Hyperspectral Image Classification on Decision level fusion

GitanjaliS.Korgaonkar
Post Graduate Student, TCET
Thakur ShyamnarayanMarg
Thakur Village, Kandivali(E).

Dr.R.R.Sedamkar
Professor & Head of Computer Engg.,TCET
Thakur ShyamnarayanMarg
Thakur Village, Kandivali(E).

KiranBhandari
Assistant Professor, TCET
Thakur ShyamnarayanMarg
Thakur Village, Kandivali(E).

ABSTRACT
In this paper different types of image classification will be studied. Decision level fusion, using a specific criterion or algorithm to integrate the classified results from different classifiers, has shown great benefits to improve classification accuracy of multi-source remote sensing images. Based on a survey to hyperspectral remote sensing classification techniques and decision level fusion algorithms, some issues on hyperspectral remote sensing image classification based on decision level fusion are explored. In this three decision level fusion methods and four schemes for input data are used to hyperspectral remote sensing image classification.

General Terms
Hyperspectral Image, Decision Fusion.

Keywords
Hyperspectral image classification, Supervised classification, Unsupervised Classification, Fusion, Decision Fusion.

1. INTRODUCTION
The "hyper" in hyperspectral means "over" as in "too many" and refers to the large number of measured wavelength bands. Hyperspectral images are spectrally overdetermined, which means that they provide ample spectral information to identify and distinguish spectrally unique materials. Hyperspectral imagery provides the potential for more accurate and detailed information extraction than possible with any other type of remotely sensed data [1].Hyperspectral images are 3D data, with a spectral signature for the scene spread over several bands. Traditionally, the high dimensional spectral information is used to perform a pixel-by-pixel classification of the scene. Band subset selection/feature extraction methods have been developed to improve the performance of parametric classifiers such as ML, Distance Classifiers and clustering methods. However, the classification accuracies of these methods do not match those developed for gray scale/color images.

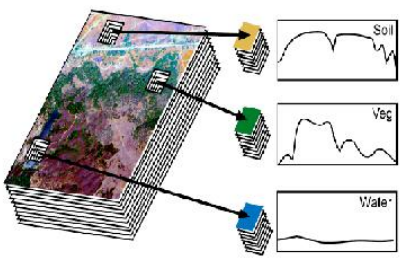

Figure1. The concept of hyperspectralimagery

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 information classes of interest.

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. Two main classification methods are Supervised Classification and Unsupervised Classification.

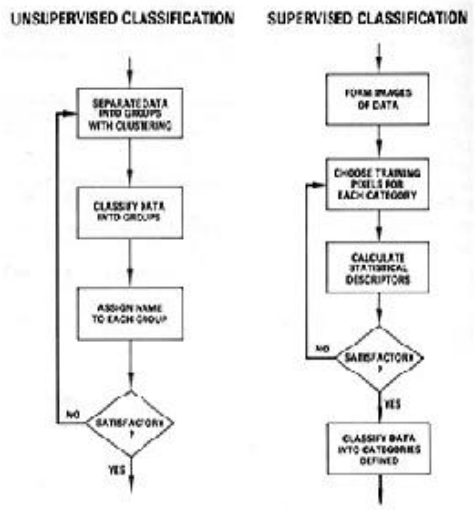

Figure 2. Types Of Classification

2.1Supervised 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 rage 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[2].

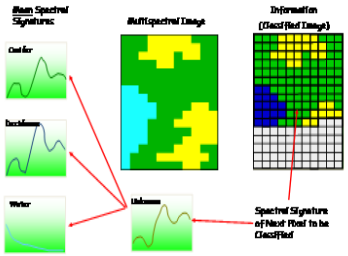

Figure 3. Supervised Classification

Following are sub-type of Supervised Classification:

2.1.1 *Parallelepiped Classification*

The parallelepiped classifier uses the class limits and stored in each class signature to determine if a given pixel falls within the class or not. The class limits specify the dimensions (in standard deviation units) of each side of a parallelepiped surrounding the mean of the class in feature space. If the pixel falls inside the parallelepiped, it is assigned to the class. However, if the pixel falls within more than one class, it is put in the overlap class. If the pixel does not fall inside any class, it is assigned to the null class.

