

## Hyperspectral Image Classification

Gitanjali S. Korgaonkar, Dr. R. R. Sedamkar

Gitanjali S. Korgaonkar is currently pursuing masters degree in electronics & telecommunication engineering in TCET, Kandivali, Mumbai University, India,

Dr. R.S. Sedamkar is received his Ph.D. degree in 201 and is currently Dean-Academics, Professor and Head of Computer Department at Thakur college of engineering and technology, Mumbai.

### Abstract

This paper proposes weighted decision fusion for classification on a sample hyperspectral image which has two characteristics 1) Spatial and 2) Spectral. Spatial is nothing but within a single band and Spectral means narrow spectral bands over a continuous spectral range. The Weighted Majority voting rule is introduced which is likely to enhance the performance. In this pixels close to the spectral centroid of a segment are assigned higher weights for voting because they are the best representatives of the segment, and pixels deviated from the centroid are not representatives and will have lower weights. Experimental results for SAMSON (Spectroscopic Aerial Mapping System with Onboard Navigation) data show that proposed scheme achieve Overall Accuracy, Kappa Coefficient and Average Accuracy better than Majority voting rule.

**Index Terms**— Hyperspectral image classification, Supervised classification, Unsupervised Classification, Weighted majority voting Rule.

### Introduction

Hyperspectral images contain a wealth of data, but interpreting them requires an understanding of exactly what properties of ground materials we are trying to measure, and how they relate to the measurements actually made by the hyperspectral sensor.[1]

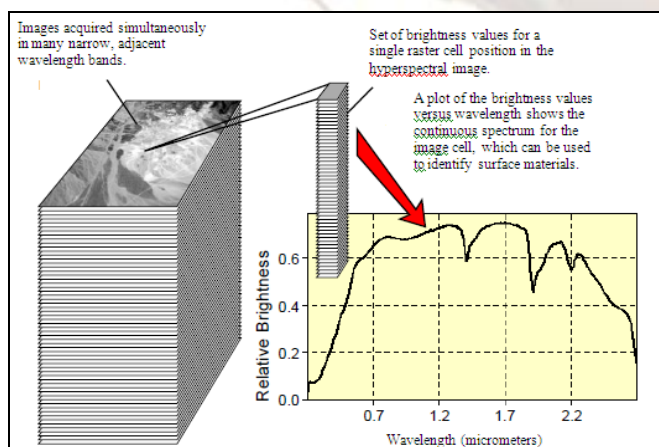


Figure 1. The Hyperspectral Imagery

### 1.1 The Imaging Spectrometer

Hyperspectral images are produced by instruments called imaging spectrometers. The development of these complex sensors has involved the convergence of two related but distinct technologies: spectroscopy and the remote imaging of Earth and planetary surfaces. Spectroscopy is the study of light that is emitted by or reflected from materials and its variation in energy with wavelength. As applied to the field of optical remote sensing, spectroscopy deals with the spectrum of sunlight that is diffusely reflected (scattered) by materials at the Earth's surface.

Instruments called spectrometers (or spectroradiometers) are used to make ground-based or laboratory measurements of the light reflected from a test material. An optical dispersing element such as a grating or prism in the spectrometer splits this light into many narrow, adjacent wavelength bands and the energy in each band is measured by a separate detector. By using hundreds or even thousands of detectors, spectrometers can make spectral measurements of bands as narrow as 0.01 micrometers over a wide wavelength range, typically at least 0.4 to 2.4 micrometers (visible through middle infrared wavelength ranges).

### 1.2 Characteristics of Electromagnetic Radiation:

Hyperspectral imagery involves the sensing of electromagnetic radiation. The electromagnetic spectrum that is used in remote sensing includes the ultraviolet (UV), visible, infrared, and microwave portions of the spectrum.

Figure 2 illustrates the wavelength regions of the electromagnetic spectrum. The spectral portions of near IR and short wave infrared (0.7-3.0  $\mu\text{m}$ ) are called the reflective infrared because measured radiation in this spectral region is mostly composed of reflected sunlight. In contrast, the IR spectrum from 5.0 to 13.0  $\mu\text{m}$  is termed thermal infrared because measurements in this spectral region are generally recording energy radiated from scene elements [2]

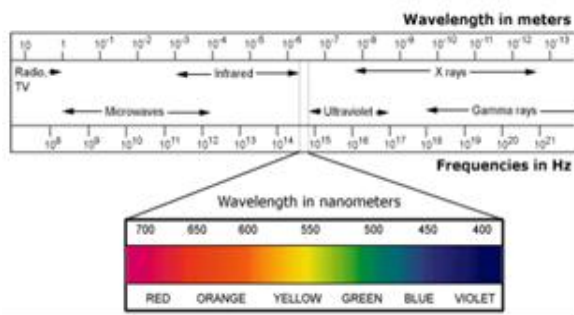


Figure 2 Wavelength regions of the electromagnetic spectrum.

