

# Feature Extraction from Satellite Imagery

Aishwarya.C.G<sup>1</sup>, Apoorva T M<sup>1</sup>, Shresta R Shetty<sup>1</sup>, Vandana B R<sup>1</sup>, Mrs. Sangeetha V<sup>2</sup>

Student of CSE Dept, K.S.I.T, Benagluru

Associate Professor, Department of CSE, K.S.I.T, Bengaluru-560109

**Abstract:** Satellite multispectral remote sensing imagery has been used over decades for feature extraction. The spectral classes of the imagery are finally translated into the different feature types in the image interpretation process (image processing). Presently, classification of all feature types is a manual process. Local and global climatic variability and change is inevitable which makes satellite imagery redundant in a short span of time. Due to the above stated reasons, we need an efficient and fast automatic feature extraction algorithm for better observing and organization of the resources of Earth. This paper proposes a technique to extract urban built-up, land/vegetation and water features from Enhanced Thematic Mapper Plus (ETM+) (Landsat 7) imagery. The study used three indices, Normalized Difference Water Index (NDWI), Normalized Difference Vegetation Index (NDVI), and Normalized Difference Built-up Index (NDBI) to represent three major features on Earth: built-up land, open water body, and vegetation, respectively. Consequently, the seven bands of an original Landsat 7 image were reduced into three thematic-oriented bands derived from above indices, which were combined to compose a new image. This reduced data correlation and redundancy between original multispectral bands, with overall accuracy ranging from 91.5 to 98.5 percent. All these feature extractions will be helpful in developing information systems on water bodies, vegetation areas and urban sprawling.

**Keywords:** Automation, remote sensing, Landsat, spectral index ratio, feature extraction.

## I. INTRODUCTION

Sustainability of water resources is one of the critical factors in maintaining human civilization as it exists today. But with the current trend in modifications in the land cover of the world, more so in the populated developing countries, the long-term sustainability of water resources is a question. Mapping and regular monitoring of surface water bodies for the availability of water resources at river basins/sub basins, regional, state and country level is an important aspect of water resources planning and management in India. Urban spatial areas have expanded in an accelerated speed during the last five decades, and rates of urban population growth are higher than the overall growth in many of the countries because of economic activity and transportation nodes. Expanded urbanized areas have encroached on surrounding valuable natural lands such as paddy fields, forestlands, or wetlands. Urban areas are dominated by built-up lands with impervious surfaces, and therefore the conversion of the nature lands into these impervious built-up lands may have significant impacts on the ecosystem, hydrologic system, biodiversity, and local climate which can result in the negative aspects such as the urban heat island phenomenon. As water gets depleted with every passing minute, so does our ability to find more water resources for consumption. This calls for quick factual statistics of water resources. The scope of urban spatial expansion and the resultant urban heat needs an accurate data on urban built-up areas in terms of size, shape, and spatial context. Therefore, technique is required to quickly reveal the data. Timely availability of the data is of great importance for urban planners, water resource planners and decision makers.

Satellite remote sensing data provide a synoptic coverage of fairly large areas at frequent intervals, enabling thereby to monitor the features and capture their geospatial and temporal variability; thus helping in the assessment of inter/intra-annual variations in feature spread. Using Enhanced Thematic Mapper Plus (ETM+) (Landsat 7) optical imagery, spatio-temporal information on features has to be generated and conserved by developing an automatic feature extraction algorithm. The algorithm will then enable quick processing of data and dissemination of feature information through a web-enabled information system. Landsat 7 sensors provides data eight bands but we primarily use four spectral bands, namely Green, Red, Near Infrared (NIR) and Short Wave Infrared (SWIR) regions of electromagnetic spectrum. The present study aims at automated feature mapping which encompasses spectral characterization of various types of water bodies, vegetation areas, urban areas and development of spectral relationships and algorithm for the extraction of such features that facilitates quick processing.

## II. RELATED WORK

H. Xu [1] calculated the Modified Normalized Difference Water Index (MNDWI) by exchanging bands such as Landsat TM band 5 for the near infrared band. The use of MNDWI (Modified Normalized Difference Water Index) is more efficient for extracting water information for water region with a background of built-up land areas. The



advantage of this study is reduction and removal of built-up land area noise. A.S. KiranKumar [2] in his paper used Advanced Wide Field Sensor (AWiFS). AWiFS operates in three spectral bands in Visible and near infrared region (VNIR) and one band in Short wave infrared region (SWIR). The spatial resolution was 56 meters and a combined swath of 730km achieved through two AWiFS cameras. The directory contains the Ortho-rectified multi-spectral AWiFS data. Ortho-rectification process is meant for correcting terrain relief errors, sensor orientation/attitude, scale verification and internal errors.

