

Land Cover Classification using Support Vector Machine

Dhriti Rudrapal
PHCET, Rasayani,
Raigad, Maharashtra, India

Mansi S. Subhedar
PHCET, Rasayani,
Raigad, Maharashtra, India

Abstract - In modern changing world, it is very important to track the current status of Earth surface for geological and ecological point of view. It can be performed by collecting information about various surface types using high definition satellite sensors and airborne SAR (Synthetic Aperture Radar) sensors. In this aspect, Hyperspectral imaging plays an important role for the classification of different land cover types due to its rich information content. In this paper, we have extracted spatial and spectral features from Samson hyperspectral image and designed a model that can perform classification task. Initially, we performed unsupervised learning which provides a good understanding about the dataset. We prepared training data under supervised learning and testing is done to verify the accuracy of the proposed model. Support Vector Machine (SVM) is employed for automatic classification of various land cover types and overall accuracy is found to be more than 90 % almost in all cases.

Keywords—Hyperspectral data, Remote sensing, Support Vector Machine

I INTRODUCTION

Land cover describes physical properties of Earth surface. The Earth surface is not homogeneous at all, rather contains variations like water, bare ground, trees, grass, asphalt etc. Land use describes how people utilize the land under social and economic activities. The most commonly known land use categories are urban and agricultural utilities. In a generalized meaning, land cover demonstrates all types of variations in the surface of Earth.

Remote sensing is defined by technique of collecting information of Earth surface from a distinct place. The information is sensed and collected in the form of electromagnetic radiations reflected from Earth surface. It is then further processed and analyzed. The reflected radiation is a factor of interaction between incident radiation and surface elements. The various parameters involved with remote sensing are energy source of luminance, radiation through atmosphere, interaction with target elements, receiving energy by sensor, processing, analysis etc. The primary source may be solar energy; secondarily synthetic aperture radar systems can also illuminate radiations towards target [1].

The electromagnetic energy travels through atmosphere and may interact with atmospheric particles. The heterogeneous Earth surface interacts with incident radiations and results in reflection, refraction and absorption of energy in various manners. The reflected part of energy is sensed by high definition sensors in the optical or sometimes infrared or microwave bands of electromagnetic spectrum. The sensed

energy is then recorded and transmitted electronically to a processing station which is stored in the form of imagery data for further analysis [2].

Hyperspectral imagery is one type of data processing under remote sensing, which can be used for classification of land cover types. A multispectral remote sensor produces an image by sensing energy with a few number of broad wavelength electromagnetic spectrum bands. On the other side, Hyperspectral image is formed by sensing electromagnetic energy using hundreds of narrow adjacent bands. The hyperspectral measurement creates an image cell with continuous spectral information. By adjusting the hyperspectral sensor with atmospheric conditions, sensed terrain spectra is compared with laboratory (reference) spectra to correctly recognize and map surface materials like various land cover and land use types.

Imaging spectrometers are used to produce these hyperspectral images. The instruments involve two distinct technologies, namely spectroscopy and remote sensed imaging. Spectroscopy is the study of emitted or reflected light from a material in terms of variation of energy with respect to wavelengths. When this spectroscopy is applied with the optical remote sensing, it deals with the reflected or scattered solar energy from variety of Earth surface materials. By applying several detectors, imaging spectrometers take several measurements in narrow bands like 0.01 micrometers for a spectrum range of typically 0.4 to 2.4 micrometers that is visible to middle infrared wavelengths. A hyperspectral image contains a rich quality of information about the object from which energy is reflected. But interpretation is quite complicated as it needs the information of what surface properties we exactly want to extract and how the hyperspectral sensors have done the measurements [3].

Processing and analysis of remotely sensed hyperspectral data involves the classification task. When we talk about the classification of an image several supervised and unsupervised techniques come into picture. We have selected Support Vector Machine (SVM) as a supervised learning technique for classification of remotely sensed hyperspectral data.

The paper is organized as follows. Section II discusses related work, section III describes proposed system, and section IV presents analysis of results. Section V concludes the work.