Based on these wavelength regions, remote sensing can be classified into three main categories: visible and reflective infrared remote sensing, thermal remote sensing, and microwave remote sensing. In visible and reflective remote sensing, the radiation that is measured has a solar origin, radiated with a peak wavelength of 0.5  $\mu\text{m}$ . In thermal remote sensing, the measured radiation comes from the observed objects. Materials with normal temperatures ( $\sim 300\text{K}$ ) emit radiation with a peak wavelength of 10.0  $\mu\text{m}$ . Finally, in microwave remote sensing, observations are generally due to reflection of energy radiated by the observing platform (i.e. radar). In this paper, hyperspectral images from the visible and reflective infrared spectrum are used. The majority of currently available sensors with material identification (ID) capability utilize this portion of spectrum.

Typically the source of energy in hyperspectral imagery is the sun. The incident energy from the sun that is not absorbed or scattered in the atmosphere interacts with the earth's surface, where it is absorbed, transmitted, or reflected. Additionally, the electromagnetic radiation has specific properties that are predictable in its interaction with materials and transmission through the atmosphere. Therefore, by measuring the electromagnetic radiation in narrow wavelength bands, the resulting spectral signatures can be used, in principle, to uniquely characterize and identify any given material.

All materials have unique spectral characteristics because they absorb, reflect, and emit radiation in a unique way. For example, in the visible portion of the spectrum, a leaf appears green because it absorbs in the blue and red regions of the spectrum and reflects in the green region. These variations in absorption, reflection, and emission are due to the material composition. Differences in spectral response due to absorption, transmission, and reflection cause materials to have a unique spectral signature. Figure 3 illustrates the spectral signatures of three pixels, respectively dominated by seawater, vegetation, and concrete. Comparing the spectra between seawater and vegetation, it is observed that they reflect similarly in the visible wavelengths but differently in the infrared portion. Also, concrete has a different spectral signature compared

to the other two in specific wavelength regions. Therefore, by knowing a material's spectral signature, it is possible for this material to be detected in pixels within an image. Libraries with the characteristic spectra of various materials of interest exist, and these spectral signatures can be compared with measured spectra in order to identify the features in an image. In the early work on spectral imagery, computational limits prevented full exploitation of such data. Computational power in the latter portion of the 1990's made routine use of spectral imagery much more practical.

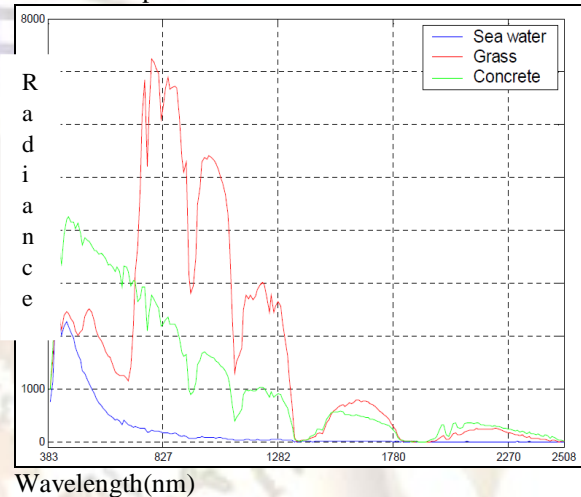


Figure 3. Hyperspectral pixel spectra.

Identifying groups of pixels that have similar spectral characteristics and determining the various features or land cover classes represented by these groups is an important part of image analysis. This form of analysis is known as classification. Visual classification relies on the analyst's ability to use visual elements (tone, contrast, shape, etc) to classify an image. Digital image classification is based on the spectral information used to create the image and classifies each individual pixel based on its spectral characteristics. The result of a classification is that all pixels in an image are assigned to particular classes or themes (e.g. water, coniferous forest, deciduous forest, corn, wheat, etc.), resulting in a classified image that is essentially a thematic map of the original image. The theme of the classification is selectable, thus a classification can be performed to observe land use patterns, geology, vegetation types, or rainfall.