The parallelepiped classifier is typically used when speed is required. The drawback is (in many cases) poor accuracy and a large number of pixels classified as ties.

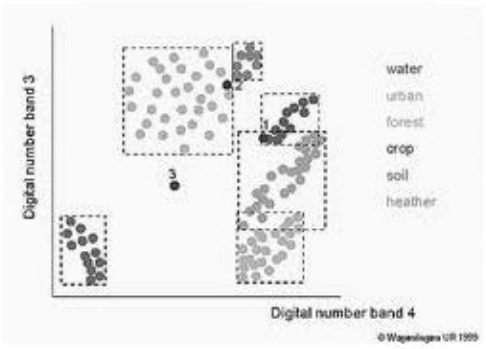

Figure 4 Parallelepiped Classification

2.1.2 *Minimum distance to mean Classification*

Minimum distance classifies image data on a database file using a set of 256 possible class signature segments as specified by signature parameter. Each segment specified in signature, for example, stores signature data pertaining to a particular class. Only the mean vector in each class signature segment is used. Other data, such as standard deviations and covariance matrices, are ignored (though the maximum likelihood classifier uses this).

The result of the classification is a theme map directed to a specified database image channel. A theme map encodes each class with a unique gray level. The gray-level value used to encode a class is specified when the class signature is created. If the theme map is later transferred to the display, then a pseudo-color table should be loaded so that each class is represented by a different color.

2.1.3 *Maximum likelihood Classification*

The algorithm used by the Maximum Likelihood Classification tool is based on two principles:

The cells in each class sample in the multidimensional space are normally distributed.

Bayes' theorem of decision making.

The Maximum Likelihood Classification tool considers both the variances and covariances of the class signatures when assigning each cell to one of the classes represented in the signature file. With the assumption that the distribution of a class sample is normal, a class can be characterized by the mean vector and the covariance matrix. Given these two characteristics for each cell value, the statistical probability is computed for each class to determine the membership of the cells to the class. When the default EQUAL a priori option is specified, each cell is classified to the class to which it has the highest probability of being a member.

.Maximum likelihood Classification is a statistical decision criterion to assist in the classification of overlapping signatures; pixels are assigned to the class of highest probability.

The maximum likelihood classifier is considered to give more accurate results than parallelepiped classification however it is much slower due to extra computations.

2.2 Unsupervised Classification

Unsupervised classification is a method which examines a large number of unknown pixels and divides into a number of

classed 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 classed 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 from 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 canprovide a remarkably rapid means of producing quality land cover data on a continuing basis[2].

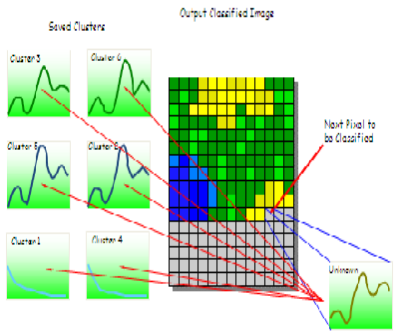

Figure 5 Unsupervised Classification

Following are Sub-type of 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 shoud be placed in a cunning way because of different location causes different result. 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 groupage 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.

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

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

where $\|x_i^{(j)}-C_j\|^2$ is a chosen distance measure between a data point $X_i^{(j)}$ and the cluster centreC_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.

