A Supervised Learning Data Mining Approach for Object Recognition and Classification in High Resolution Satellite Data

Mais Nijim, Rama Devi Chennuboyina, Waseem Al Aqqad

Abstract**—**Advances in spatial and spectral resolution of satellite images have led to tremendous growth in large image databases. The data we acquire through satellites, radars, and sensors consists of important geographical information that can be used for remote sensing applications such as region planning, disaster management. Spatial data classification and object recognition are important tasks for many applications. However, classifying objects and identifying them manually from images is a difficult task. Object recognition is often considered as a classification problem, this task can be performed using machine-learning techniques. Despite of many machine-learning algorithms, the classification is done using supervised classifiers such as Support Vector Machines (SVM) as the area of interest is known. We proposed a classification method, which considers neighboring pixels in a region for feature extraction and it evaluates classifications precisely according to neighboring classes for semantic interpretation of region of interest (ROI). A dataset has been created for training and testing purpose; we generated the attributes by considering pixel intensity values and mean values of reflectance. We demonstrated the benefits of using knowledge discovery and data-mining techniques, which can be on image data for accurate information extraction and classification from high spatial resolution remote sensing imagery.

Keywords**—**Remote sensing, object recognition, classification, data mining, waterbody identification, feature extraction.

I. INTRODUCTION

SATELLITE imagery acts as an important resource for various remote-sensing applications. Identifying and classifying the objects in images is the main research problem in the field of remote sensing. Feature extraction and classification can be done using image processing and patterns are recognized through data mining techniques [1], which can provide relevant knowledge for object recognition in satellite data.

Remote sensing technology gives the advanced technical means for recognizing the spatial distribution of land cover, water and other objects like vegetation; each image consist of valuable information and objects inside it, extracting useful patterns from it manually is very complicated. The most frequently used method to generate such objects is image segmentation [2]; this process will group the neighboring pixels of an image into similar regions through the subdivision with determined criteria of similar characteristics. In a segmentation process, each object created has spectral features, texture, morphological attributes and contextual features, which can be used to analyze image. The objects in an image are further assigned to a particular class by comparing different objects with the defined patterns.

The object classification is performed by considering the objects features uniformly this is called object-oriented classification. Using data mining techniques can do automatic recognition of objects and classification in an image. First, for the water body recognition the distribution of water should be extracted [3], for solving this problem using remote sensing satellite information various methods are proposed. The aim of this thesis is to demonstrate an effective classification for Landsat ETM+ imagery using Support Vector Machines.

II. RELATED WORK

Extracting water body information using remote sensing and data mining has gained importance for water survey, protection of wetlands, disaster prevention, monitoring water resources etc. There are several methods for collecting water body information like threshold method, exponential method, decision tree and other spectral methods. For gathering water body data single and multi-band methods are used to obtain the reflectance of surface body, which is useful in extracting vital information for image preprocessing [3]. This information is helpful in calculating ratio and spectral relationships. NDWI commonly termed as Normalized Difference Water Index is used for Ratio based index.

Knowledge based on spectrums plays a crucial role in extracting image information from remote objects. Reflection of electromagnetic waves by the objects on land surface and their reflection is utilized by satellites for collecting information about remote objects [4]. Water bodies have low reflectance value unlike other objects surfaces. Its reflectance value is 4% to 5% and declines to 2% to 3% at 580nm as water bodies its reflectance is about 4% to 5%, while decrease to 2% to 3% at 580nm as water bodies absorb at a bandwidth of 740 to 2500mm which is equivalent to band 4 and band 7. This particular bandwidth is very useful in determining the vegetation, soil, buildings and other surface objects.

Commonly in image extraction, few land data sets are collected from many satellite images as a source of data which will assist in extracting key information for extracting important information from water body images like threshold values, spectral ratios, indexes like NDWI, NWI. All this

M. N. is with the Electrical Engineering and Computer Science Department Texas A&M University-Kingsville (e-mail: mais.nijim@tamuk.edu).
W. A. A. is with the Prince Sultan University Communication and Network Engineering (e-mail: walaqqad@psu.edu.sa).

information is then classified using supervised or unsupervised learning.

