ON THE PERFORMANCE OF SUPERVISED CLASSIFIERS FOR CROP IDENTIFICATION AND ESTIMATION USING MULTI-SPECTRAL IMAGERY

Asif Ali Khan*, Nasru Minallah*, Shahbaz Khan**

ABSTRACT

The objective of this research is to investigate crop estimation using SPOT-5 satellite imagery. We specifically considered tobacco as our pilot crop and compared the obtained results with manually delineated calculations. For this research SPOT-5 imagery of 2.5m spatial resolution, was provided by Space and Upper Atmosphere Research Commission (SUPARCO), space agency of Pakistan. After preprocessing, which is a preparatory step in analyzing and classifying satellite imagery to improve classification results and reduce the efforts and processing time, different supervised classifiers namely Maximum Likelihood approach, Neural Network and Minimum Distance Classifier have been used to classify the imagery. Training data for classifiers has been collected through multiple field surveys using GPS receivers. The results obtained clearly show that the performance of maximum-likelihood classifier is better than the other considered counterparts. Also it is indicated that the newly developed system offer an efficient, reliable and faster approach for estimation of tobacco crop.

KEY WORDS: *Crop Estimation, SPOT-5 imagery, Maximum Likelihood classifier, Neural Network, Region of Interest (ROI), Minimum Distance Classifier*

INTRODUCTION

Pakistan is the 7th largest tobacco growing country in the world[1]. Pakistan Tobacco Board (PTB), a government organization is regulating tobacco crop growing in Pakistan. The PTB play an important role to measure and estimate the total yield of tobacco. This yield measurement and estimation of tobacco crop is very important for government agencies and tobacco regulatory authorities to strictly monitor the tobacco production and to keep its production within allowed limits. Furthermore it also helps in tracking proper tax collection on the basis of tobacco crop yield.

Conventional strategies based on manual measurements are currently used within Pakistan to estimate the overall tobacco yield which is not only time consuming but also prone to human errors. Furthermore, controlled monitoring of tobacco crop within remote villages of Pakistan is only possible through remote sensing because uncooperative attitude of the rural community to limit tobacco production and to deposit Tax returns on Tobacco yield is a major limitation in precise yield data collection using manual measurements.

In this paper, we investigate the application of remote sensing for crop identification and estimation using different approaches such as Maximum Likelihood classification, Neural Network and Minimum Distance Classifier.

Since the development of remote sensing systems, its use is highly considered by both the research and development community for agricultural purposes. Data derived from these systems has been used for crop mapping, crop type identification and crop area estimation. Crop productivity i.e. its type and spatial area coverage is the most important factor for accurate crop yield estimation.

General classification and image processing techniques are used for mapping and identification of crop areas from multi temporal satellite imagery. Different segmentation algorithms are applied on the imagery for object based classification. Several classification techniques such as maximum likelihood approach, Support Vector Machine (SVM), minimum distance and nearest neighbor classifiers etc. are used for identifying different objects.

Satellite imagery has been used for different purposes in agriculture. The urban vegetation has been calculated using different satellite images[2]. The objective of the research was to investigate and calculate the natural resources and vegetation in settled areas using different satellite images. Johnsan et al[3], used Spot-5 imagery to map banana plantation using object-oriented approach. Harris[4] used Landsat imagery for identifying agricultural

*Department of Computer Systems Engineering University of Engineering & Technology Peshawar, Pakistan
** Department of Telecommunication Engineering. University of Engineering & Technology Peshawar, Pakistan

changes in two selected areas of Oman using both supervised and unsupervised classification methods. The end results of both classification methods were satisfactory.

Two different dated SPOT-5 images have been used for crop area identification[5]. The area of interest in the research was Sanliurfa city, a south-eastern region of Turkey. Two different classification methods namely object-based classification and pixel-based classification were used to identify and estimate crop type and area and the results were compared using different methods e.g. Kappa coefficient (statistical measure used for accuracy assessment) and confusion matrix which is a tabular layout allowing visualization of performance of algorithm.

