A Study on Classification Techniques used in Hyperspectral Imaging

[1]Aswathy S M,[2]Sumithra M D

[1]M.Tech Student,[2]Assistant Professor
[1]Department of Computer Science and Engineering,
[1]LBSITW, Poojappura, Trivandrum, Kerala, India

***Abstract*— Hyperspectal imaging is the area where the power of digital imaging and spectroscopy is combined. Hyperspectral camera captures radiance or intensity for each pixel in the form of large number of contiguous spectral bands. The continuous spectrum for each pixel in the image can be used to identify the objects in a scene. Hyperspectral imaging is used in wide variety of applications mainly in astronomy, agriculture, geosciences, biomedical imaging and surveillance. Image classification is one of the important classifications of hyperspectral imaging (HSI). A wide range of classification techniques are available. Since hyperspectral image contains rich amount of information, accuracy of classification is an important issue. The first section of this paper presents the concept of hyperspectral imaging and next session presents various supervised and unsupervised techniques used in hyperspectral image classification.**

***IndexTerms*— Hyperspectral imaging, supervised classification unsupervised classification**

I. INTRODUCTION

Hyperspectral imaging (HSI) is a recent breakthrough in the area of remote sensing. Remote sensing is the capturing of information about objects in an area without making any physical contact with that object. In early days hyperspectral images are accessed only by remote sensing experts. Nowadays hyperspectral image analysis has become one of the growing technologies in remote sensing by combining the power of digital imaging and spectroscopy. Human eye can see objects only in the visible light of electromagnetic spectrum where hyperspectral imaging collects information across wide range of electromagnetic spectrum from ultraviolet to long infrared. Hyperspectral camera captures radiance or intensity for each pixel in the form of large number of contiguous spectral bands. Hyperspectral imaging provides a third dimension to a reflected image that contains object's spectral data. In hyperspectral imaging hyper indicates 'too many' and indicates large number of measured wavelengths. Hyperspectral imaging is a collection of spectral details over an area using remote sensors where each image data represents dozens or hundreds of narrow, adjacent wavelength bands. The continuous spectrum for an image cell can be prepared from these spectral bands. This spectral information helps to uniquely identify and distinguish similar objects [1].

Hyperspectral imaging technology uses the ability of spectral imaging to capture the energy radiated from material's surface. The radiated information can be used to derive the information about materials such as texture, chemical composition etc. The received energy is transformed into electric impulse using optical sensors. These impulses are translated into series of images and each image represents many spectral bands over a defined range. The spectral data is made into a form of hyperspectral image cube like a set of images layered on top of one another. The data cube is a combination of three dimensions (x,y,λ) where x and y are two spatial dimensions and λ represents a range of wavelength.

Reflectance is the percentage of electromagnetic power that is reflected from a material. The reflectance spectrum is the plot of reflectance as a function of wavelength. The energy measured from various parts of electromagnetic spectrum in discrete form is called a spectral band. The spectral information can be compared with fields or laboratory reflectance spectra in order to uniquely identify and distinguish spectrally similar materials