Many natural disasters are caused due to water bodies, therefore extracting water body information has gained importance over time and became a curtail research area in Remote Sensing. As a result, accuracy in extracting image information has become more crucial part of remote sensing.

To address these problems in image extraction a lot of research has been done and the frequently used techniques include analyzing spectrums, calculating maximum likelihood for image objects, calculating threshold ratios and spectral ratios and defining training dataset from these information for classifying the objects. However, extraction of image information from water and surface objects has been highly effective with the use of remote sensing technologies. These technologies provide solutions to the drawbacks discussed above such as cost and time. This application of satellite data for the extraction of water source has gained importance for the past 20 years.

III. Area of Interest

The study area is located in New Orleans, LA, USA, specifically located between latitude 29.9511 and longitude 90.0715. It mainly includes complex urban areas of New Orleans, south of Lake Pontchartrain, banks of Mississippi River and Lake Borgne. New Orleans city is located on the banks of Mississippi River, 105 miles from Gulf of Mexico [5].

The area exhibits high-level water and land heterogeneity, which consists of different land, covered areas, and water sources. Common water sources are fresh water river areas and seawater areas of the Gulf of Mexico, which share similar spatial and spectral reflectance values.

This paper Landsat 7 Ortho-rectified ETM+ pan sharpened image acquired on September 17, 2000 from the U.S. Geological Survey (USGS) which provides complete coverage of research area. The USGS (EROS) Center has all data archives of satellite Imagery, from the beginning. All Landsat satellite data in USGS EROS archive can be downloaded of free as cost. The Landsat ETM+ sensor gives eight-bands of multispectral scanning, which has the efficiency of generating high-resolution imagery information of the entire surface of earth [5].

IV. Proposed Methodology

The approach proposed includes the following major stages: image segmentation, training set generation, data mining, and classification of image data using support vector machines, and validating the classification on independent samples. The sequence of stages needed to classify objects is shown in Fig. 2 depicts how image analysis and Data mining are integrated and applied to Landsat ETM+ images.

A. Image Preprocessing

Image preprocessing is the initial step for the classification process and this can be done using various methods of spatial enhancement techniques and spectral transformations. By transformation of pixel intensity and saturation values the spectral data can be preprocessed for better results. Principal component analysis (PCA) and morphological operations were done on the original raw data [6].

At the time image is captured, there may be many problems such as low contrast, less background light, no proper focus of image objects etc. To overcome such drawbacks Contrast Stretching is utilized as it helps in identifying low contrast regions in the captured images.

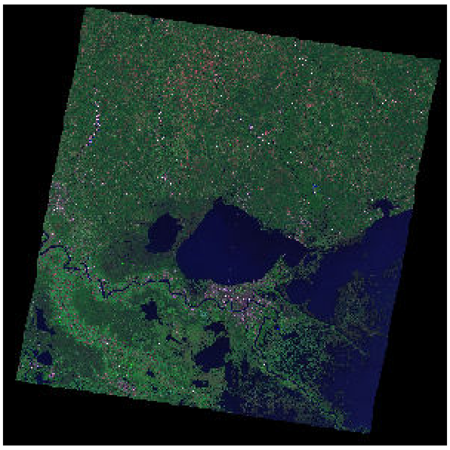

Fig. 1 Area of interest, color composite ETM+ image of band 3(Red), band 2(Green), band 1(Blue)