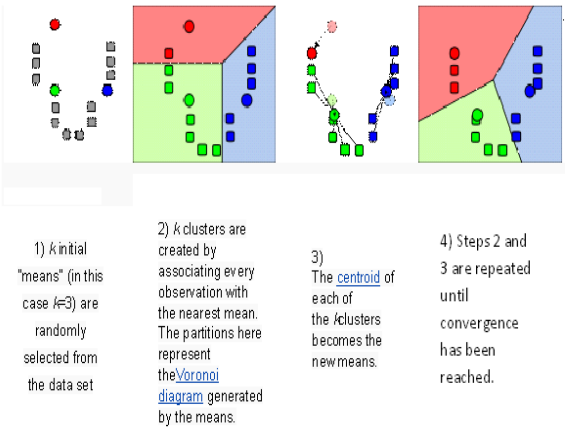

Figure 6 Steps for K-means Algorithm

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.

2.2.2 *ISODATA Algorithm*

In general, both of them assign first an arbitrary initial cluster vector. The second step classifies each pixel to the closest cluster. In the third step the new cluster mean vectors are calculated based on all the pixels in one cluster. The second and third steps are repeated until the "change" between the

iteration is small. The "change" can be defined in several different ways, either by measuring the distances the mean cluster vector have changed from one iteration to another or by the percentage of pixels that have changed between iterations.

"Iterative Self-Organizing Data Analysis Technique" The ISODATA algorithm has some further refinements by splitting and merging of clusters .Clusters are merged if either the number of members (pixel) in a cluster is less than a certain threshold or if the centers of two clusters are closer than a certain threshold. Clusters are split into two different clusters if the cluster standard deviation exceeds a predefined value and the number of members (pixels) is twice the threshold for the minimum number of members.

The ISODATA algorithm is similar to the k-means algorithm with the distinct difference that the ISODATA algorithm allows for different number of clusters while the k-means assumes that the number of clusters is known a priori. The objective of the k-means algorithm is to minimize the within cluster variability. The objective function (which is to be minimized) is the *sums of squares distances* (errors) between each pixel and its assigned cluster center.

$$SS_{\text{distances}} = \sum_{\forall x} [x - C(x)]^2 \quad (2)$$

where $C(x)$ is the mean of the cluster that pixel x is assigned to.

Minimizing the $SS_{distances}$ is equivalent to minimizing the Mean Squared Error (MSE). The MSE is a measure of the within cluster variability.

$$MSE = \frac{\sum_{\forall x} [x - C(x)]^2}{(N-c)b} = \frac{SS_{\text{distances}}}{(N-c)b} \quad (3)$$

Where N is the number of pixels, c indicates the number of clusters, and b is the number of spectral bands. Note that the MSE is not the objective function of the ISODATA algorithm. However, the ISODATA algorithm tends to also minimize the MSE.

K-means (just as the ISODATA algorithm) is very sensitive to initial starting values. For two classifications with different initial values and resulting different classification one could choose the classification with the smallest MSE (since this is the objective function to be minimized). However, as we show later, for two different initial values the differences in respects to the MSE are often very small while the classifications are very different. Visually it is often not clear that the classification with the smaller MSE is truly the better classification.

From a statistical viewpoint, the clusters obtained by k-mean can be interpreted as the Maximum Likelihood Estimates (MLE) for the cluster means if we assume that each cluster comes from a spherical Normal distribution with different means but identical variance (and zero covariance).

This touch upon a general disadvantage of the k-means algorithm (and similarly the ISODATA algorithm): k-means works best for images with clusters that are spherical and that have the same variance. This is often not true for remote sensing images. For example, a cluster with "desert" pixels is compact/circular. A "forest" cluster, however, is usually more or less elongated/oval with a much larger variability compared to the "desert" cluster. While the "desert" cluster is usually very well detected by the k-means algorithm as one distinct cluster, the "forest" cluster is often split up into several smaller clusters. The way the "forest" cluster is split up can vary quite a bit for different starting values and is thus arbitrary.

3. FUSION

Fusion is the act or process of combining or associating data or information regarding one or more entities considered in an explicit or implicit knowledge framework to improve one's capability (or provide a new capability) for detection, identification, or characterization of that entity.