In classifying an image we must distinguish between spectral classes and information classes. Spectral classes are groups of pixels that have nearly uniform spectral characteristics. Information classes are various themes or groups we are attempting to identify in an image. Information classes may include such classes as deciduous and coniferous forests, various crop types, or inland bodies of water. The objective of image classification is to match the spectral classes in the data to the in-

formation classes of interest.

Different conditions like the sun illumination, snow cover, other atmospheric conditions, crop growth in a particular season etc. conditions effect the classification process.

## 2. TYPES OF CLASSIFICATION

Image classification is perhaps the most important part of digital image analysis. It is very nice to have a "pretty picture" or an image, showing a magnitude of colors illustrating various features of the underlying terrain, but it is quite useless unless to know what the colors mean. Image classification or the 'partition of the feature space' can be done in two ways: Supervised and Unsupervised classification. In the supervised classification, the operator defines his classes of interest by selecting the training samples in an image on the basis of his/her knowledge of the area. In the unsupervised classification, the image is partitioned into homogenous spectral clusters using some clustering algorithm. These spectral clusters are accomplished on the basis of some spectral similarities. In the classification application, we need to preserve the features that are useful for discrimination among classes.

### 2.1 Supervised Classification

With supervised classification, we identify examples of the Information classes (i.e., land cover type) of interest in the image. These are called "training sites". The image processing software system is then used to develop a statistical characterization of the reflectance for each information class. This stage is often called "signature analysis" and may involve developing a characterization as simple as the mean or the range of reflectance on each bands, or as complex as detailed analyses of the mean, variances and covariance over all bands. Once a statistical characterization has been achieved for each information class, the image is then classified by examining the reflectance for each pixel and making a decision about which of the signatures it resembles most[3].

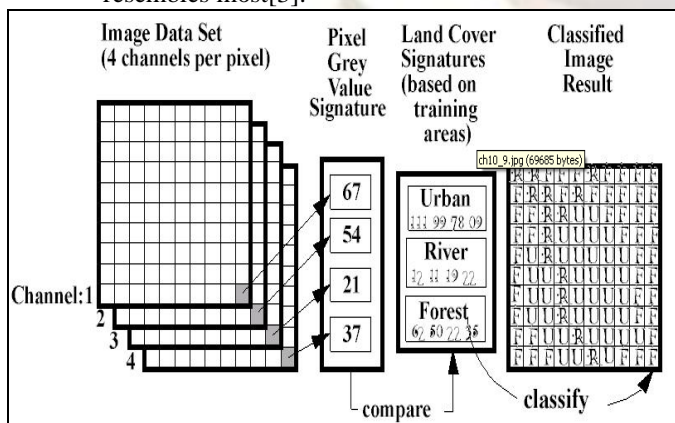


Figure 4. Supervised Classification

### 2.1.1. Support Vector Machine

One of the best algorithm used for supervised classification is Support Vector Machine(SVM). A support vector machine (SVM) is a concept in statistics and computer science for a set of related supervised learning methods that analyze data and recognize patterns, used for classification and regression analysis. The standard SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the input, making the SVM a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on. SVMs characterize classes using a geometrical criterion rather than statistical criteria. They seek a separating hyperplane maximizing the distance with the closest training samples for two classes. This approach allows SVMs to have a very high capability of generalization and, as a consequence, only require a few training samples. For non linearly separable data, SVMs use the kernel trick to map the data onto a higher dimensional space where they are linearly separable [4].

Here, we consider multiclass SVMs without any feature reduction of the original hyperspectral data. The standard

Gaussian radial basis kernel with L2-norm distance is used. This algorithm was proved to provide interesting classification accuracy, even in the case of limited training set [5]. Note that other kernels could be considered, such as the spectral angle mapper which basically computes the angle between two vectors in the vector space [5][6][7]. In the following, we present the used classifier and start by briefly recalling the general mathematical formulation of SVMs. Starting from the linearly separable case, optimal hyperplanes are introduced, then the classification problem is modified to handle non-linearly separable data and a brief description of multiclass strategies is given. Finally, kernel methods are presented.