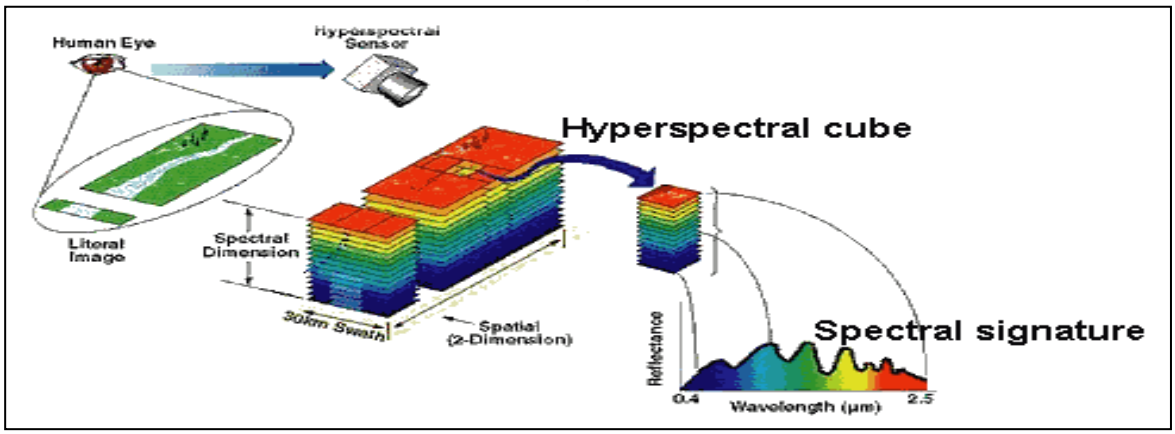

Fig 1 Hyperspectral Imaging

. Hyperspectral data is used in many applications such object detection, material mapping and identification. Reflectance statistics is used for classifying image pixels. Hyperspectral imaging helps to identify minerals for mining and oil industries. Figure 1 shows how hyperspectral image is taken by hyperspectral sensors. Following session describes various classification techniques used in hyperspectral image analysis.

II. CLASSIFICATION TECHNIQUES

Classification of remotely sensed data is the process of assigning data with similar characteristics into group of classes with the intension of uniquely identifying objects within an image. Hyperspectral image classification process produces thematic maps of earth surface objects such as road, roof, soil and building etc and the selected categories are distinguishable in an image. Classification is done as follows. The image pixels are clustered into several classes such that pixels with similar characteristics are placed into same class. The aim of classification is to automatically classify image pixels into different land cover classes. An important issue regarding hyperspectral image classification is the large amount of data captured by hyperspectral imaging systems. Classification can be done in two ways: based on pixel information and based on training samples. The analysis involves dimensionality reduction followed by classification to study the model in a low dimensional feature space. Representative samples from land cover classes are called training samples.

Hyperspectral images can be classified based on the training samples and pixel information. Classification classes are defined based on the characteristics and objective of the image data. In supervised classification user selects sufficient reference data as training sites from an image which are various land cover classes. The spectral signatures of the training sites are used for classifying land cover classes. The main steps involved in supervised method are feature selection, labeling and classifying. In order to reduce the data dimensionality the image is transformed into a feature image. The spectral characteristics are used to train the classifier for transforming data points into thematic maps. The unsupervised method does not have any known priori. Based on the spectral characteristics such as mean, standard deviation etc the pixels in an image are grouped. This group of pixels is called clusters. User or computer automatically identifies these clusters and which bands to use. Finally the spectral clusters are made into information classes by the analysts. Per-pixel classifier is a pixel based classification where the entire scene is processed pixel by pixel. It considers the spectral information of all training samples but ignore the effects of mixed pixels. Maximum likelihood classifier is an example for per-pixel classifier.

A.Maximum Likelihood Classifier (MLC)

In remote sensing applications Maximum Likelihood Classifier (MLC) is the most commonly used classifier. It is a parametric classifier where a sample that shows maximum likelihood is assigned into the corresponding class. Maximum Likelihood classifier work based on two concepts: A normally distributed multidimensional data space is assumed for each sample in the class and Bayes' theorem is used for decision making. It considers variance and covariance of the pattern for classifying undefined sample into one of the classes. The spectral characteristics of an unknown pixel such as mean vector and covariance matrix can be identified from the normally distributed data space. The statistical probability for each class is computed using the characteristics in order to determine the class membership. The unclassified sample is assigned to one with highest probability value. To perform classification a probability threshold is defined for each pixel. The undefined pixel becomes unclassified if its highest probability is smaller than the threshold. The maximum likelihood function, L_k for a pixel belonging to class k in the image is calculated as follows:

$$L_k = P(k/\mathbf{X}) = P(k)*P(X/k) / \Sigma P(i)*P(\mathbf{X}/i) \quad (1)$$

where P(k) denotes prior probability of class k and P(X/k) denotes conditional probability to observe X from class k. The likelihood function can also expressed as follows due to mathematical reasons:

$$gi\ (x) = \ln p(wi) - \tfrac{1}{2} \ln |\Sigma i| - \tfrac{1}{2} (x-m_i)t\ \Sigma i^{-1}\ (x-mi) \quad (2)$$

where i represents class, x represents n-dimensional data (where n is the number of bands) , p(wi) represents probability that class wi occurs in the image and is assumed the same for all classes , |Σi| represents determinant of the covariance matrix of the data in class wi , Σi^{-1} represents its inverse matrix and m_i denotes mean vector. The problem is that when the dimension of the data increases the size of the training sample also increases. This causes practical problems and thus classification accuracy decreases.