II RELATED WORK

Hyperspectral imagery is widely used for environmental monitoring in many of the recent works. A hyperspectral image contains information about spatial, spectral, multitemporal and multisensory domain. Sarath *et al.* discussed about several supervised and unsupervised techniques for hyperspectral image classification like Maximum Likelihood Classification, Minimum Distance Classification and Parallelepiped Classification. It was stated that Maximum Likelihood Classifier may show better accuracy than Parallelepiped Classification, but for faster applications Parallelepiped Classification is preferred over Maximum Likelihood and Minimum Distance Classification [4]. The classification algorithms may be per pixel, sub pixel, per field, contextual and multiple classifiers based. Due to high dimension and multiclass problems of hyperspectral data, classification becomes a challenging task. A Gaussian mixture model is proposed in [5], which adopts nonnegative matrix factorization with locality preserving ability for dimension reduction to overcome high dimension issue. T. Moughal proposed SVM to deal with multiclass problems of hyperspectral image and found overall accuracy of 78.39%. SVM being a binary classifier creates an optimum hyper-plane between one class of interest and other classes [6]. J. Anthony and R. Crompton demonstrated the success of Support Vector Machine for remote sensing hyperspectral image classification. They verified performance with Aviris hyperspectral data and found accuracy of about 90 % [7]. In certain applications like urban area mapping, integration of spatial and spectral properties is mandatory as sufficient spatial resolution is necessary to distinguish small spectral classes [8]. Gustavo *et al.* proposed a semi-supervised graph based method, designed to handle the special characteristics of hyperspectral images, which produces better classification maps [9].

We need to consider some important aspects while hyperspectral image processing. First of all, the proposed system should be able to deal with high dimensional data volume. Secondly, the model needs to perform satisfactorily with limited training data.

III METHODOLOGY

The complete work-flow of land cover classification can be demonstrated using figure 1.

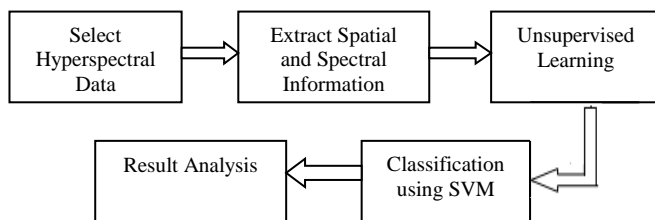


Figure 1: Proposed Method

A Dataset Selection

We have selected Samson hyperspectral image in this work for classification. This data is acquired by Florida Environmental Research Institute under remote sensing using high definition sensor of Samson. This is a push-broom visible to near-infrared hyperspectral sensor.

The data is in units of Remote Sensing Reflectance with a scale factor applied to it. The sensor was coupled with a GPS system to gather its positional information. The dataset is available with atmospheric correction done by TAFKAA, an algorithm applied to reduce the effect of atmospheric disturbances. The Samson dataset is freely available in unclassified format, therefore we have selected this in our work [10].

B Spatial and Spectral Information

The spatial and spectral properties of a hyperspectral image can be explained by figure 2.

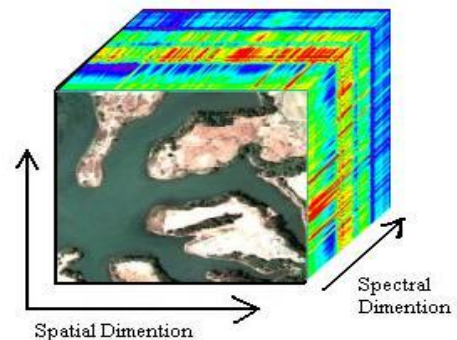


Figure 2: Dimensions of a hyperspectral image

In spatial dimension, SAMSON hyperspectral image consist of 952 lines and 952 samples. That means we can consider an image of 952×952 dimension. Each pixel of this image contains reflection in terms of brightness. In spatial domain pixels are classified based on their relationship with surrounding pixels. But it is a critical task as most of the pixels are not of pure type; rather they are mixed pixels or sub-pixels. The pixels having similar reflections come under a single class and that class represent a particular land type that is actually present in Earth surface.