Spot-4 satellite imagery of 20m multi spectral resolution and 10m panchromatic resolution has been used to identify and estimate the forest fire damage[6]. The core of the research was to assess the damage caused by fire on 10th August, 2009 in the Izmir district of Turkey. Orthorectification, level slicing and classification like concepts of image processing have been applied in the research.

The application spectrum of remote sensing is very wide e.g. in Russia, remote sensing is used to assess carbon emission due to forest fire[7]. A Normalized Difference Vegetation Index (NDVI) image of pre and post fire satellite imagery has been produced and used for assessment. Similarly, some researchers have extensively used this for crop yield estimation. Others used aerial photograph for efficient crop management. Many others such as[8-15] have used satellite imagery for detecting and monitoring land cover and change detection, analysis of vegetation trends, classifying urban regions, classification of cultivated lands etc.

DATASET USED AND STUDY AREA

Study Area

In Pakistan, the province of Khyber Pakhtunkhwa produce 77% of the total tobacco crop. The districts that we have targeted for this research are Mardan and Swabi, pilot regions of Khyber Pakhtunkhwa. Both these districts contribute significantly to the total tobacco yield in the province.

Imagery Description

There are various types of imaging systems providing different types of sensing data e.g. Panchromatic imaging system provides grayscale images, multispectral imaging system captures information in multiple frequency bands and so on. SPOT (satellite for earth observation) is an imaging system operated by French space agency, Centre National d'EtudesSpatiales (CNES) which has both panchromatic and multispectral sensors. The imagery used in our project is multi band SPOT-5 imagery provided by Space and Upper Atmosphere Research Commission (SUPARCO), space agency of Pakistan. The acquisition date as reported by SUPARCO is 28th July, 2013 (a shining sunny day in the region). This particular period of the year is selected for imagery acquisition because most of the tobacco is in mature phase during this period. Figure 1 shows the actual imagery provided by SUPARCO.

The multi band SPOT-5 imagery consists of 4 different bands located in Green, Red, Near Infrared (NIR) and mid infrared regions of the spectrum at 2.5 m pixel resolution.

Due to summer period, mostly the imagery is cloud-free. The area of individual scene is 60 x 60 Km with 2.5 meter pixel resolution.

APPROACH

In almost every classification problem, the general steps are as follows.

Image/Data Acquisition

The details of the imagery used in the research has already been presented in the previous section.

Preprocessing

Preprocessing is a preparatory step in analyzing and classifying satellite imagery which not only improves the classification results but also reduces the efforts and processing time of overall process. Every single operation that takes place on the imagery before the actual classification process is considered as preprocessing. Preprocessing steps in our case are radiometric calibration, Image stretching, geometric rectification and atmospheric correction.

Radiometric influence results in a distorted sensor signal which has vegetation features information. Clouds, dust and other atmospheric factor also temper the actual signal which results in erratic classification results later in classification process. ENVI 5.0 has number of built-in functions to remove all these effects. FLASH and QUICK atmospheric corrections functions have been applied to the imagery before feeding it to the classifier.

Geometric correction also needs to be applied before image analysis. For image-to-map geometric rectification, the ENVI 'Image-to-map registration' function has been applied.

Classification

After preprocessing the imagery, the next step is to classify the data. There are two main types of classification.

Supervised Classification

In supervised classification, the classifiers are first provided with the training samples. The test data is then provided to the trained classifiers for classification purposes.

Unsupervised Classification

Unlike supervised classification, no trainings is required in this type of classification. All you need to do is to provide your data to the classifier. The unsupervised classifiers analyze the data and find some patterns in it. It then classifies the data in different classes based on the similar patterns found.

In this paper, we have focused on supervised classification and used 3 different classifiers namely Neural-network, minimum distance and maximum likelihood classifier.