B. Neural Network Classifier

Neural network consists of processing units called neurons that are organized in layers. Each unit takes input signal and process with the help of some function to produce output. Output from each unit ifs fed to the next layer. At the output layer, one node for each class will be defined. The output value, O_i for the neuron is defined as follows:

$$O_i = K\varphi .\ \sum_{j=1}^{N} W\,i,j\ Xj - \Theta i \quad (3)$$

k denotes a constant,φ is a nonlinear function, Wi,j represents synaptic weights and Θi is the threshold. The general concept is that the input is 1 if the neuron is active and it is 0 if the neuron is inactive. The weights are determined through iterative training procedure and output is produced for each sample .During training phase the class for each sample is known and correct values are assigned to those output nodes. The network's calculated value is compared with the correct values in order to calculate the error. This error value is used to adjust the weights. This helps to find a closer correct value for the output value. The network takes all the input to perform classification and gives class representation for each input set.

C. Bayesian Network

Bayesian classification is a supervised classification that uses a network with directed acyclic graph. Arcs in the graph represent causal influences within features and lack of arcs indicates conditional independencies. There are mainly two tasks involved in the learning procedure of Bayesian network. The DAG structure is learned first and after that parameters are determined. The probabilistic parameters are computed first and then they are encoded into a set of tables. Each node in the structure has associated conditional probability table that describes the parameters. The joint distribution is calculated from these tables. The structure of the network that is assumed to be correct is taken as the known structure. The parameters in the conditional table are learned by determining exponential number of parameters from the data.

D. Independent Component Analysis

The Characteristics of independent component analysis (ICA) can be exploited for hyperspectral image classification. Independent components (ICs) are retrieved from the original data and projected onto independent space. For each class data are centered and optimum transform is derived using independent component analysis. A multivariate density function for each component is computed using a nonparametric kernel density estimator. Density function is determined from the product of estimated marginal densities. The classification assignment is performed by submitting the results into the Bayesian classification rule. Based on conditional probabilities and priori probability the Bayes classifier is designed. Following are the general assumption taken by ICA,

- The components of density function are statically independent.
- There should be at most one signal with Gaussian distribution.
- The total number of independent components and the number of unknown mixtures are equal.

In the case of high dimensional data, important information may be hidden in low density regions. The issue with this method is that by using multidimensional estimator some important information for classification may not get retrieved.

E.Classification using Support Vector Machine (SVM)

In [5] Farid Melgani and Lorenzo Bruzzone introduced support vector machines for hyperspectral image classification.Support vector machine provides higher classification accuracy than other pattern recognition methods such as maximum likelihood and neural networks. Different proportions between the testing and training samples is taken as the basis for estimating the accuracy of SVMs. SVM is a supervised learning model that defines two categories of classes from the given set of training examples. SVM builds a classification model that assigns new sample into one category or other. The training samples of separate classes are divided by a clear gap that is as wide as possible. To perform the classification task support vector machine exploits the concept of margin maximization instead of estimating the statistical distribution of classes.

SVM defines a decision boundary that keeps the data points maximally away from each of the two categories. The margin of the classifier is determined by the distance to the closest data points. Training samples are classified into two classes where these classes are separated by a linear separator called hyperplane. A number of hyperplanes can be defined for a classification task. The one hyperplane that represents largest separation between two classes is the best choice for a hyperplane. A hyperplane that maximizes the distance from it to the nearest data points on each side of a surface is called the maximum- margin hyperplane. The decision function for a SVM is specified by the small set of data points that indicates the position of the linear separator. These data points are known as support vectors.