#### Linear SVM

For a two-class problem in a n-dimensional space  $\mathbb{R}^n$ , we assume that l training samples,  $x_i \in \mathbb{R}^n$ , are available with their corresponding labels  $y_i = \pm 1$ ,  $S = \{(x_i, y_i) \mid i \in [1, l]\}$ . The SVM method consists of finding the hyperplane that maximizes the margin i.e., the distance to the closest training data points in both classes. Noting  $w \in \mathbb{R}^n$

$\mathbb{R}^n$  as the vector normal to the hyperplane and  $b \in \mathbb{R}$  as the bias, the hyperplane  $H_p$  is defined as  
 $(w, X) + b = 0, \forall X \in H$  (1)

Where  $\langle W, X \rangle$  is the inner product between  $w$  and  $x$ . If  $x \notin H_p$  then  $f(x) = \langle W, X \rangle + b$  is the distance of  $x$  to  $H_p$ . The sign of  $f$  corresponds to decision function  $y = \text{sgn}(f(x))$ . The optimal Parameters ( $w, b$ ) are found by solving

$$\min \left[ \frac{\|w\|^2}{2} + c \sum_{i=1}^l \xi_i \right] \quad (2)$$

Subject to

$$y_i((W, X_i) + b) \geq 1 - \xi_i, \xi_i \geq 0 \forall i \in [1, l] \quad (3)$$

The solution vector is a linear combination of some samples of the training set, whose  $\alpha_i$  is non-zero, called Support Vectors. The hyperplane decision function can thus be written as:

$$y_u = \text{sgn} \left( \sum_{i=1}^l y_i \alpha_i \langle X_u, X_i \rangle + b \right) \quad (4)$$

The maximum margin linear classifier is the linear classifier with the, um, maximum margin. This is the simplest kind of SVM (Called an LSVM)

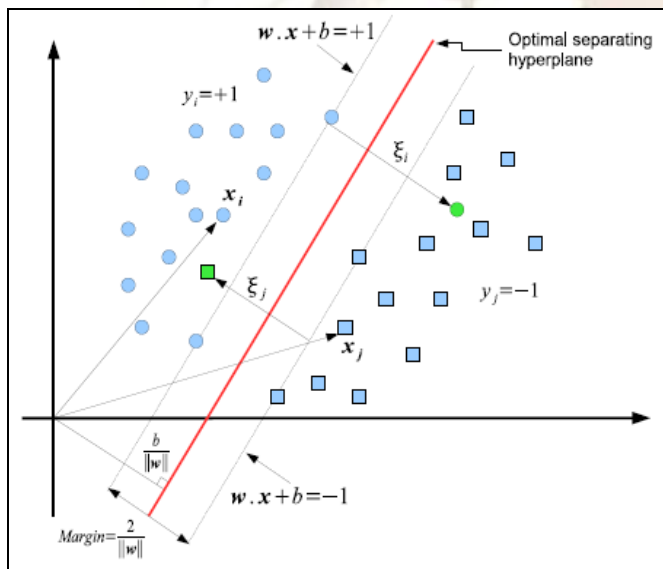


Figure 5 Maximum Margin Of SVM

There is one non separable vector in each class  
Non linear support vector machine

If the training data is not linearly separable, there is no straight hyperplane that can separate the classes. In order to learn a nonlinear function in that case, linear SVMs must be extended to nonlinear SVMs for the classification of nonlinearly separable data. The process of finding classification functions using nonlinear SVMs consists of two steps. First, the input vectors are transformed into high-dimensional feature vectors where the training data can be linearly separated. Then, SVMs are used to

find the hyperplane of maximal margin in the new feature space. The separating hyperplane becomes a linear function in the transformed feature space but a nonlinear function in the original input space. Let  $x$  be a vector in the  $n$ -dimensional input space and  $j(\cdot)$  be a nonlinear mapping function from the input space to the high-dimensional feature space. The hyperplane representing the decision boundary in the feature space is defined as follows.

$$w \cdot j(x) - b = 0 \quad (5)$$

where  $w$  denotes a weight vector that can map the training data in the high dimensional feature space to the output space, and  $b$  is the bias. Using the  $j(\cdot)$  function, the weight becomes

$$w = \sum \alpha_i y_i \varphi(x_i) \quad (6)$$

The decision function of

$$F(X) = \sum_{i=1}^m \alpha_i y_i X_i \cdot X - b \quad (7)$$

$$\text{becomes} \quad F(X) = \sum_{i=1}^m \alpha_i y_i \varphi(X_i) \cdot \varphi(X) - b \quad (8)$$

Note that the feature mapping functions in the optimization problem and also in the classifying function always appear as dot products, e.g.,  $j(x_i) \cdot j(x_j)$ .  $j(x_i) \cdot j(x_j)$  is the inner product between pairs of vectors in the transformed feature space. Computing the inner product in the transformed feature space seems to be quite complex and suffer from the curse of dimensionality problem. To avoid this problem, the kernel trick is used. The kernel trick replaces the inner product in the feature space with a kernel function  $K$  in the original input space as follows.

$$K(u, v) = j(u) \cdot j(v) \quad (9)$$

The Mercer's theorem proves that a kernel function  $K$  is valid, if and only if, the following conditions are satisfied, for any function  $y(x)$ .

$$\int \int k(u, v) \Psi(u) \Psi(v) dx dy \leq 0 \quad (10)$$

$$\text{Where } \int \Psi(x)^2 dx \leq 0$$

The Mercer's theorem ensures that the kernel function can be always expressed as the inner product between pairs of input vectors in some high-dimensional space, thus the inner product can be calculated using the kernel function only with input vectors in the original space without transforming the input vectors into the high dimensional feature vectors.