There are three types of Fusion:

3.1 Observation fusion

It involves fusing information from different sensors of the same physical phenomenon, such as image intelligence (IMINT), and also fusing information from sensors of different phenomena, such as fusing laser imaging detection and ranging, hyperspectral (images recording visible plus infrared and/or ultraviolet light), and overhead persistent infrared.

Image intelligence is a technical method of gathering intelligence. It is based on photography and visual monitoring from aeroplanes and (in big states) from satellites. IMINT is entirely authentic, without risks, but it is very expensive.

LIDAR (Light Detection And Ranging, also LADAR) is an optical remote sensing technology that can measure the distance to, or other properties of a target by illuminating the target with light, often using pulses from a laser.

3.2 Object/feature fusion

It involves fusing different data types from different INTs, such as fusing IMINT and SIGINT to yield information resources that are more powerful, flexible and accurate than the original sources.

Signals intelligence (often contracted to SIGINT) is intelligence-gatheringby interception of signals, whether between people ("communications intelligence"—COMINT), whether involving electronic signals not directly used in communication ("electronic intelligence"—ELINT), or combinations of the two.

3.3 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 [2],[7].

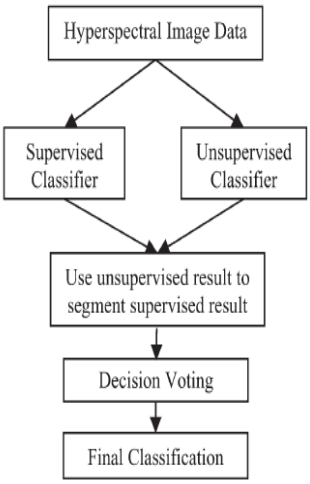

Figure 7 Decision Fusion Approach

In this three decision level fusion methods and four schemes for input data are used to hyperspectral remote sensing image classification.

The first scheme is the most common way in which the original hyperspectral dataset is used by different classifiers. The second scheme is an improved way, in which all classifiers still use identical input dataset, but the dataset consists of both the original data and texture features derived from original data. In the third scheme, all wavebands are divided into different groups based on inter-band correlationanalysis, and each group of data together with texture feature are used to a specific classifier, which means that the input for multiple classifiers are different but everygroup of data should be a representative subset of original data. In the fourth scheme, the first ten components derived by MNF transformation to original data and texture features are used as input of different classifiers.

The Minimum (or Maximum) Noise Transform (MNF) is a modification to principal components analysis that normalizes each band of the hyperspectral image by its noise level prior to processing. This acts to reduce the influence of noise in the transformed images as the noisier bands of the hyperspectral image are deemphasized. Generally noise is calculated by using "shift-difference" statistics. In this method, the difference between adjacent pixels is assumed to be an estimate of noise.

In decision level fusion, three fusion methods are experimented: improved evidence theory, linear consensus and SVM combiner.

3.3.1 Improved D-S evidence theory

Evidence theory was also called Dempster-Shafer evidence theory. Compared with Bayesian theory, D-S evidence theory assigns probability to sets and can handle the uncertainty caused by unknown factors. D-S evidence theory uses discrimination framework, confidence function, likelihood function and probability allocation function to represent and process knowledge. For remote sensing image, different classifier may generate different classified labels, which result in the generation of evidence with high contradiction, so the modified evidence combination is applied to classification integration of hyperspectral remote sensing images.

There are three main reasons why the D-S evidence theory should be taken into account when it comes to information fusion. First of all, since the D-S evidence theory supports the representation of both imprecision and uncertainty, it is considered to be a more flexible and general approach than the traditional probability theory. Secondly, D-S offers the possibility of coming up with the probabilities of a collection of hypotheses, whereas a classical probability theory only deals with one single hypothesis. Finally, the major strength of the D-S theory is its ability to deal with ignorance and missing information [6].

