVM is that SVMs

International Journal of Scientific & Engineering Research Volume 5, Issue 4, April-2014
ISSN 2229-5518

were originally developed for binary problems only [14]. But in case of hyperspectral images, the classification usually involves multiple classes. This problem can be approached through two methods. Several binary classifiers can be combined using an appropriate algorithm like one against one or one against all. The other is designing a multiclass classifier directly. The former method is preferred since designing a multiclass classifier involves optimizing several parameters simultaneously which would cause the classifier to become unstable. Both one against all and one against one algorithms provide almost the same accuracy. In this paper the one against all algorithm is implemented.

The main objective in SVM is to define a hyperplane that effectively demarcates the given classes. In case of a binary case, the hyperplane defined is of the form

$$f(x) = w.x + b \quad (11)$$

Where $w \in R^d$ is the normal vector to the hyperplane and $b \in R^d$ is the bias. The hyperplane has to be optimally defined such that the distance from the hyperplane to the closest training sample is maximized. The optimal hyperplane can be determined as the solution of the following convex quadratic programming problem.

$$minimize: 0.5\ ||W||^2$$
$$subject\ to: \ y_i(w.x_i + b) \geq 1\ i = 1,2 \ldots . N \quad (12)$$

This classical linearly constrained optimization problem can be translated (using a Lagrangian formulation) into the corresponding dual problem. The Lagrange multipliers involved in the obtained equation can be estimated using quadratic programming (QP) methods. The discriminant function associated with the optimal hyperplane becomes an equation depending both on the Lagrange multipliers and on the training samples, i.e.

$$f(x) = \sum_{i \in s} \alpha_i y_i (x_i, x) + b \quad (13)$$

In order to incorporate even the non-linear relations of the training samples, we introduce the kernel trick in the linear SVM to form the discriminant function as given below.

$$f(x) = \sum_{i \in S} \alpha_i y_i\, K(x_i, x) + b \quad (14)$$

Where $K(x_i, x)$ is the kernel matrix of the data. Thus the efficiency of classification is higher in the case of kernel based SVM when compared to linear SVM. The kernel used is the Gaussian radial basis function given by the equation.

$$K(x_i, x) = \exp(-\gamma \left|\left|x_i - x\right|\right|)^2 \quad (15)$$

4 Simulation Results

4.1 AVIRIS Datasets

AVIRIS is an acronym for the Airborne Visible Infrared Imaging Spectrometer. It is a unique optical sensor that delivers calibrated images of the upwelling spectral radiance in 224 contiguous spectral channels (also called bands) with wavelengths from 400 to 2500 nanometers (nm).

The test hyperspectral image used in this paper is captured by the AVIRIS sensor over the Indian Pines test site in Northwestern Indiana and consists of 145 x 145 pixels and 224 spectral reflectance bands in the wavelength range 0.4–2.5 10^{-6} meters. This scene is a subset of a larger one. The Indian Pines scene contains two-thirds agriculture, and one-third forest or other natural perennial vegetation. There are two major dual lane highways, a rail line, as well as some low density housing, other built structures, and smaller roads. Since the scene is taken in June some of the crops present, corn, soybeans, are in early stages of growth with less than 5% coverage. The ground truth available is designated into sixteen classes and is not all mutually exclusive.

The ground truth of a hyperspectral image defines the classes to which each pixel of the hyperspectral image belongs. This information is very important to define the Support Vector Machine since the grouping matrix is formed using this information only. Also the ground truth is required to measure the accuracy of the defined SVM. Part of the pixels of the Indian pines Hyperspectral Dataset is used for training the SVM. The training matrix and the grouping matrix is derived from this image and its ground truth which is shown in figure 4 and 5. Salinas-A, which is a subset of the Salinas image is used for testing the SVM.

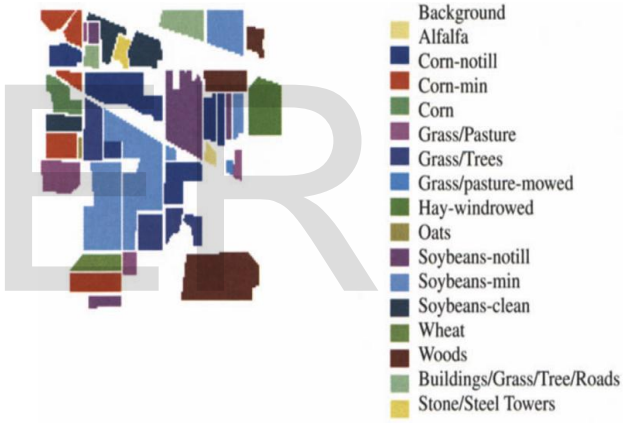

Fig. 4. Thematic map of the Indian Pines Hyperspectral data

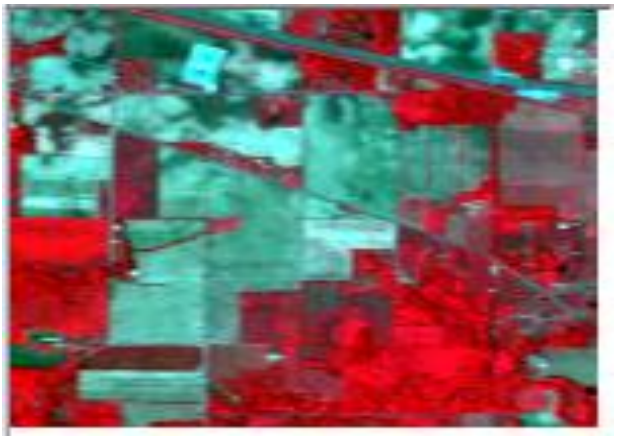

Fig. 5. Indian Pines Hyperspectral image

International Journal of Scientific & Engineering Research Volume 5, Issue 4, April-2014
ISSN 2229-5518