In spectral dimension, each pixel contains detailed information about spectrum of incoming light. The spectral dimension of selected Samson hyperspectral image is ranging from 400 nm to 900 nm. The range is from visible to near infrared in electromagnetic spectrum. The spectrum is divided in 156 discrete bands having bandwidth of 3.2 nm. Spectral analysis can be done by various methods. Colour composition is one of the simple methods in which a colour image is displayed by combining three of the hyperspectral channels. Figure 3 shows the Samson Hyperspectral image used in our work. The image is displayed using 3 bands out of available 156 different wavelengths: for red, green and blue, selected bands are 70, 50 and 40 accordingly. In hyperspectral image, usually a pixel contains reflections from number of materials, which is referred to as mixed pixel. In spectral analysis, mixed pixels are classified by spectral unmixing process, in which spectrum of a mixed pixel is broken into several pure spectral components. These components are referred as end members and also known as class types or signatures. Different end members are water, vegetation, land etc. The fractional amount of these end members present in a mixed pixel is called as fractional abundance. So, a set of corresponding fractional abundance indicates the compositions of a mixed pixel [11].



Figure 3: Samson Hyperspectral Data

C Unsupervised Learning

An unsupervised system represents a particular input pattern into a statistical structure of input patterns which gives us an idea up to what aspect of structure the input need to capture. Unsupervised learning is sometimes called clustering. Clustering is the technique of finding similarity in the data. It groups similar (near to each other) data types in a single cluster and distinct (far from each other) data types into different clusters. In unsupervised learning, no class values are assigned. Clustering is a data mining technique used almost in every field. The result of clustering depends on algorithm used and area of application. There are two algorithms: partition clustering and hierarchical clustering. The clustering quality can be increased by minimization of intra-cluster distance and maximization of inter-cluster distance [12].

K-means is one of the important and most useful clustering techniques used in unsupervised learning. K-means is a partitioning clustering algorithm. The given data is partitioned in 'k' different clusters, where 'k' is specified by user. Each cluster has a centroid; the cluster center. The algorithm first chooses 'k' data points randomly or initial centroid. Then each data point is assigned to its closest centroid. If the convergence criterion is not met, reassign the data points and continue the process [13].

As an unsupervised learning, we have used K-means clustering in this paper. It helps to acquire initial knowledge about the dataset and helps for preparing the training dataset that is used further in supervised classification. Figure 4 displays K-means output for hyperspectral dataset used in this study, considering 'k' equal to 21 i.e. the number of clusters.

D Classification Process

There are basic two phases of classification: training and testing. Hyperspectral data consists of high information content but available training data is very less.