Tool Used

There are many available tools that can be used for multi spectral imagery analysis and classification purposes. In our research, we have used ENVI 5.0.

ENVI, from ITT Visual Information Solutions is a visualization software platform for viewing and analysis of hyper spectral imagery. ENVI includes a number of mineral and vegetation spectral libraries which can be used for spectrum identification, feature extraction, anomaly detection, target finding and material mapping. The spectral Library Builder within ENVI allows for your specific field spectral reflectance data files to be added to these data bases and re-sampled to match the hyper spectral or multi-bands of the airborne or satellite images.

Results and discussions

The result of classifying the study area for tobacco estimation using SPOT-5 imagery is shown in Figure 2. This show that the aerial photograph gives valuable information and can be used for agricultural purposes specifically for crop identification and area estimation. Training data collection and accuracy assessment procedure is discussed before diving into the classifiers performance and results.

Collection of Training Data

As discussed earlier, supervised classification needs training samples in the first step. For gathering training sample, two surveys/field visits were conducted to different parts of district Mardan in order to collect training samples. Coordinates of around 60 tobacco fields were recorded using GPS receivers. Different Region of Interest (ROI) were then marked on the imagery using these coordinates. All supervised classifiers consider these ROIs as their training data and classify the imagery to multiple classes accordingly. The number of classes in the classification result is equal to the number of ROIs in the ROI file.

Accuracy Assessment Procedure

Various assessment procedures are present to assess the accuracy of different remote sensing techniques. In this research, we are using the ground-truth data collected in field surveys using GPS receivers to assess and compare the performance of the considered classifiers.

The 3 classification methods were evaluated based on

Table 1. Performance Comparison of different Classifiers

Class	Tobacco (%age)	Other Vegetation (%age)	Settled Area (%age)	Water (%age)	Hills (%age)	Overall Accuracy (%age)	Kappa Coefficient
Minimum Distance	**69.20**	**22.43**	**76.69**	**84**	**70.18**	**79.25**	**0.7490**
Maximum Likelihood	100	91.63	93.24	100	99.31	96.84	0.9657
Neural Net	96.53	76.43	80.41	100	96.10	92.32	0.9070

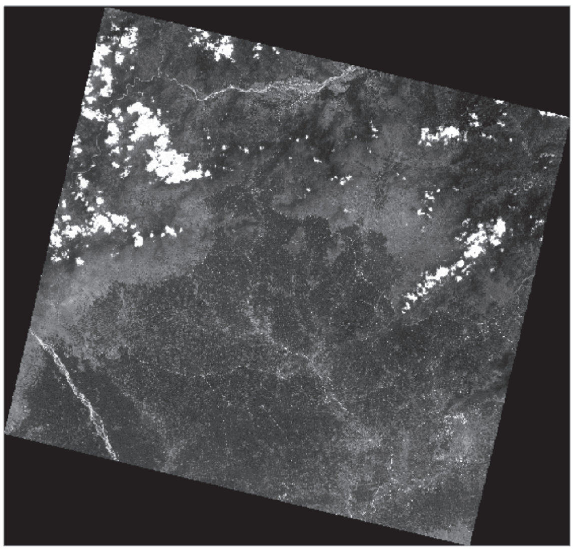

Figure 1. Spot 5 imagery of 2.5m spatial resolution

Table 2. Class Distribution Summary

Class Name	Area (Hectares)
Tobacco	**16,885.3288**
Other Vegetation	88,673.5237
Settled Area	56,784.4737
Hilly Area	47,931.9212
Water	580.8250
Unclassified	4,792.6781

accuracy assessment procedure and ground truth verification. The imagery has been classified in 5 different classes including tobacco crop, other vegetation, settled area, water and hills. The classification result and its accuracy is summarized in Table 1.