Consider a set of training data points with N vectors from a d- dimensional feature space. A target is associated to each vector x_i. The given data set can be defined as follows:

$$D= \{(x_i, y_i) \mid x_i \in R^d , y_i \in \{-1, 1\}\}_{i=1}^{N} \quad (4)$$

Assume that the two classes are linearly separable. From this assumption one can find atleast one hyperplane that maximizes the margin and classifies class with one target from those having other target. A hyperplane can be defined by a vector $w \in R^d$ and a bias b where the vector is normal to the hyperplane. The decision rule for the assignment of data points can be defined by a function as

$$F(x)= w.x + b \quad (5)$$

Since the training data are linearly separable two planes and selected so that there exists no points in between them. The region surrounded by these hyperplanes is called margin. The hyperplanes are described by following equations

$$wx_i + b = 1 \quad (6)$$
$$wx_i + b = -1 \quad (7)$$

The distance between the two hyperplane is equal to 2/ ||w||. In order to avoid the data point from falling in between these hyperplane, two constraints are defined

$w.x_i - b \geq 1$ for x_i of the first class or (8)
$w.x_i - b \leq -1$ for x_i of the second. (9)

This can be written as follows:

$$Y_i(w.x_I - b) \geq 1 , \quad \text{for all } 1 \leq i \leq n. \quad (10)$$

This can be put together into an optimization problem:

$$\text{Minimize } \|w\| \quad \text{subject to} \quad y_i(w.x_i - b) \geq 1 \qquad (11)$$

Figure 2 shows the concept of support vector machine in a linearly nonseparable case. White and black circles represents the two classes "+1" and "-1," respectively. The extra circles on hyperplanes are known as support vectors.

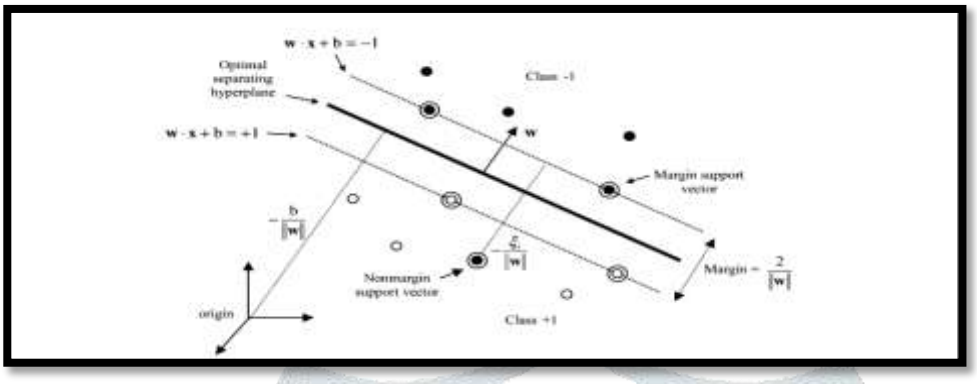

Fig 2. Support Vector Machine

The drawback of SVM is that it is mainly developed for solving binary classification problems.

F.Classification using Nearest Neighbor Classifier

Nearest neighbor classification is a simple supervised method that predicts the label of a testing sample. Nearest neighbor classifier is a non parametric classifier. It does not assume a normally distributed dataset and image classes are separated without using statistical parameters. Nearest neighbor (NN) algorithms classifies a new sample based on the distance to the nearest training samples. According to a given distance measure it tries to find the training sample nearest to the testing sample and then former's category is assigned to the latter sample. Consider a data set with training samples $X = \{x_i\}_{i=1}^{n}$ in R^d (d-dimensional feature space) and class labels $w_i \in \{1,2,\ldots., C\}$ where C indicates the number of classes and n is the total number of training samples. Suppose n_l is the number of available training samples for the l^{th} class where $\sum_{l=1}^{C} nl = n$. There are various metrics available for determining the distance to the nearest data points. Most commonly used distance metric is Euclidean distance. According to this metric the distance between training sample x_i and the given testing sample y is given by

$$d(x_i, y) = \| x_i - y \|_2^2 \qquad (12)$$

k-NN classifier is the extension of the NN classifier. It chooses k nearest training samples from the training data X. In d-dimensional feature space training samples are represented in terms of vectors and each sample is associated with class label. The figure 3 shows the concept of k-NN classification.

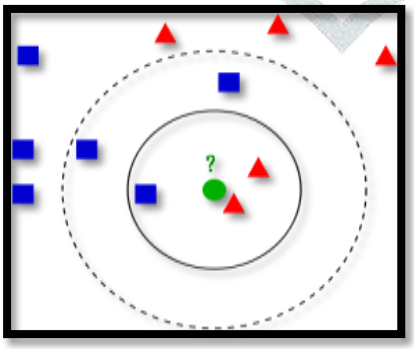

Fig 3 k-NN classification.