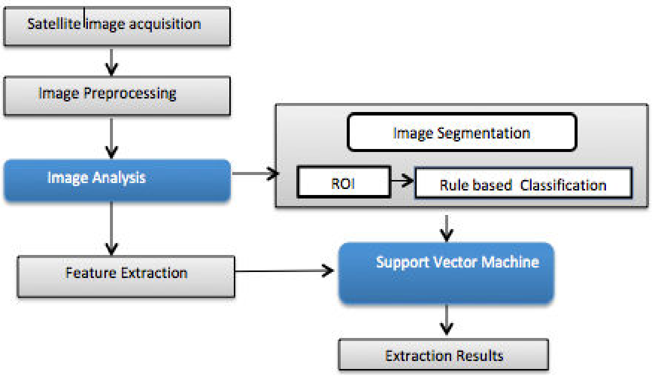

Fig. 2 Flow Chart illustrating Image analysis and Data mining approach proposed to classify land and water from Landsat ETM+ image

B. Data Mining

Data mining, which can be defined as Knowledge Discovery in Database (KKD), can be used for the extraction of useful information from a huge collection of data [7]. By using data mining techniques we can recognize rules, which describe properties of data, objects clustered together and frequent patterns.

Analysis on data and its interpretation requires new and advance techniques like Data mining, which helps in extracting desired and useful information from a given dataset. New areas in Data mining have evolved to enhance knowledge discovery using known or unknown information called Knowledge Discovery Databases (KDD). Data Mining is often considered as identification of patterns from the given data. In

KDD, additional steps are involved such as data preparation, selection of required datasets, data cleaning and utilizing already now knowledge. Unknowingly using data mining techniques can be harmful and derives unknown and meaningless patterns, which are irrelevant to user [7].

C. Support Vector Machine

Support vector machines concept was originally developed by Vapnik for object classification; it has gained great importance later on [8]. Kernel based techniques such as Simple vector machines, navies bayes classification and Gaussian classification etc proved to be good classification techniques. Among all these classification techniques, SVM has gained lot of importance as it can be easily implemented over the rest of classification techniques and it can also be used to calculate density and regression function. SVM is a subset of supervised learning algorithms, which is useful in predicting and classifying objects. In the SVM algorithm, the dataset to be classified is organized linearly or non-linearly. SVM classification techniques can be used for classifying text, recognizing patterns in text data, detecting relevance in data sets. SVM is a very useful technique both in the field of Data Mining as it has the ability and knowledge of data to assign labels to a set of classified data and segregating them from rest of the data. It is also a highly useful technique in Machine learning [8].

Support vector machines are binary classifiers [9], which linearly separate two classes. For an example in the area of remote sensing, the image objects can be classified by considering the pixel value, individual pixels in multi spectral image. Each pixel in the image is considered as pattern vector and the measurements are done numerically. Feature vector elements are the variable numerical values of the pixels in the satellite image with multi spectral bands. Fig. 3 illustrates a examples of a two-class classification problem. Support vector machines needed training variables that describe the specifications of the maximum margin hyperplane. The vectors line on the maximum margin is the support vector that differentiates the two classes. Kernel values and the mathematical calculations are the two important functions of support vector machines to function efficiently.

In this scenario the classification is demonstrated between two-classes, the classes are being P, N where yi =+1,-1.The support vector machine algorithm creates the maximum margin for classification, which can separate two given classes with maximum separation surface [9].

High resolution Landsat 7 ETM+ imagery is used to evaluate and understand the spatial and temporal reflectance values. In our experiments, we consider only two types of the classes: Land and water.

Waterbody has considerably low reflectance in the visible, but it rapidly decreases to zero as wavelength increases into infrared. Compared with this, the reflectance of blue is higher than green, and green is higher than red, so, the clear water will almost always display blue or green.