The accuracy of individual class as well as the

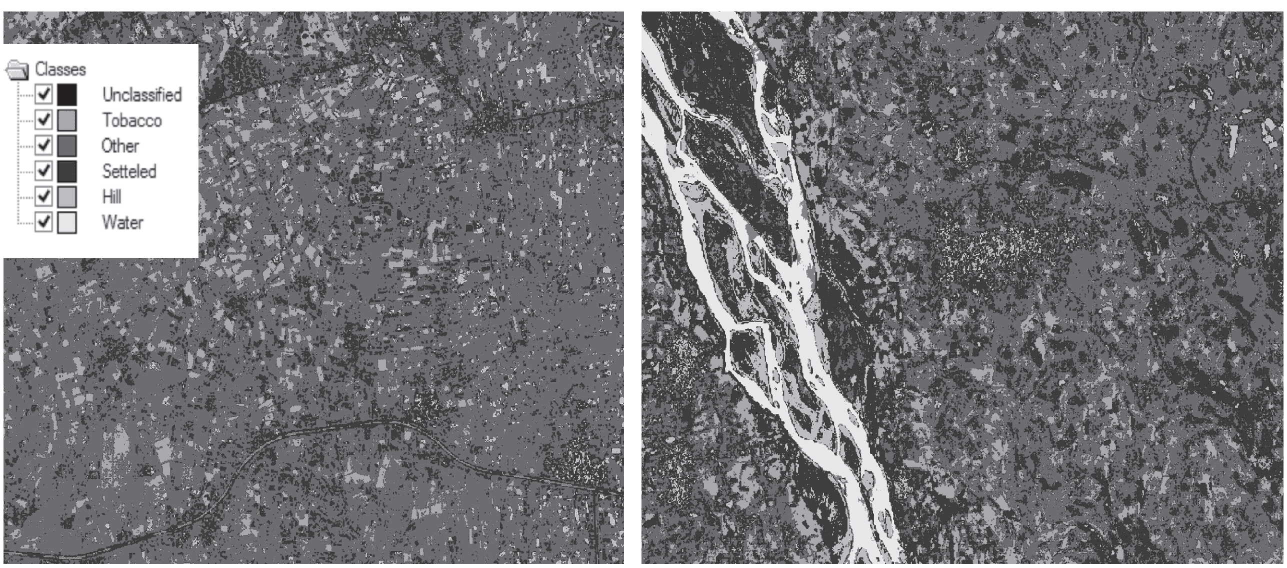

Figure 2. CLassification Result using Maximum Likelihood Classifier

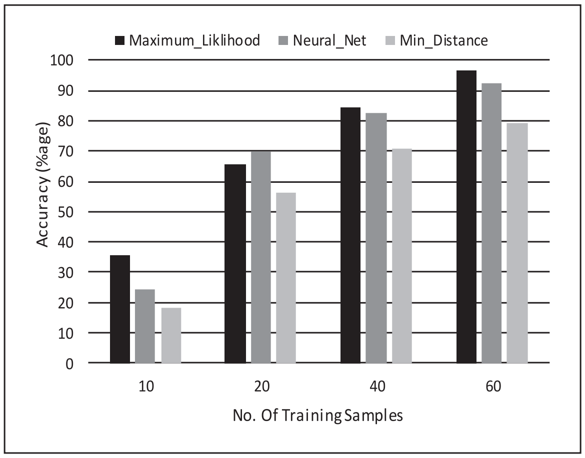

Figure 3. Classifiers Performance vs Training Samples

overall accuracy and Kappa coefficient of the classifiers is presented. The overall accuracy also considers the under-estimation and over-estimation problem. Under-estimation means if a Tabaco crop has not been classified as tobacco and over-estimation is vice versa i.e. if the field is not actually tobacco but it has been classified to tobacco by the classifier.

Both class level and overall accuracy from Table 1 shows that maximum-likelihood gives better performance than the other two considered classifiers. The class distribution summary of the overall imagery as per maximum-likelihood classifier is presented in Table 2.