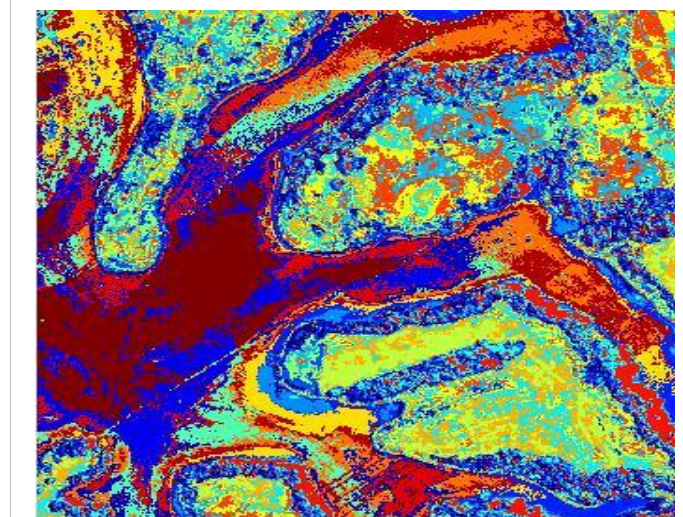



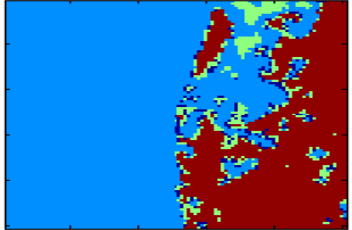

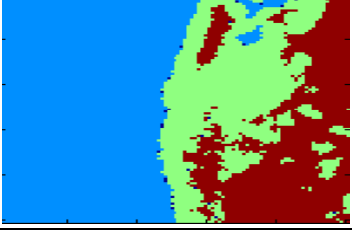

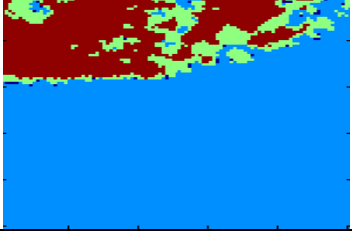
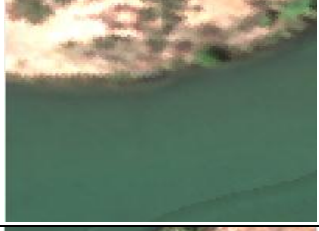
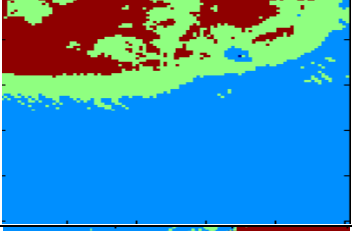

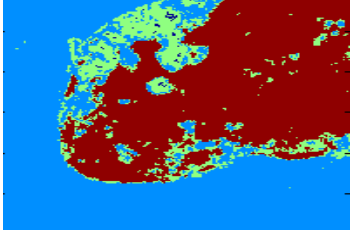

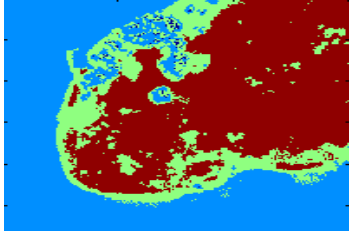




Figure 4: K-Means output (21 Clusters)

SVM shows good results even when the amount of training data is very less. For a given training set, each sample is marked to put into one of the two categories. SVM is a non-probabilistic binary linear classifier. The training algorithm is based on assigning new examples into one of the two categories, making the decision simple and error free. SVM creates a hyper-plane to map samples of different classes with a clear gap that is as wide as possible. SVM uses concept of kernel approach which map data to a high dimensional space through a non-linear transformation. SVM do not need the feature selection in the pre-processing stage as it can work on full dimensionality of hyperspectral data. By designing kernel function, SVM can be applicable to any complex data types [14].

IV RESULT ANALYSIS

We have carried out experiments in Matlab R2010a environment on 2GHz machine with 4GB of RAM. The output of SVM is shown in table I. The test samples are considered by spatial and spectral variation of hyperspectral data. For each test sample, SVM output is shown. The thematic image represents classified output. Each color in thematic map belongs to a theme or class. In this study, four classes are considered namely land, water, vegetation, manmade. For each observation, average accuracy, and overall accuracy is calculated. The average accuracy is calculated by taking mean of accuracies of four individual classes. On the other hand, overall accuracy is found by selecting best result from four classes. The Samson hyperspectral data consists of 156 spectral bands and spatial dimension of 952 lines with each line of 952 samples. From table, it is found that accuracy varies with location change i.e. spatial variation. Accuracy also changes with spectral variations and better results are obtained for higher bands. In most of the cases, we are getting more than 82.32% of average accuracy and 93.23 % of overall accuracy.

TABLE I: CLASSIFICATION OUTPUT WITH SUPPORT VECTOR MACHINE

Observation Number	Test Data		SVM Output (Classified view)	Accuracy	
	Selected scene	Spectral & Spatial position		Average (%)	Overall (%)
1		Band 10 Lines: 300-400 Samples: 300-400		72.2363	97.7086
2		Band 90 Lines: 300-400 Samples: 300-400		92.1543	93.2193
3		Band 20 Lines: 400-500 Samples: 400-500		74.6857	90.2669
4		Band 80 Lines: 400-500 Samples: 400-500		86.9485	86.8998
5		Band 30 Lines: 200-500 Samples: 300-600		79.9162	95.8066
6		Band 80 Lines: 200-500 Samples: 300-600		88.0090	95.4850
<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> Land</div> <div style="text-align: center;"> Water</div> <div style="text-align: center;"> Vegetation</div> <div style="text-align: center;"> Man-made</div> </div>					

V CONCLUSION

Remote sensing is a useful tool for understanding current status of Earth surface as a demand of geologists and sociologists. Hyperspectral data can be used under remote sensing for land cover classification due to rich information contents. The aim of classification is to create a thematic map representing different classes which are actually present at the ground level. Classification is performed in supervised method. Prior to that unsupervised learning using K-means clustering is performed to understand the database. SVM is preferred because as a binary classifier SVM creates a hyper-plane between a class of interest and rest of classes. Also, SVM gives good results when less training data is available while dealing with hyperspectral data. Testing results show good accuracy with SVM. Further we will continue the work with other classification algorithms like Neural Network (NN) under supervised scheme.