Representation of Evidence

If Θ is the set of $\theta N(\theta\ N\ \epsilon\Theta)$ corresponding to *N* identifiable objects, then

Θ is the space of hypotheses called a frame of discernment.

$\Theta = \{\theta 1, \theta 2, \ldots, \theta N\}$.

These singleton hypotheses are assumed to be mutually exclusive and exhaustive. The

D-S theory allows considering any subset of Θ. For example, 2Θ denotes the power set

ofΘ . A subset that contains at least two elements of Θ is called a compound hypothesis.

A key point of the D-S evidence theory is the mass function *m* (basic probability

assignment, or BPA) that is defined on 2Θ as

$m: 2^{\Theta} \rightarrow [0,1]$.

Here, *m* is defined for every element *A* of 2Θ such that the mass value *m*(*A*)

belongs to the [0, 1] interval with the following property:

$m(\Phi) = 0$,

$$\sum_{A \in 2^{\Theta}} m(A) = 1, \qquad (4)$$

Evidence Combination

The procedure for aggregating multiple evidence from different sources defined onthe same frame of discernment by means of the previously defined mass functions is animportant issue in the D-S theory. This can be seen as a problem of information fusion.Two bodies of evidence *m*1 and *m*2 with focal elements *A*1 ,L, *Ai* and *B*1 ,L, *B j* ,respectively, can be combined to yield a new mass function *m* by a combination rule. The

D-S evidence theory provides a method to compute the orthogonal sum $m = m1 \oplus m2$ oftwo bodies of evidence, according to the Dempster's combination rule [1], by

$$m(\phi) = 0,$$
$$m(A) = \frac{\sum_{A_i \cap B_j = A} m_1(A_i) m_2(B_j)}{1 - K} \quad \forall A \subseteq \Theta, \qquad (5)$$

Where

$$K = \sum_{A_1 \cap B_j = \phi} m_1(A_1) m_2(B_j)$$ and K<1

Conflict Problems

One attractive feature of the Dempster's rule is that the normalization produces convergence toward the dominant opinion. In other words, concordant items of evidence reinforce each other by assigning mass in the null set to the focal elements. For example, consider a situation in which the bodies of evidence have two mass functions as follows:

$m1(\{a\}) = 0.5$, $m1(\{a, b\}) = 0.1$ and $m1(\{c\}) = 0.4$,

$m2(\{a\}) = 0.6$ and $m2(\Theta) = 0.4$,

where$\Theta = \{a, b, c\}$. According to the Equation (5) of the Dempster's rule, thecombination result is:

$m(\{a\}) = 0.74$, $m(\{a, b\}) = 0.05$ and $m(\{c\}) = 0.21$.
Example 1:In the multisensor data fusion system, suppose that the frame of discernment
Θ is {missile *a*, helicopter *h*, fighter *f*}. The sensors of radar and infrared rayprovide two bodies of evidence *m*1 and *m*2, respectively,
$m1(\{a\}) = 0.96$, $m1(\{h\}) = 0.04$, $m1(\{f\}) = 0.00$,
$m2(\{a\}) = 0.00$, $m2(\{h\}) = 0.02$, $m2(\{f\}) = 0.98$.
The combination *m* can be processed in the manner of Table 1. From Table 1, the value of *K* is 0.9992. According to Equation (5), the results *m* of combination can be figured out as follows:
$m(\{a\}) = 0/(1-K) = 0$,
$m(\{h\}) = 0.0008/(1-K) = 1$,
$m(\{f\}) = 0/(1-K) = 0$.
Thus, according to the Dempster's combination results, the radar and infrared ray sensors offer little belief to the helicopter, but the final target identification supports the helicopter 100%. This appears counter-intuitive.