The relationship of classifiers accuracy and number of training sample is presented in Figure 3. The performance of the classifiers is directly dependent on the number of training samples. With lesser training data, the accuracy of the classifiers degrades whereas with more training examples it boosts as shown in Figure 3.

Beyond 60, increasing the number of training samples does not affect the accuracy rather it require more computational power which is not desirable.

REFERENCES

1. *British American Tobacco Group. "History of tobacco." Pakistan Tobacco Company. Published: 07/05/2010 16:49:15 GMT. Last Accessed on 02 February, 2014. < http://www.ptc.com.pk/group/sites/PAK_7SHBXN.nsf/vwPagesWebLive/DO7T5LR5?opendocument&SKN=1>*

2. *Mathieu, R.; Aryal, J.; Chong, A.K 2007. Object-Based Classification of Ikonos Imagery for Mapping Large-Scale Vegetation Communities in Urban Areas, vol. 7, pp. 2860-2880.*

3. *Johansen, K., Phinn, S., Witte, C., Philip, S. and Newton, L., 2009. Mapping banana plantations from object-oriented classification of SPOT-5 imagery, vol. 75, pp. 1069–1081.*

4. *Harris, R., 2003. Remote sensing of agriculture change in Oman International Journal of Remote Sensing, vol. 24 (23), pp. 4835–4852.*

5. *Ormeci, C., Alganci, U., and Sertel, E., 2010 Identification of Crop Areas Using SPOT – 5 Data, FIG Congress 2010 Facing the Challenges – Building the Capacity, Sydney, Australia, 11-16 (Online Proceeding)*

6. *U. ALGANCI, E. Sertel, C. Ormeci, 2009. Forest Fire Damage Estimation Using Remote Sensing and GIS", 30th EARSeL syposium.*

7. *A.S. Isaev, G.N. Korovin, S.A. Bartalev, D.V. Ershov, A. Janetos, E.S. Kasischke, 2002. Using remote sensing to assess Russian forest fire carbon emissions, Climatic Change, vol. 55, pp. 235–249.*

8. *K. Green, D. Kempka, and L. Lackey, 1994. Using remote sensing to detect and monitor land-cover and land-use change", Photogrammetric E ngineering and Remote Sensing, pages.331-337.*

9. *Hountondji, Y. C., Sokpon, N., and Ozer, P 2006. Analysis of the vegetation trends using low resolution remote sensing data in Burkina Faso (1982-1999) for the monitoring of desertification, International Journal of Remote Sensing 27, pp. 871-884.*

10. *Y. V. Venkatesh and S. Kumar Raja, 2003. On the classification of multispectral satellite images using the multilayer perceptron, Pattern Recognit., vol. 36, no. 9, pp. 2161–2175.*

11. *A. K. Shackelford and C. H. Davis, 2003. A combined fuzzy pixel-based and object-based approach for classification of high-resolution multispectral*

data over urban areas,"IEEE Trans. Geosci. Remote Sens., pp. 2354–2363.

12. *Pierce, L. E., Bergen, K. M., Dobson, M. C., & Ulaby, F. T, "Multitemporal land cover classification using SIR-C/X-SAR imagery", Remote Sensing of Environment, vol. 64, pp. 20–33.*

13. *Biro, K., Pradhan, B., & Buchroithner, M 2009. Delineation of cultivated land use area using multi-sensor satellite data: a case study from Gadarif Region Sudan", 3rd Workshop of the EARSeL Special Interest Group on Land Use and Land Cover, 25–27, Bonn, Germany.*

14. *Lunetta. E and Elvidge. C 1999. Remote Sensing Change Detection: Environmental Monitoring Methods and Applications. London, U.K.*

15. *A. Baraldi, V. Puzzolo, P. Blonda, L. Bruzzone, and C. Tarantino 2006. Automatic spectral rule-based preliminary mapping of calibrated Landsat TM and ETM+images," IEEE Trans. Geosci. Remote Sens., vol. 44, pp. 2563–2586.*