REFERENCES

- [1] Antonio Di Gregorio, Louisa J. M. Jansen, "Land Cover Classification System (Lccs): Classification Concepts And User Manual", Environment and Natural Resources Service, Africover, East Africa project and FAO Land and Water Development Division, ISBN 92-5-104216-0
- [2] James B. Campbell, "Introduction to Remote Sensing", The Guilford Press, New York.
- [3] Randall B. Smith, "Introduction to Hyperspectral Imaging", with TNTmips, ©MicroImages Inc., 1999-2012.
- [4] T. Sarath, G. Nagalakshmi, S.Jyothi, "A Study on Hyperspectral Remote Sensing Classifications", International Journal of Computer Applications (0975-8887), International Conference on Information and Communication Technologies (ICICT- 2014).
- [5] Antonio Plaza, Jon Atli Benediktsson, Jon Atli Benediktsson, Joseph W. Boardman, Jason Brazile, Lorenzo Bruzzone, Gustavo Camps-Valls, Jocelyn Chanussot, Mathieu Fauvel, Paolo Gambah, Anthony Gualtieri, Mattia Marconcini, James C. Tilton, Giovanna Trianni, "Recent advances in techniques for hyperspectral image processing", Elsevier Journal, Remote Sensing of Environment 113 (2009) S110-S122.
- [6] T A Moughal, "Hyperspectral image classification using Support Vector Machine", 6th Vacuum and Surface Sciences Conference of Asia and Australia (VASSCAA-6), Conference Series 439 (2013) 012042.
- [7] J. Anthony Gualtieri, Robert F. Crompt, "Support vector machines for hyperspectral remote sensing classification", AIPR Workshop: Advances in Computer-Assisted Recognition, 221 (January 29, 1999); doi:10.1117/12.339824
- [8] Antonio Plaza, Jon Atli Benediktsson, Jon Atli Benediktsson, Joseph W. Boardman, Jason Brazile, Lorenzo Bruzzone, Gustavo Camps-Valls, Jocelyn Chanussot, Mathieu Fauvel, Paolo Gambah, Anthony Gualtieri, Mattia Marconcini, James C. Tilton, Giovanna Trianni, "Recent advances in techniques for hyperspectral image processing", Elsevier Journal, Remote Sensing of Environment 113 (2009) S110-S122.
- [9] Gustavo Camps-Valls, Tatyana V. Bandos, Dengyong Zhou, "Semi-Supervised Graph-Based Hyperspectral Image Classification", IEEE Transactions On Geoscience And Remote Sensing, Vol. Xx, No. Y, Month Z 2007
- [10] <https://www.optics.org/confluence/display/opticks/SampleData-Samson>
- [11] Bolanle Tolulope Abe, "Ensemble Classifiers for Land Cover Mapping", Ph.D., thesis on Philosophy, Faculty of Engineering and the Built Environment, University of the Witwatersrand. Zoubin Ghahramani, "Unsupervised Learning", Advanced Lectures on Machine Learning, 2004, LNAI 3176. c Springer-Verlag.
- [12] Zoubin Ghahramani, "Unsupervised Learning", Advanced Lectures on Machine Learning, 2004, LNAI 3176. c Springer-Verlag
- [13] www.cs.uic.edu/~liub/teach/cs583.../CS583-unsupervised-learning.
- [14] J. Anthony Gualtieri, Robert F. Crompt, "Support vector machines for hyperspectral remote sensing classification", AIPR Workshop: Advances in Computer-Assisted Recognition, 221 (January 29, 1999); doi:10.1117/12.339824