4.2 Implementation

Initially the Principal Components of the hyperspectral image i.e., the Indian image are extracted using KPCA. The first four principal components extracted are shown in figure 6. Also the independent components are also extracted from the same test data and displayed as figure 7. We can observe from figure6 that the maximum information of the hyperspectral image is present in the first few principal components itself but incase of figure7 it can be observed that the pixels of different classes is grouped in different independent components. For example the roadways are clearly demarcated in the third Independent component.

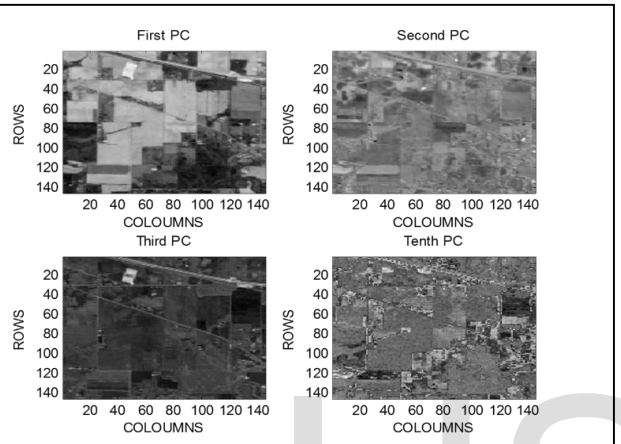

Fig. 6. Principal Components extracted using KPCA

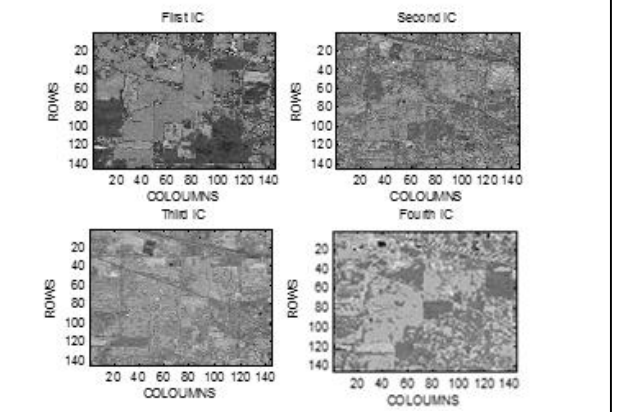

Fig. 7. Principal Components extracted using ICA

The SVM is trained using the ground truth that is available. After it has been applied both the principal components and the independent components are used to test the SVM thus obtaining a classification map as the result. The Indian pines data set originally has 16 classes. But the classes with very few pixels are discarded. The classification of the various classes of the HSI is achieved using the Support Vector Machine. Since there are 16 classes, this is a multiclass problem.

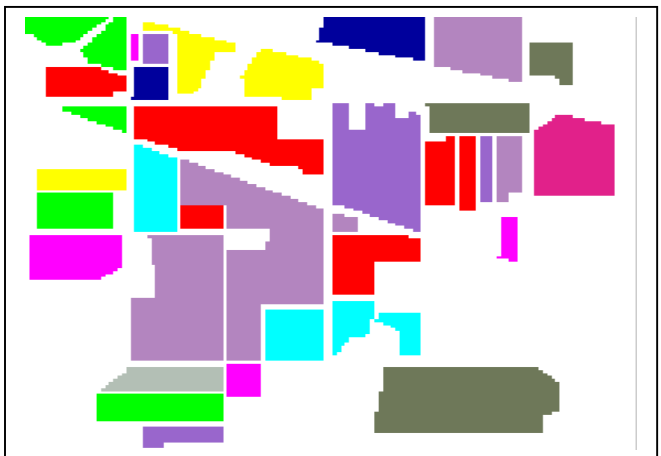

Fig. 8. Classification output using KPCA-SVM.

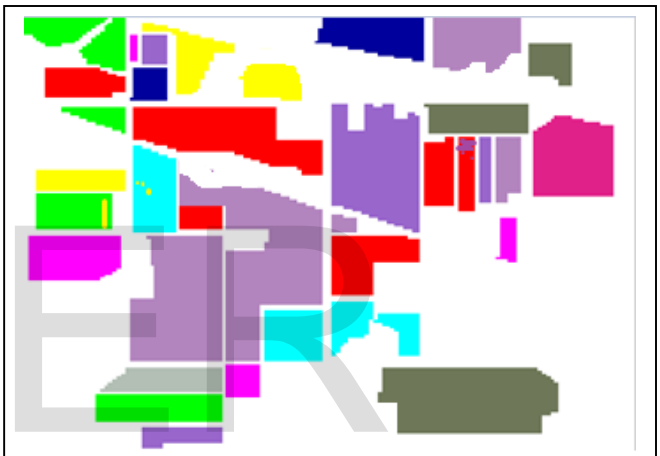

Fig. 8. Classification output using ICA-SVM.