Table 1 Dempster'sCombination[6]

m1 / m2	{a} 0.96	{h} 0.04	{f} 0.00
{a} 0.00	{a} 0.000	ϕ 0.000	ϕ 0.000
{h} 0.02	ϕ 0.0192	{h} 0.008	ϕ 0.000
{f} 0.98	ϕ 0.9408	ϕ 0.0392	{f} 0.000

Combining Evidence Based on Modified Dempster's Rule
Since there are flaws in Dempster's combination rules, the combination result can be unacceptable if the bodies of evidence highly conflict with each other. Therefore, it is of great importance to modify the D-S combination rules and mend the flaws for when the conflicting bodies of evidence are combined in data fusion for target identification. Under the closed-world assumption, we can modify Dempster's combination rules by using a new rule of evidence combination that not only corrects the counter-intuitive effect but also incorporates the result of the uncertainty coming from conflicting bodies of evidence. Therefore, Dempster's combination rule can be modified as:

$$m(\phi) = 0,$$
$$m(C) = \sum_{A_i \cap B_j = C} m_1(A_i) m_2(B_j) + \sum m_\phi(C), \forall C \subseteq \Theta. \quad (6)$$

3.3.2 Linear consensus method

Consensus theory [4],[5] involves general procedures with the goal of combining single probability distributions to summarize estimates from multiple experts, with the assumption that the experts make decisions based on Bayesian decision theory. The combination formula obtained is called a consensus rule. The consensus rules are used in classification by applying a maximum rule, i.e., the summarized estimate is obtained for all the information classes, and the pattern *X* is assigned to the class with the highest summarized estimate. Probably, the most commonly used consensus rule is the linear opinion pool (LOP), which is based on a weighted linear combination of the posterior probabilities from each data source. Another consensus rule, the logarithmic opinion pool (LOGP), is based on the weighted product of the posterior probabilities.
The LOGP differs from the LOP in that it is unimodal and less dispersed. Also, the LOGP treats the data sources independently. The weighting schemes in consensus theory should reflect the goodness of the input data. The simplest approach is to give all the data sources equal weights. Also, reliability measures that rank the data sources according to their goodness can be used as a basis for heuristic weighting Furthermore, the weights can be chosen to not only weight the individual sources but also the individual classes. For such a scheme, both linear and nonlinear optimization can be used.
The consensus theory is a popular method of multi-classifier combination, and it is suitable to integrate multiple outputs of the category probability generated by all member classifiers. Two commonly used models are: linear consensus model and logarithm consensus model. Equation is the principle of linear consensus model:

$$T_j(X) = \sum_{i=1}^{N} p_i(C_j \mid X) * \lambda_{ij} \quad (7)$$

Each classifier is regarded as an expert, and the output element corresponding to X is its membership degree, or confidence level or probability to every class. TjXis the membership of the unlabeled pixel X to class j after combining multiple classifiers. Pi(Cj\X) is the probability or the confidence level of X belonging to class j by the ithclassifier.λ_{ij} is the classification accuracy (producer accuracy) of classifier i to class j and it represents the important degree as the weight.

3.3.3 SVM as the combiner of decision level fusion

Support Vector Machine (SVM) is an implementation of the latest generation of machine learning algorithms based on recent advances in statistical learning. These supervised methods are used for classification and regression.
SVMs discriminate two classes by fitting an optimal linear separating hyperplane (OSH) to the training samples of two classes in a multidimensional feature space. The optimization problem being solved is based on structural risk minimization and aims to maximize the margins between the OSH and the closest training samples the so called support vectors. For linearly not separable cases, the input data are mapped into a high-dimensional space in which the new distribution of the samples enables the fitting of a linear hyperplane[3].