Let the green circle be the test sample. There are two set of classes: red triangles and blue squares. The test sample belongs either to the first class or to the second class. Consider k=3, indicated in solid line circle. There are two red triangles and one square inside the solid line. Hence the green circle should be assigned to the red triangle class. If k is chosen as 5, then test sample is classified to the first class because there are three blue squares and only two red triangles inside the dashed outer circle. The

good choice for k is selected based on the data set on which experiment is conducting. The effect of noise for the sample classification can be reduced by the selecting larger values for k.

There are mainly two stages involved: training phase and classification phase.The feature vectors are stored in training phases and also the class labels for the training samples are assigned. The value of k contains a predefined value given by the user. An undefined test sample is classified to a label that is most frequently appearing among the k nearest training samples to the test point. The class label of unknown sample is finally determined based on the majority voting.

G. Classification using Local Mean based NN classifier (LMNN)

The local mean based nearest neighbor classifier is an extension of K-NN classifier. The basic idea is that from each class the local mean vector of k nearest training samples in each class is determined. Euclidean distance method is used as the distance metric. First the k nearest training samples from each class is selected. Let k nearest training sample in the lth class be $\{X_l^1, X_l^2, .. X_l^k\}$, then the local mean vector y_l is calculated as follows

$$Y_l = \frac{1}{k}\sum_{j=1}^{k} x_l^k \qquad (13)$$

After calculating the mean vector per class, the class that minimizes the residual is selected as the class label of y.

$$\text{Class } (y) = \arg\min_{l=1,\ldots C} r_l(y) \qquad (14)$$

Where $r_l(y)$ is theresidual between the mean vector and the testing sample. When k=1, LMNN is equivalent to 1-NN classifier.

H. Classification Using Sparse Representation

Nowadays sparse representation of signals is used in many signal processing applications. Using only a few coefficients, most important information in a signal can be indicated by sparse representation. The important information can be stored in a certain basis or dictionary. Nowadays, sparse data representations are incorporated with compressed sensing in order to perform various applications in pattern recognition. The applications are based on the fact that the signals in the same class lie in low dimensional feature space.

In the case of hyperspectral imaging, remote sensors collect large number of spectral bands. In [8] Yi Chen, Nasser M Nasrabadi and Trac D Tran introduced a classification method that uses sparsity of the input with respect to a given training dictionary. A hyperspectral data pixel of an image can be represented by a linear combination of a few training samples from a training dictionary. A sparse vector is defined for the undefined pixel in the form of sparse representation where the nonzero entries in the sparse vector represent the weights of the selected training samples. The test spectral pixel's sparse vector and its associated weights can be recovered by solving an optimization problem. Based on the spectral characteristics of the recovered sparse vector, the class of the unknown pixel is determined.

A. Sparsity Model

Sparsity model assumes that the spectral characteristics of a pixel lie in a low dimensional space. Consider a dataset with M distinct classes and each of mth class contains N_m training samples $\{a_j^m\}_{j=1,\ldots Nm}$. Suppose x denotes a B-dimensional hyperspectral pixel observation. If the observed pixel lies in the mth class, then its spectral characteristics can be approximated by the training samples $\{a_j^m\}_{j=1,\ldots Nm.}$ in that mth class. The test pixel can be represented as the linear combination of these training samples as follows

$$\begin{aligned} x &\approx \alpha_1^m a_1^m + \alpha_2^m a_2^m + \ldots\ldots + \alpha_{Nm}^m a_{Nm}^m \\ &= [\, a_1^m \; a_2^m \ldots\ldots \; a_{Nm}^m \,] \, [\alpha_1^m \; \alpha_2^m \; \ldots\ldots \; \alpha_{Nm}^m]^T \\ &= A^m \alpha^m \end{aligned} \qquad (14)$$

where A^m represents B ×N class dictionary whose columns are the training samples in the mth class and α^m is an unknown sparse vector whose elements are the weights of corresponding training samples in A^m. The undefined test pixel can be represented in terms of the union of the M subspaces related with M classes. Then the test sample is represented as

$$x = A^1\alpha 1 + A^2\alpha^2 + \ldots\ldots + A^M\alpha^M = A\alpha \qquad (15)$$

where A is a B ×N is a structured dictionary that contains training samples from all classes and α represents N-dimensional sparse vector that is built by concatenating the sparse vectors { αm }m=1,…M The test pixel can be written in terms of k dictionary samples. These are associated with k nonzero entries $\alpha_{\Lambda K}$, k=1,….k.