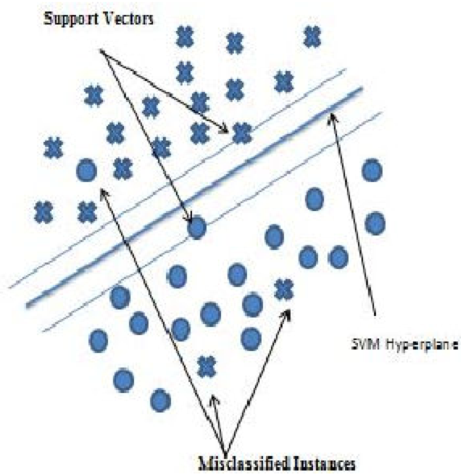

Fig. 3 Example of SVM [9]

V. Results

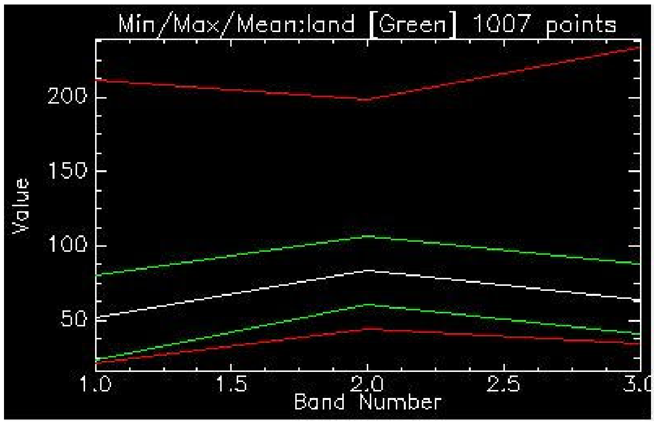

Fig. 4 Mean Values for Land by bands

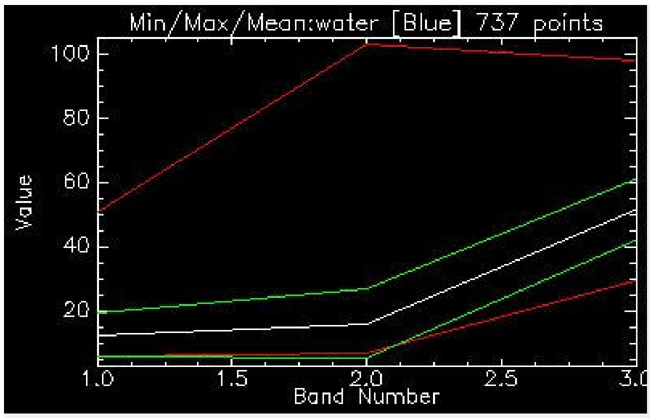

Fig. 5 Mean Values for Water by bands

When the water is clear, it will reflect green and blue colors. On the other hand, when the water is deep it will display the color black. Furthermore, waterbody has almost no reflectance in NIR in color infrared model, so it will shows black. For Salty water with many sediment or nutritious, it will appears to be in green in NC, but blue in CIR depends on the materials it has. CIR is better when compared to than NC model to display waterbodies.