One vs. All algorithm is implemented to solve the multiclass problem. In this algorithm, initially one class is separated out and all the other classes are grouped together. Thus in this way the multiclass problem is reduced to a binary class problem, SVM is then implemented. This process is repeated until all the classes are individually classified. SVM involves the training and testing stages. In the training process the features used for classifying form the rows of the training matrix. The SVM is defined in this stage. The classes to which these features belong to form the grouping matrix. Converting the raw data into the training and grouping matrix is the training process. In the testing stage, a HSI is classified into the classes without apriori knowledge of its ground truth. The Thematic map of Indian pines data set is shown in figure 5 and the results of classification of the hyperspectral image are shown in figure 8. An accuracy of 95% is achieved by using this algorithm.

5 Conclusion

In this paper multiclass classification problem in the premise of

International Journal of Scientific & Engineering Research Volume 5, Issue 4, April-2014
ISSN 2229-5518

hyperspectral images is addressed. Two feature extraction techniques, Kernel Principal Component Analysis and Independent Component Analysis techniques are compared. The KPCA and ICA are implemented to extract the principal components and independent components respectively. Though the KPCA-SVM algorithm provides an improved accuracy compared with ICA-SVM, the difference in accuracy is very less. On the other hand, on comparing the complexity, when ICA is used as the feature for classification, the computation time reduces. This is because, in KPCA-SVM algorithm, the kernel matrix computation is complex and time consuming. After feature extraction, SVM is implemented. The classification is carried out using Non-linear Kernel based SVM wherein the Gaussian RBF kernel is used. Efficiency is calculated by comparing the classified result obtained from the SVM with the known ground truth. The KPCA-SVM algorithm provides an overall accuracy of 95% while ICA-SVM provides 93.25%

REFERENCES

[1] Gary Shaw and Dimitris Manolakis, Massachusettes Institute of Technology, Lincoln Laboratory, "Signal Processing for Hyperspectral Image Exploitation", *IEEE Signal Processing magazine*,2002.

[2] G. F. Hughes, "On the mean accuracy of statistical pattern recognizers," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 55–63, 1968

[3] J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Anal. MachineIntell.*, vol. 18, pp. 763–767, July 1996

[4] Q. Jackson and D. A. Landgrebe, "An adaptive classifier design for highdimensional data analysis with a limited training data set," *IEEE Trans.Geosci. Remote. Sensing*, vol. 39, pp. 2664–2679, Dec. 2001

[5] Chein-I Chang, "A Joint Band Prioritization and Band-Decorrelation Approach to Band Selection for Hyperspectral Image Classification." , *IEEE transactions on geoscience and remote sensing*, vol. 37, no. 6, november 1999.

[6] Farid Melgani and Lorenzo Bruzzone, "Classification of Hyperspectral Remote Sensing Images With Support Vector Machine", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 42, no. 8, Aug. 2004.

[7] C.-I.Chang, Q.Du, T.M.Tu, and L.G.Althouse, "A Joint Band Prioritization and Band-Deccorelation Approach to Band Selection for Hyperspectral Image Classification,"*IEEE Trans. Geoscience Remote Sensing*, vol. 31, pp. 2631-2640, Nov. 1999

[8] BScholkopf, S.Mika, C.J.C.Burges, P.Knirsch, and KR.Miiller, "Input Space versus Feature Space in Kemel-Based Methods," *IEEE Trans. Neural Networks*, vol. 10, pp. 1000-1017, Sep. 1999.

[9] A. Hyvarinen, J. Karhunen, and E. Oja, Independent Component Analysis, *Wiley*, New York, 2001

[10] A. Villa, J. Chanussot, C. Jutten, J. A. Benediktsson, and S. Mossaoui, "On the use of ICA for hyperspectral image analysis,"in *Proc. IEEE IGARSS '09*, 2009

[11] K-R.Miiller, S.Mika, G.Ratsch, and K.Tsuda, "An Introduction to Kernel-Based Learning Algorithms," *IEEE Trans. Neural Networks*, vol. 12, pp. 181-201, Mar. 2001.

[12] A.Ruiz and E.L.T.Pedro, "Nonlinear Kemel-Based Statistical Pattem Analysis," *IEEE Trans. Neural Networks*, vol. 12, pp. 16-32, Jan. 2001.

[13] B.Scholkopf, "The Kemel Trick for Distances," *Microsoft Research Technical Report*, MSR-TR-2000-5 I, pp. 1-10, May 2000.

[14] V.N. Vapnik, Statistical Learning Theory. New York, NY, *USA:Springer*, 1998.