The classification function becomes:

$$F(X) = \sum_i \alpha_i K(X_i, X) - b \quad (11)$$

Since  $K(\cdot)$  is computed in the input space, no feature transformation will be actually done or  $j(\cdot)$  will be computed, and thus the weight vector  $w = \sum_i \alpha_i y_i$  will not be computed either in nonlinear SVMs.

The followings are popularly used kernel functions.

- Polynomial:  $K(a,b) = (a \cdot b + 1)^d$
- Radial Basis Function (RBF):  $K(a,b) = \exp(-g \|a-b\|^2)$
- Sigmoid:  $K(a,b) = \tanh(ka \cdot b + c)$ [8].

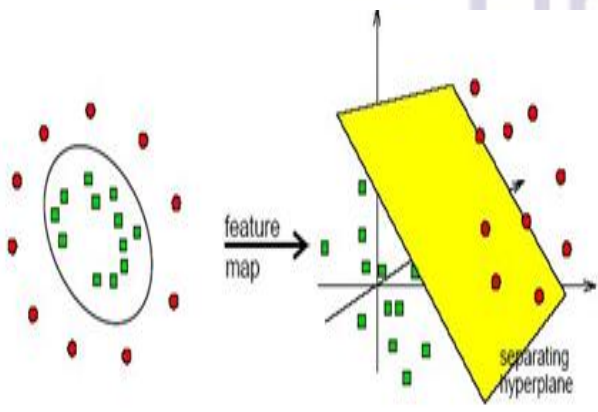


Figure 6 Non Linear SVM 's Separation

The kernel function transforms the data into a higher dimensional space to make it possible to perform the separation.

## 2.2 Unsupervised Classification

Unsupervised classification is a method which examines a large number of unknown pixels and divides into a number of classes based on natural groupings present in the image values. Unlike supervised classification, unsupervised classification does not require analyst-specified training data. The basic premise is that values within a given cover type should be close together in the measurement space (i.e. have similar gray levels), whereas data in different classes should be comparatively well separated (i.e. have very different gray levels).

The classes that result from unsupervised classification are spectral classes which based on natural groupings of the image values, the identity of the spectral class will not be initially known, must compare classified data to some form of reference data (such as larger scale imagery, maps, or site visits) to determine the identity and informational values of the spectral classes. Thus, in the supervised approach, to define useful information categories and then examine their spectral separability; in the unsupervised approach the computer determines spectrally separable class, and then define their information value.

Unsupervised classification is becoming increasingly popular in agencies involved in long

term GIS database maintenance. The reason is that there are now systems that use clustering procedures that are extremely fast and require little in the nature of operational parameters. Thus it is becoming possible to train GIS analysis with only a general familiarity with remote sensing to undertake classifications that meet typical map accuracy standards.

With suitable ground truth accuracy assessment procedures, this tool can provide a remarkably rapid means of producing quality land cover data on a continuing basis[3].