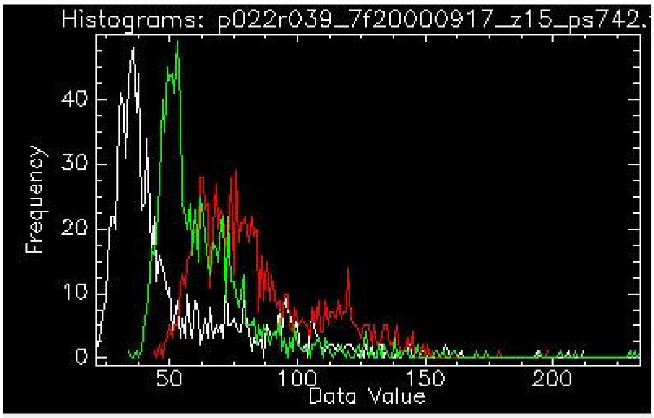

Fig. 6 Histogram for Land (R, G and B) bands

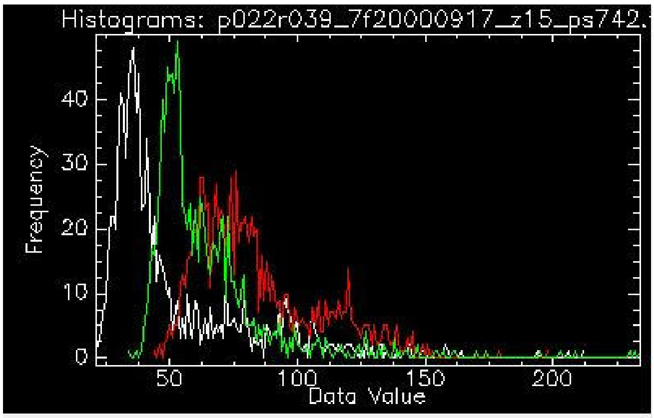

Fig. 7 Histogram for water (R,G and B) bands

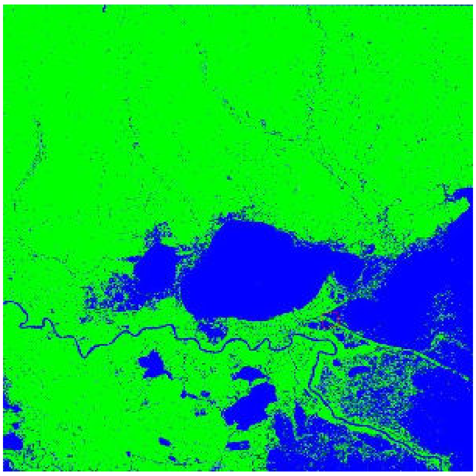

Fig. 8 SVM Classified Image from Landsat ETM+

Water bodies exhibit special features in different situation and color models. Deep water and clear water absorbs almost all of the NIR energy while reflecting somewhat more green and red light, if the water is deep and free of sediment, it will appears to be black. If there is substantial suspended sediment, it will appear in relatively dark shade of blue and green.