Paul Shane Frazier and Kenneth John Page [3] ascertained the accuracy of digital image processing techniques in order to map water bodies by using Landsat 5 TM data. The results of the analysis were compared to 6 band classification over the same area. The water boundaries that delineated by the digital classification procedures were compared to water boundaries delineated by color aerial photography of the same area taken as the TM data. The study showed that Landsat TM data maps the water bodies accurately.

### III. PROBLEM STATEMENT

Visual interpretation of satellite data provides the best delineation of water bodies, urban and vegetable areas of varied sizes but is time consuming, especially when working with high resolution data. The simple and common approach of unsupervised classification which uses an interactive self-organizing data analysis technique provides results with very low accuracy, when there is spectral overlap between water bodies with other classes. The supervised classification provides more accurate and reliable outputs compared to unsupervised method. But the supervised method may vary when used for high resolution data. The supervised method requires sufficiently large spectral training data sets and is not fully automated method. Further it does not take into account the spatial features of the objects.

The method of partial characterization classifies the features based on their texture namely smooth or rough. Because the water bodies exhibit a smooth texture when compared to the other landscape features like vegetation and buildings in the satellite imagery, they can be easily extracted using the fractal method. However, the method does not take into account the spectral features of the objects; hence different classes with varied spectral characteristics but with similar textures are classified as one class. The results may also vary significantly with image resolution.

The threshold method is by far the most widely used algorithms for the extraction of different features from satellite imagery. For instance, the threshold method is based on the fact that the reflected radiance of water in short-wave-infrared (SWIR) band is lesser than that of other objects like vegetation, buildings, bare soil, and roads. Each pixel is sent for the threshold test and is then classified as water body along with some other objects that does not truly belong to water class, providing false positive results.

Whatever be the approach, generally the user is interested in a method which is fast, accurate, and automated. Towards this objective and to upgrade the existing system, the current study in this paper puts forward an automatic approach to extract water bodies from a satellite imagery using spectral ratio indices for classification.

Satellite images obtained from remote sensing is the only way to detect and study any feature locally or globally. To date, numerous classification techniques for the extraction of features from satellite data have been reported based on the spectral separability of the features in the visible and infrared regions of the spectrum. We are adopting the mathematical formulae that deal with the band combinations of the electromagnetic spectrum. Normalized Difference Water Index (NDWI), Normalized Difference Vegetation Index (NDVI) and Normalized Difference Built-Up Index (NDBI).

### IV. SYSTEM DESIGN

We began to work with Java as it is widely used in all our applications today. Unfortunately, we were unable to download necessary library files such as media.wrapper class for image extraction, rasterization, etc. Hence, we searched for alternatives. Python programs when compared to Java programs are 3-5 times shorter. This is because of the Python's built-in high-level data types and its dynamic typing. For instance, a Python programmer need not have to waste time to declare the types of arguments or variables and Python's powerful polymorphic list and dictionary types. Because of the run-time typing, Python's run time works harder than Java's. Thus, Python was used instead of Java.

Satellite imagery in our study has been obtained from Landsat 7 or Landsat Enhanced Thematic Mapper Plus (ETM+). These images consist of 8 spectral bands with a spatial resolution of 30 meters for Bands 1 to 7. The resolution for Band 8 (panchromatic) is 15m. All bands can collect one of the two gain settings namely high or low for increased dynamic range and radiometric sensitivity, while Band 6 collects both high and low gain for all scenes. Approximate scene size is 170km north-south by 183km (east-west).

Out of the 8, we have made use of 4 bands of the ETM+ satellite images, that is, Band 2, band 3, Band 4 and Band 5. For each spectral ratio index, we have calculated a difference between two of the four distinct bands of the ETM+ bands.

The different band designations for (ETM+) images are given in Table 1.

Landsat 7 Bands	Wavelength (micrometers)	Resolution (meters)
Band 1 – Blue	0.45-0.52	30
Band 2 – Green	0.52-0.60	30
Band 3 – Red	0.63-0.69	30
Band 4 – Near Infrared (NIR)	0.77-0.90	30
Band 5 – Short Wave IR (SWIR)	1.55-1.75	30
Band 6 – Thermal IR	10.40-12.50	60 * (30)
Band 7 – Mid IR	2.09-2.35	30
Band 8 – Panchromatic	0.52-0.90	15

Table 1: Band designations of Landsat 7.