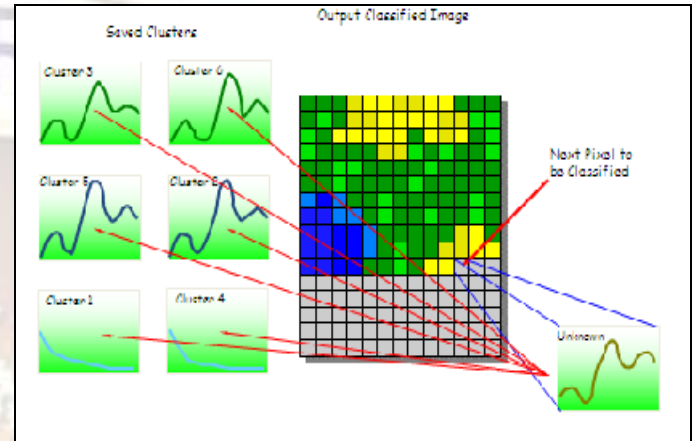


Figure 7 Unsupervised Classification

### 2.2.1 K-Mean Algorithm

K-means is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume  $k$  clusters) fixed a priori. The main idea is to define  $k$  centroids, one for each cluster. These centroids should be placed in a cunning way because of different location causes different results. So, the better choice is to place them as much as possible far away from each other. The next step is to take each point belonging to a given data set and associate it to the nearest centroid. When no point is pending, the first step is completed and an early group age is done. At this point we need to re-calculate  $k$  new centroids as barycenters of the clusters resulting from the previous step. After we have these  $k$  new centroids, a new binding has to be done between the same data set points and the nearest new centroid. A loop has been generated. As a result of this loop we may notice that the  $k$  centroids change their location step by step until no more changes are done. In other words centroids do not move any more.[9]

Finally, this algorithm aims at minimizing an objective function, in this case a squared error function. The objective function:

$$J = \sum_{j=1}^k \sum_{i=1}^n \|x_i^{(j)} - c_j\|^2 \quad (12)$$

where  $\|X_i(j) - C_j\|^2$  is a chosen distance measure between a data point  $X_i(j)$  and the cluster centre  $C_j$ , is an indicator of the distance of the  $n$  data points from their respective cluster centres.

The algorithm is composed of the following steps:

- 1) Place  $K$  points into the space represented by the objects that are being clustered. These points represent initial group centroids.
- 2) Assign each object to the group that has the closest centroid.
- 3) When all objects have been assigned, recalculate the positions of the  $K$  centroids.
- 4) Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

Although it can be proved that the procedure will always terminate, the  $k$ -means algorithm does not necessarily find the most optimal configuration, corresponding to the global objective function minimum. The algorithm is also significantly sensitive to the initial randomly selected cluster centres. The  $k$ -means algorithm can be run multiple times to reduce this effect.

$K$ -means is a simple algorithm that has been adapted to many problem domains. As we are going to see, it is a good candidate for extension to work with fuzzy feature vectors.

### 3. DECISION FUSION

A decision fusion approach is developed to combine the results from supervised and unsupervised classifiers. The final output

takes advantage of the power of a support-vector machine- based supervised classification in class separation and the capability of an unsupervised classifier, such as  $K$ -means clustering, in reducing trivial spectral variation impact in homogeneous regions [10][11][12].

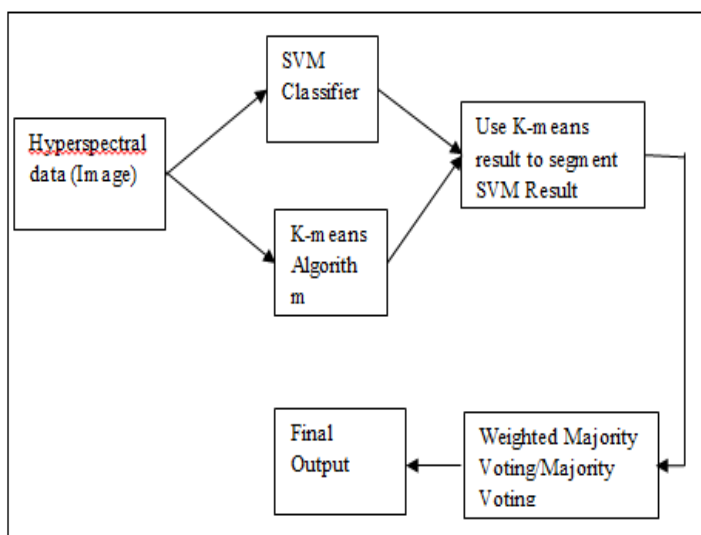


Figure 8 Decision Fusion Approach

Decision fusion has been widely used to improve the overall classification accuracy. Most decision fusion approaches mainly focus on supervised classifiers as base learner, i.e., all classifiers need training, so the classification results can only be as good as training data. To avoid the possible negative influence from the limited quality of training data, we are motivated to propose a method that can combine supervised and unsupervised classifiers.

A supervised classifier may provide better classification than an unsupervised classifier. However, in addition to training data limitation, a supervised classifier may result in over-classification for some homogeneous areas. Although it may be less powerful, an unsupervised classifier can generally well classify those spectrally homogeneous areas. Thus, fusing supervised and unsupervised classification may yield better performance since the impact from trivial spectral variations may be alleviated and the subtle difference between spectrally similar pixels may not be exaggerated. Although individual classifiers are pixel-based, the final fused classification has a similar result to an object-based classifier.