Here k represents the sparsity level of α. The index set represented by $\Lambda_k = \{ \lambda_1, \lambda_2, \ldots \lambda_k \}$ indicates the support of α. $\alpha_{\Lambda k}$ denotes k-dimensional vector that consists of α entries.

$$x = \alpha_{\lambda 1} a_{\lambda 1} + \alpha_{\lambda 2} a_{\lambda 2} + \ldots\ldots\ldots + \alpha_{\lambda M} a_{\lambda M} = A_{\Lambda K}\, \alpha_{\Lambda K} \qquad (16)$$

B. Reconstruction and Classification

The test sample x's α can be reconstructed by solving an optimization problem. Consider a dictionary of training samples A. The α representation is obtained by solving following problem:

$$\alpha' = \arg\min \| \alpha \|_0 \quad \text{subject to } A\alpha = x. \qquad (17)$$

The above constraint can be relaxed to an equality one, in order to approximate the errors in empirical data as follows:

$$\alpha' = \arg\min \|\alpha\|_0 \quad \text{subject to} \quad \| A\alpha - x \| \le \sigma \qquad (18)$$

where σ denotes the error tolerance. The minimization of error within a certain sparsity level can be also indicated as

$$\alpha' = \arg\min \| A\alpha - x \|_2 \quad \text{subject to} \quad \| \alpha \|_0 \le K_0 \qquad (19)$$

Where K_0 denotes the given upper bound on the sparsity level. These problems can be solved by using pursuit algorithms like orthogonal matching pursuit. From the characteristics of recovered sparse vector, the class of test pixel is determined. The error between the training sample and test sample in the mth class is defined as the mth residual. The residual for the observation is calculated with help of $\bar{\alpha}^m$ recovered sparse coefficients associated with training samples in mth class as

$$r^m(x) = \|x - A^m \bar{\alpha}^m\|_2 \quad \text{where } m=1,2,.....M \tag{20}$$

The class that provides minimal residual is selected as the class for the test pixel,

$$\text{class}(x) = \arg\min r^m(x) \quad \text{where } m=1,2....M \tag{21}$$

I. Collaborative Representation based Nearest Neighbor Classifier (CRNN).

In sparse representation based classification, the test sample is represented by training samples of each class where in collaborative representation the test sample is indicated as a linear combination of all training samples. In collaborative representation (CR)-based nearest neighbor classification the test sample is taken in a collaborative form and weights for the representation is determined by l_2-norm minimization-derived closed form solution. Let the parameter α denotes the weighting coefficients and let it be a n× 1 matrix. If a statistical learning model is complex and shows more error instead of relationship then overfitting problem occurs. Inorder to avoid overfiting, regularization techniques are introduced. Regularization is an optimizing process of some representation in order to fix the training data correctly and aims to minimize the computational complexity. There exists l_2 norm regularization which minimizes a loss function. In terms of error function, l_2 norm loss function is known as least squares error. The difference between a target value and estimated values are determined first and then the sum of squares of these estimation is calculated. The weight vector α can be solved using l_2-norm regularization as follows:

$$\alpha = \arg\min_\alpha \| y - X\alpha \|_2^2 + \lambda \|\Gamma_y \alpha\|_2^2 \tag{22}$$

where y is the testing sample, X denotes the training samples, Γ_y denotes biasing Tikhonov matrix and λ is the global regularization parameter. The minimization between the regularization term and residual part is minimized by λ parameter. The regularization parameter, Γ_y is built in the form of

$$\Gamma_y = \begin{bmatrix} \|y - x^{(1)}\|_2 & & 0 \\ & \ddots & \\ 0 & & \|y - x^{(n)}\|_2 \end{bmatrix} \tag{23}$$

where $x^{(1)}$, $x^{(2)}$ ….. $x^{(n)}$ denotes the columns of the training sample matrix X. The closed form solution for the weight vector can be solved as

$$\alpha = (X^T X + \lambda^2 \Gamma_y^T \Gamma_y)^{-1} X^T y. \tag{24}$$

The k largest elements are selected from the calculated α values. The number of weights associated with each class is determined from that and the testing sample's label is found based on majority voting.

$$\text{class}(y) = \arg\max_{l=1,2,...C} N_l(\alpha) \tag{25}$$

where $N_l(\alpha)$ represents the no: of weights belonging to *l*th class.The testing sample is dissimilar to the associated training sample if the calculated weight parameter αis near 0. The training sample that gives the highest weight is the most similar sample to the testing sample.