The use of the normalized difference water index (NDWI) method maximizes the reflectance properties of water by minimizing the low reflectance of near infrared (NIR) and maximizing the reflectance in the green wavelength. Studies show that this method yields better results for deeper and worse for shallower parts of the water body.

$$NDWI = \frac{(Band2 - Band4)}{(Band2 + Band4)}$$

Similarly, the use of NDVI maximizes the reflectance properties of vegetation by minimizing NIR and maximizing the reflectance in the red wavelength. For NDBI, the method maximizes the reflectance properties of water by minimizing the low reflectance of (NIR) and maximizing the reflectance in the short-wave infrared (SWIR) wavelength.

$$NDVI = \frac{(Band3 - Band4)}{(Band3 + Band4)}$$

$$NDBI = \frac{(Band5 - Band4)}{(Band5 + Band4)}$$

The entire working of our project can be depicted in the Figure 1:

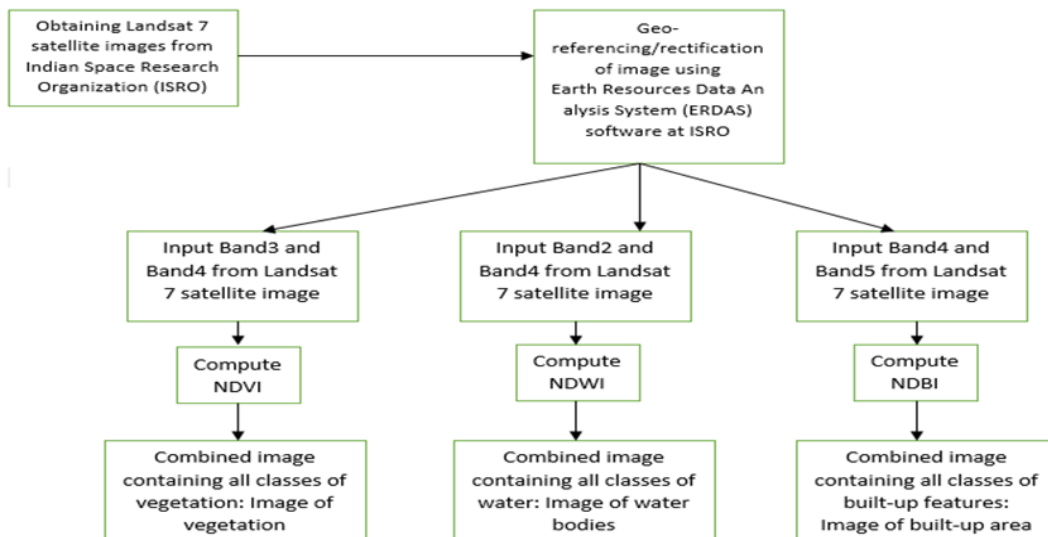


Figure 1: Flowchart depicting the working of automatic feature extraction algorithm

## V. REQUIREMENTS

System requirements consist of both hardware and software. Hardware requirements we have used are as follows:

- Intel i3 or above
- 4GB RAM
- 1GB graphic card (or above) for clearer image

Since we have worked with Python 3.4, there are a number of libraries we have used such as:



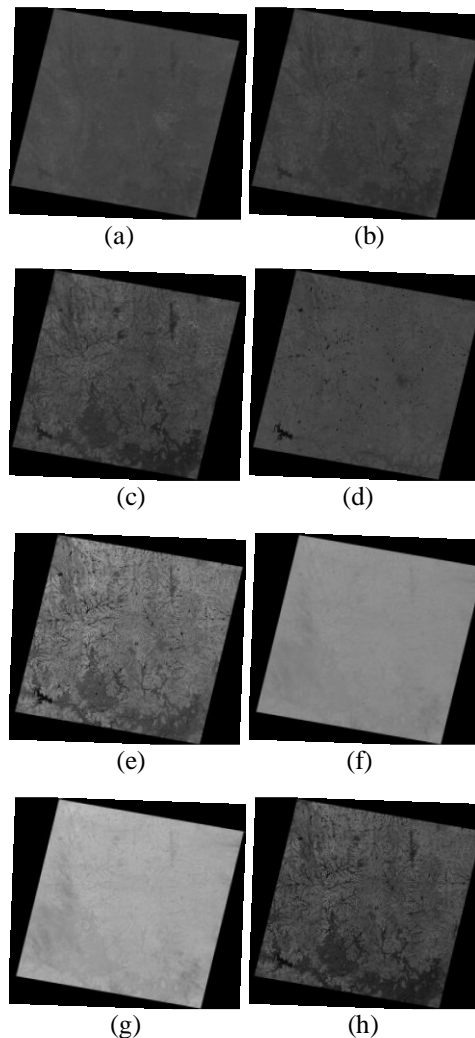
- **NumPy:** It is a fundamental package used for scientific computing in Python language. NumPy is a Python library which provides a multi-dimensional array object, various derived objects including masked arrays and matrices, and a collection of routines for fast operations on arrays, including logical, mathematical, sorting and selecting, I/O, shape manipulation, basic statistical operations, basic linear algebra, random simulation and much more.
- **The Geospatial Data Abstraction Library (GDAL):** GDAL is a computer software library. It is used for reading and writing raster and vector geospatial data formats. GDAL is released a free software license by the Open Source Geospatial Foundation. GDAL provides an abstract data model to the calling application for all supported formats. It can also be built with a variety of useful command line interface utilities for data translation and processing. Projections and transformations are supported by the PROJ.4 libraries.
- **The Open Source Geospatial Foundation (OSGeo):** OSGeo is a non-government non-profit organization. Its mission is to support and promote the development of open geospatial technologies and data.
- **Pillow:** It is a friendly fork of PIL (Python Imaging Library). It is more user friendly than PIL and is a must have for anyone who works with images.