### 4. VOTING RULE

In this paper, we used a weighted majority voting (WMV) rule to further improve classification accuracy. The final fusion output is dependent on not only the classification accuracy of the supervised classifier and the chosen unsupervised classifier but also the fusion rule. The MV rule is the most frequently used. It is simple and requires no training. The MV rule may have some limitations. For instance, it only counts the number of pixels in each class in a segment, but does not consider pixel "quality" since it treats all the pixels equally; as a result, the spectral variations from pixel to pixel is completely ignored in a relatively homogeneous segment. Intuitively, pixels close to the spectral centroid of a segment should be assigned higher weights for voting because they are the best representatives of the segment, and pixels deviated from the centroid are not representatives and should have lower weights. The resulting WMV rule can alleviate the impact from outliers, whose spectral signatures are quite different from the rest of pixels although they are grouped into the same segment. To gauge such deviation, the Mahalanobis distance can be used to measure the distance between a pixel and the centroid.

#### 4.1 Accuracy Assessment:

Accuracy assessment is an important step in the classification process. The goal is to quantitatively determine how effectively pixels were grouped into the correct feature classes in the area under investigation. The land cover types derived from digital image interpretation and analysis requires validation with data obtained from ground

verification. For the accuracy assessment, a coefficient of agreement between classified image data and ground reference data were calculated using Kappa. The accuracy was determined by kappa statistics in order to test whether any difference exists in the interpretation work.

Briefly, Kappa statistic considers a measure of overall accuracy of image classification and individual category accuracy as a means of actual agreement between classification and observation. The value of Kappa lies between 0 and 1, where 0 represents agreement due to chance only. Meanwhile 1 represents complete agreement between the two data sets. Negative values can occur but they are spurious. It is usually expressed as a percentage (%). [13] claimed that the Kappa statistic has been shown to be a statistically more sophisticated measure of classifier agreement and thus gives better interclass discrimination than overall accuracy.

## 5. RESULT AND ANALYSIS

The software implementation of the algorithm is written in a Matlab environment using Matlab7.10 software. This hyperspectral dataset was generated by the SAMSON sensor. The SAMSON components are integrated with the flight management system to enable real-time management of data collection and flight orientation.

The HSI sensor incorporates improved spectral alignment, radiometric calibrations and atmospheric correction procedures. The charge coupled device (CCD) used in the SAMSON sensor is a thinned, backside-illuminated CCD (produced by Sarnoff Imaging). The CCD manufacturing process greatly increases the quantum efficiency from about 5% to 60% at 400 nm, and from about 40% to 85% at 700 nm, with a band width of 3.2nm. The data was collected by the Florida Environmental Research Institute as part of the GOES-R sponsored experiment. The instrument flown during the collect is the SAMSON, a push-broom, visible to near IR, hyperspectral sensor. This sensor was designed and developed by FERL. The following paper describes the basic design of the sensor: Kohler et al. (2006)[14]. This dataset utilized the new radiometric calibration technique specifically designed to characterize and correct the stray light found within the sensor. The following paper describes the approach: Kohler et al. (2004). It has samples = 952, lines = 952, bands = 156.



Figure 9 Hyperspectral Image

### 5.1 Classification output:

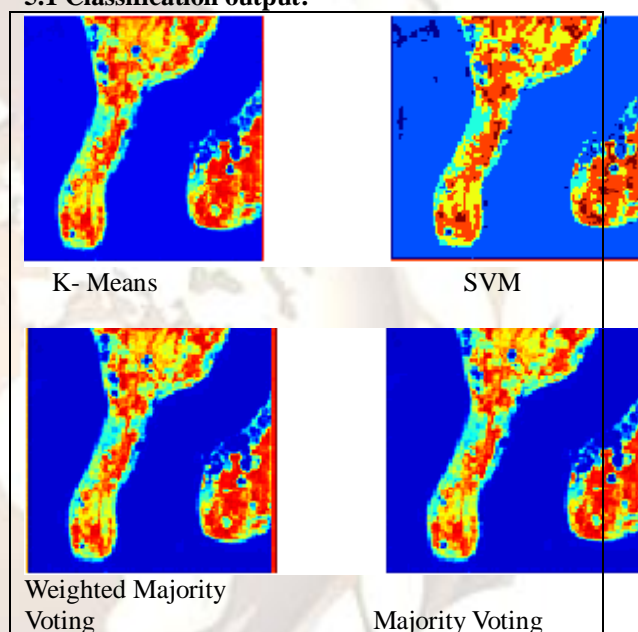


Figure 10. Classified Output for band 75

Bands	OA=	AA=	OA=	AA=
	90.5886	90.585	99.0039	99
	Majority Voting Kappa Coefficient		Weighted Majority Voting Kappa Coefficient	
1	91.4158		99.9080	
10	90.8216		99.2586	
25	91.0987		99.5614	
45	90.7436		99.1733	
75	88.8225		97.0738	
50	90.3567		98.7505	