J.Unsupervised Classification Techniques

Based on the statistical information involved in an image, the original image is classified into several spectral classes. ISODATA and K-means clustering algorithm are the main two approaches used in supervised technique. Different number of clusters is allowed in ISODATA algorithm. In order to calculate a class in ISODATA maximum likelihood decision rule is used. A threshold is defined such that if the number of pixels in the cluster is smaller than that value or if the centers of two clusters are closer than that threshold then those clusters are merged. If the cluster standard deviation exceeds a predefined value and the number of pixels is twice the threshold for the minimum number of pixels, then clusters are split into two different clusters.
Clustering is the process of partitioning an observation into certain number of clusters. K-means clustering attempts to partition n observation into k clusters. First step is to find k cluster centers. Then following steps are repeated until convergence occurs. Each data point is assigned to its closest cluster center. The position of each cluster is set to the mean of all data points belonging to that cluster. The relationship between the clusters and classes are not ensured in unsupervised technique.

III. CONCLUSION

Hyperspectral imagery is a dynamic area in remote sensing. Hyperspectral sensors provide ample information to accurately identify materials from a surface. Hyperspectral imaging has developed new opportunities in the field of image analysis and image classification. Spectral and spatial features are used for hyperspectral image classification. A major problem in hyperspectral image analysis is that spatial resolution gets worse when the spectral resolution increases. Low spatial resolution causes issues in classification, since the pixels may contain more than one land cover classes. Identifying the uncertainty and high dimensionality are the important factors in classification accuracy. Many classification algorithms based on supervised and

unsupervised techniques has developed. Advanced research is needed in hyperspectral imagery to improve the classification accuracy and for better analysis of hyperspectral images.

IV. ACKNOWLEDGMENT

I am thankful to my guide Mrs. Sumithra M D, Assistant Professor of Computer Science and Engineering, for her guidance and encouragement for the paper work.

REFERENCES

[1] Peg Shippert, "Why Use Hyperspectral Imagery", Photogrammetric Engineering & Remote Sensing, April, 2004.

[2] Randall B. Smith, "Introduction to Hyperspectral Imaging" MicroImages, Inc.

[3] Mann Bruce "Locality-Preserving Dimensionality Reduction and Classification for Hyperspectral Image Analysis" Ieee Transactions On Geoscience And Remote Sensing, Vol. 50, No. 4, April 2012.

[4] R.Ablin, C.Helen Sulochana "A Survey of Hyperspectral Image Classification in Remote Sensing" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 8, August 2013.

[5] Farid Melgani, Lorenzo Bruzzone, "Classification of Hyperspectral Remote Sensing Images With Support Vector Machines", IEEE Transactions On Geoscience And Remote Sensing, Vol. 42, No. 8, August 2004.

[6] Alberto Villa, Jón Atli Benediktsson, Jocelyn Chanussot, Christian Jutten, " Hyperspectral Image Classification With Independent Component Discriminant Analysis", IEEE Transactions On Geoscience And Remote Sensing ,2011.

[7] Wei Li, Member,Qian Du,Fan Zhang,Wei Hu, "Collaborative Representation-Based Nearest Neighbor Classifier for Hyperspectral Imagery". IEEE Geoscience and Remote Sensing letters, vol. 12, no. 2, February 2015.

[8] Yi Chen, Nasser M. Nasrabadi, and Trac D. Tran," Hyperspectral Image Classification Using Dictionary-Based Sparse Representation" IEEE Transactions On Geoscience And Remote Sensing, 2011.

[9] Qiang Du "Unsupervised real-time Constrained linear Discriminant anaylsis to hyperspectral image classification" ScienceDirect, Pattern Recognition 40 (2007) 1510-1519.

[10] Wei Li, Eric.W.Tramel,Saurabh Prasad, James E.Fowler "Nearest Regularized Subspace for Hyperspectral Classification", Transactions On Geoscience And Remote Sensing, Vol.52,No.1,January 2014.