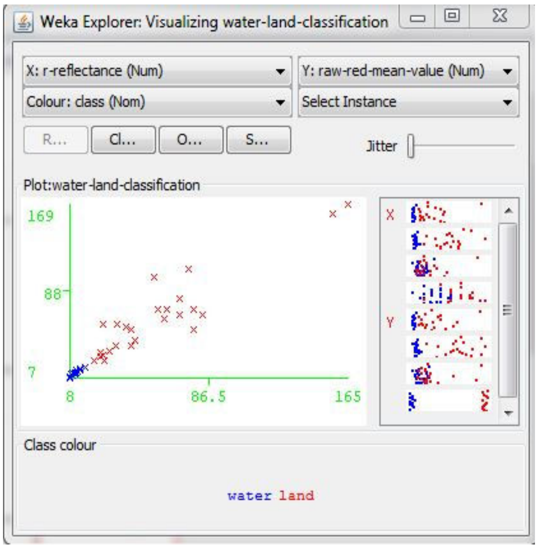

Fig. 9 Plot Red-reflectance

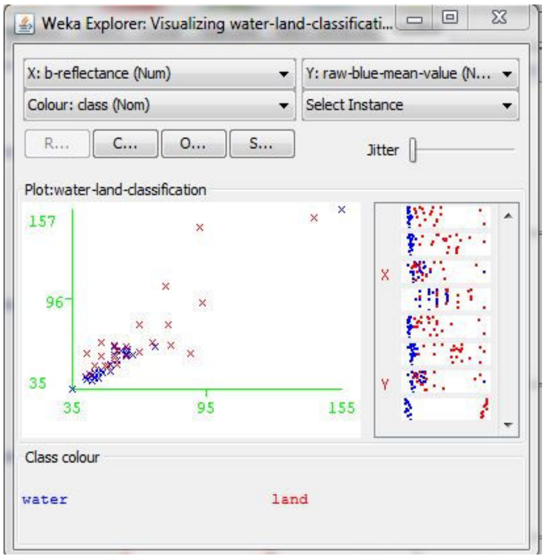

Fig. 10 Plot Green Reflectance

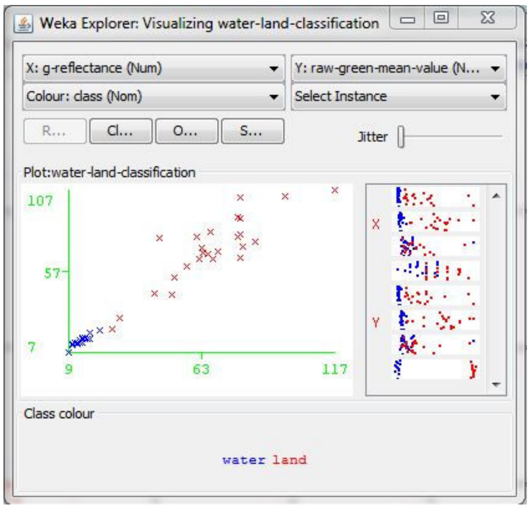

Fig. 11 Plot Blue Reflectance

VI. Conclusion

In this paper, we proposed a classification system for remote sensing Landsat 7 ETM+ satellite imagery using supervised classifiers. The data mining approach for knowledge discovery is considered as a promising approach to find information in the high volumes of satellite data. It employed an object based algorithm for features extraction, and supervised classifier for pattern recognition. For efficient classification and outlier detection support vector machines are being used in the felid of data mining. In our research it is shown that the various types of water body objects in the study area can be extracted from Landsat 7 ETM+ images efficiently using Support vector machines with different nonlinear kernel. In the proposed approach, statistical learning method is employed by the classifier to generate the hypothesis with the available data.

References

[1] A Adnan A. Y. Mustafa, Linda G. Shapiro and Mark A. Ganter,"3D Object Recognition from Color Intensity Images", 13th Int. Conf. on Pattern Recognition, Vienna, Austria, pp. 25-30, /August, 1996.

[2] Wang Xiang-yang, Sun Wei-wei a, Wu Zhi-fang, Yang Hong-ying, Wang Qin-yan “Color image segmentation using PDTDFB domain hidden Markov tree model” Applied Soft Computing 29 (2015) 138–152.

[3] Yang Haibo, Wang Zongmin, Zhao Hongling, Guo Yu “Water body Extraction Methods Study Based on RS and GIS ” 2011 3rd International Conference on Environmental Science and Information Application Technology (ESIAT 2011).

[4] Yuqiang Wang, Renzong Ruan, Yuanjian SHE, Meichun YAN “Extraction of Water Information based on Radarsat Sar and Landsat ETM+ ” 2011 3rd International Conference on Environmental Science and Information Application Technology (ESIAT 2011).

[5] U. S. Geological Survey (USGS)” Landsat Orthorectified ETM+ Pan Sharpened” Sioux Falls, SD USA, USGS Earth Resources Observation and Science Center (EROS), https://lta.cr.usgs.gov/Tri_Dec_GLOO.

[6] Xiaoxiao Lia, B., Soe W. Myintb, Yujia Zhangb, Chritopher Gallettib, Xiaoxiang Zhangc,Billie L. Turner II “Object-based land-cover classification for metropolitan Phoenix, Arizona, using aerial photography” International Journal of Applied Earth Observation and Geoinformation 33 (2014) 321–330.

[7] Usama Fayyad, Gregory Piatetsky-Shapiro, and Padhraic Smyth “From Data Mining to Knowledge Discovery in Databases “American Association for Artificial Intelligence. All rights reserved. 0738-4602-1996.

[8] I. Witten, E. Frank “Data Mining: practical Machine Learning Tools and Techniques”.

[9] J. Zahang, W. Hsu and M. L. Lee. “An information-Driven Framework for image Mining”, in proceedings of 12th International Conference on Database and Expert Systems Applications.