100	89.0256	97.2958
125	91.4878	99.9867
145	89.0309	97.3015
150	89.0639	97.3376
156	89.1105	97.3886

Table 1. Comparison of different Bands

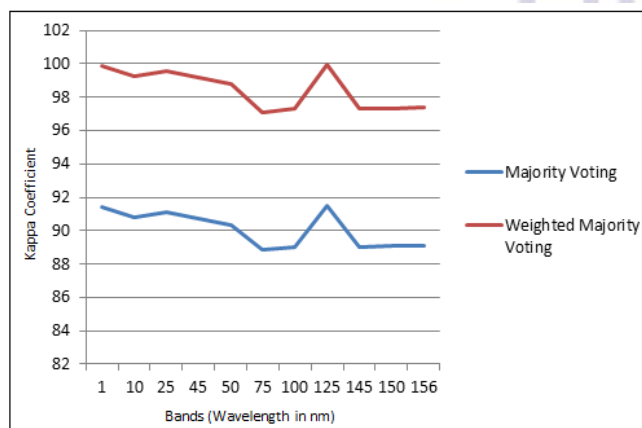


Figure 11. Kappa Coefficient of hyperspectral image at different wavelengths

## 6. CONCLUSION

Decision Fusion is the most widely used methods for Classification of hyperspectral image. In this paper, we used a weighted decision fusion approach. The final output take advantage of the power of supervised classification in class separation and the capability of unsupervised classifier in reducing the impact from spectral variations in homogeneous regions.

The Weighted Decision Fusion Method were examined using SAMSON hyperspectral sample data, which shows that the classification accuracy and the kappa coefficient is better than that using MV method.

## REFERENCES

- [1] Documentation Tutorials on hyperspec by Radall B. Smith
- [2] Peg Shippert "Introduction to Hyperspectral Image Analysis" Earth Science Applications Specialist Research Systems, Inc
- [3] <http://www.sc.chula.ac.th/courseware/2309507/Lecture/remote18.htm>
- [4] B. Scholkopf and A. J. Smola, Learning with Kernels. MIT Press, 2002.
- [5] M. Fauvel, J. Chanussot, and J. A. Benediktsson, "Kernels Evaluation for

multiclass classification of hyperspectral-remote sensing data," in Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, Toulouse, France, 2006.

- [6] M. L. G. Mercier, "Support vector machines for hyperspectral image classification with spectral-based kernels," in Geoscience and Remote Sensing Symposium, vol. 1. IGARSS '03. Proceedings, July 2003.
- [7] N. Keshava, "Distance metrics and band selection in hyperspectral processing with application to material identification and spectral libraries," IEEE Trans. on Geoscience and Remote Sensing, vol. 42, pp. 1552-1565, July 2004.
- [8] <http://people.sabanciuniv.edu/berrin/cs512/lectures/11-svmtutorial2.pdf> by Hwanjo Yu and Sungchul Kim
- [9] [http://home.dei.polimi.it/matteucc/Custering/tutorial\\_html/kmeans.html](http://home.dei.polimi.it/matteucc/Custering/tutorial_html/kmeans.html)
- [10] H. Yang, B. Ma, and Q. Du, "Decision fusion for supervised and unsupervised hyperspectral image classification," in Proc. IEEE Geosci. Remote Sens. Symp., Jul. 2009, pp. 948-951.
- [11] M. Petrakos, J. A. Benediktsson, and I. Kanellopoulos, "The effect of Classifier agreement on the accuracy of the combined classifier in decision level fusion," IEEE Trans. Geosci. Remote Sens., Vol 39, no. 11, pp. 2539-2546, Nov. 2001.
- [12] A. Cheryadat, L. M. Bruce and A. Mathur, "Decision Level Fusion with best-bases for hyperspectral Classification," in Proc. IEEE Workshop Adv. Tech. Anal. Remotely Sensed Data, 2003, pp. 399-406
- [13] Fitzgerald, R. W. and Lees, B. G. 1994. Assessing the classification accuracy of multi-sources remote sensing data. Remote Sensing of Environment, 47: 362-368.
- [14] [www.opticks.org/confluence/display/opticks/sample+data](http://www.opticks.org/confluence/display/opticks/sample+data)