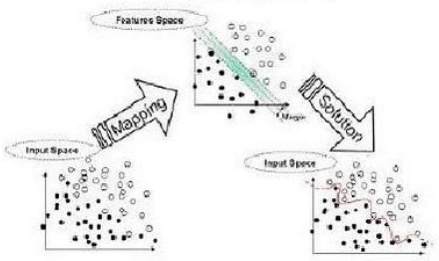

Figure 8 Support Vector Machine

3.3.3.1 Linear SVM

For a two-class problem in a n-dimensional space R^n, we assume that l training samples, xi 2 R^n, are available with their

corresponding labels yi = ±1, S = {(xi, yi) | i 2 [1, l]}. The SVM method consists of finding the hyperplane that maximizes the margin (see Fig. 1), i.e., the distance to the closest training data points in both classes. Noting w 2 R^n as the vector normal to the hyperplane and b ϵR as the bias, the hyperplaneHp is defined as

$$\langle \mathbf{w}, \mathbf{x} \rangle + b = 0, \ \forall \mathbf{x} \in H_p \quad (8)$$

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

$$\min \left[\frac{\|\mathbf{w}\|^2}{2} + C \sum_{i=1}^{l} \xi_i \right] \quad (9)$$

subject to

$$y_i(\langle \mathbf{w}, \mathbf{x}_i \rangle + b) \geq 1 - \xi_i, \ \xi_i \geq 0 \ \forall i \in [1, l] \quad (10)$$

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

$$y_u = sgn\left(\sum_{i=1}^{l} y_i \alpha_i \langle \mathbf{x}_u, \mathbf{x}_i \rangle + b \right) \quad (11)$$

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

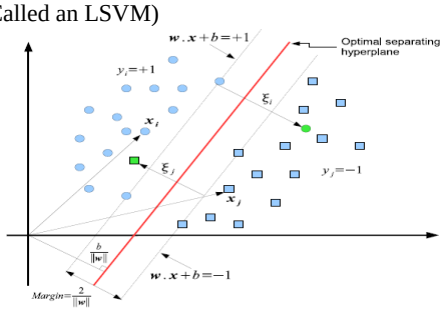

Figure 9 Maximum Margin Of SVM

There is one non separable vector in each class

3.3.3.2Non linear support vector machine

Using the so-called Kernel Trick, one can generalize SVMs to non-linear decision functions. This way, the classification capability is improved. The idea is as follows. Via a non-linear
mapping Φ data are mapped onto a higher dimensional space F

$$\begin{aligned} \Phi : \ \mathbb{R}^n &\longrightarrow F \\ \mathbf{x} &\longmapsto \Phi(\mathbf{x}). \end{aligned}$$

The SVM algorithm can now be simply considered with the following training samples:

$$\Phi(S) = \{(\Phi(\mathbf{x}_i), y_i) \mid i \in [1, l]\}.$$

It leads to a new version of where the scalar product is now: $\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle$.
Hopefully, for some kernels function k, the extra computational cost is reduced to:

$$\langle \Phi(\mathbf{x}_i), \Phi(\mathbf{x}_j) \rangle = k(\mathbf{x}_i, \mathbf{x}_j).$$

Using kernels, it is possible to work implicitly in F while all the computation are done in the input space.

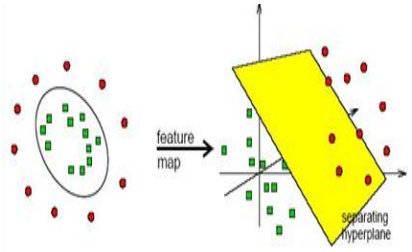

Figure 10 Non Linear SVM 'sSeperation

The kernel function transforms the data into a higher dimensional space to make it possible to perform the separation.
The structure of SVM Combiner is shown as Figure below:

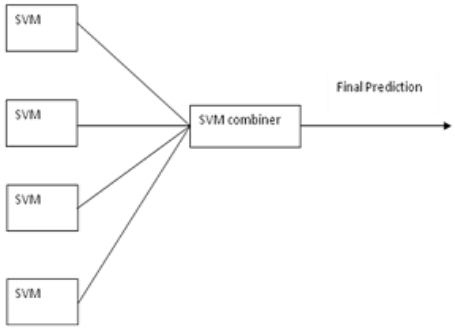

Figure11 SVM Combiner

4. RESULT AND ANALYSIS

The second scheme which includes spectral and texture features and the fourth scheme using MNF transformation and texture feature obtain higher classification accuracy than original remote sensing only. The texture and MNF transformation have positive effects to classification, and MNF transformation may overcome the Hughes phenomena to a certain extent.In this paper,the different bands of image are read from raw data which is obtained from WeoGeo using this URL:
http://market.weogeo.com/#/original_data_map/osucoas_hyperspec_goesr_sep11_e
This hyperspectral dataset was generated by the SAMSON sensor. It covers the spectral range of 400nm-900nm 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 FERI. The following paper describes the basic design of the sensor: Kohler et al. (2006)

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 12 Hyperspectral Image

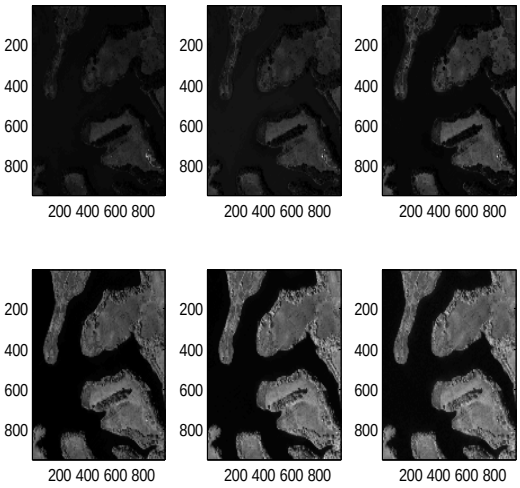

Figure 12 Output at different bands

5. CONLUSION

The advantage of classification is obvious, we can get the physical meaningful reflectance or temperature and their multivariate spreads, we can know the estimate the area coverage for each class, which is important for quantitative analysis.

The decision fusion method combines the advantages of both supervised classification and unsupervised classification.SVM combiner fusion and the improved evidence theory method have
obtained the best accuracy separately, so they are suitable for further uses.Also we can get different band images from raw data which is obtained from WeoGeo.

6.REFERENCES

[1] Peg Shippert" Introduction to Hyperspectral Image Analysis" Earth Science Applications Specialist Research Systems, Inc

[2] Ding, W. and Marchionini, G. 1997 A Study on Video Browsing Strategies. Technical Report. University of Maryland at College Park

[3] B. Waske and J. A. Benediktsson, "Fusion of support vector machines for classification of multisensor data," IEEE Trans. Geosci. Remote Sens., vol. 45, no. 12, pp. 3858–3866, Dec. 2007

[4]J. A. Benediktsson and P. H. Swain, "Consensus theoretic classification methods," IEEE Trans. Syst., Man, Cybern., vol. 22, pp. 688–704, July/Aug. 1992.

[5] J. A. Benediktsson, J. R. Sveinsson, and P. H. Swain, "Hybrid consensus theoretic classification," *IEEE Trans. Geosci. Remote Sensing*, vol. 35,pp. 833–843, July 1997.

[6] Tzu-Chao Lin," Improving D-S Evidence Theory for Data Fusion System,"

[7] 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 Sensing*, vol.39, no.11, pp.2539-2546, Nov 2001.

[8] Peijun Du, Wei Zhang, Shubi Zhang and Junshi Xia,'"Hyperspectral Remote Sensing Image Classification Based OnDecision Level Fusion,"Geoscience and RemoteSensing Symposium,2009 IEEE International,IGARSS 2009

Table 2 Classification Accuracy[8]

Sr.No.	Classification method	The First Scheme		The Second Scheme		The Third Scheme		The Fourth scheme	
		Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
1	Evidence Theory	92.18%	0.90	92.67%	0.91	90.79%	0.88	94.00%	0.92
2	Linear Consensus	92.53%	0.91	92.67%	0.91	89.88%	0.87	92.32%	0.90
3	SVM Fusion	92.67%	0.91	96.86%	0.96	90.65%	0.88	93.44%	0.92