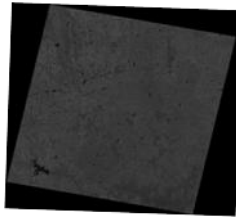
The other technologies we have incorporated in our project are:

- Framework: Django
- IDE: Text editor
- Browser: Google Chrome, Mozilla Firefox, Windows 7 and above.

## VI.IMPLEMENTATION

As said earlier, the Landsat 7 has 8 bands. However, Band 6 can be collected as high band and as low band. Bands of a typical Landsat 7 image were downloaded at the Indian Space Research Organization, Bangalore. We downloaded a scene of the city of Bangalore, as it was in January 2000.





(i)

Figure 2: Band under (a)blue wavelength, (b)green wavelength, (c)red wavelength, (d)near IR wavelength, (e)short-wave IR wavelength, (f)thermal IR (low) wavelength, (g)thermal IR (high) wavelength, (h)middle IR wavelength, (i)panchromatic. All images depict the scene in and around the city of Bangalore (Jan 2000).

The implementation is given in the form of the following algorithm:

```
//import libraries
//open images
{
extract bands for that spectral image(rasterization)
    for ndwi, extract band2 and band4
    for ndvi, extract band3 and band4
    for ndbi, extract band4 and band5
    calculate the spectral ratio indices
    if index value for both bands > 0

{
calculate min and max values of ratio of each pixel
display output image based on min, max values
}
}
```

## VII. RESULTS

The jpeg image we generate after computing the index for extraction of water bodies is given in figure 3. It gives all the water bodies, small or big, turbid or clear, within the scene size of 183km by 170km of ground area.

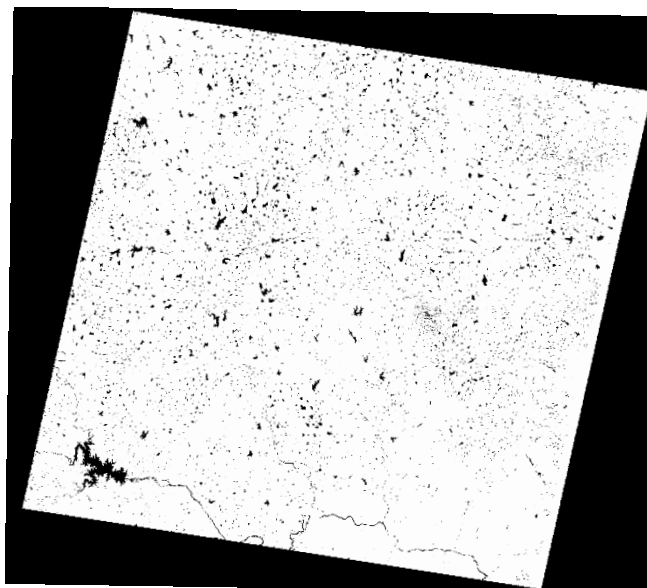


Figure 3: All water bodies extracted from the satellite image. Also seen is the Cauvery river at the bottom of the image.

The jpeg image we generate after computing the index for extraction of vegetation is given in figure 4. It displays all classes of vegetation, barren or agricultural or non-agricultural, within the scene size of 183km by 170km of ground area.

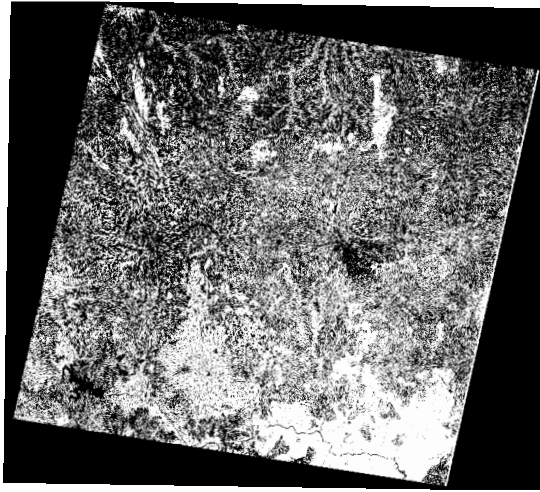


Figure 4: Image shows all vegetation types in that area.

The jpeg image we generate after computing the index for extraction of built-up area is given in figure 5. It displays all classes of man-made features, within the scene size of 183km by 170km of ground area.

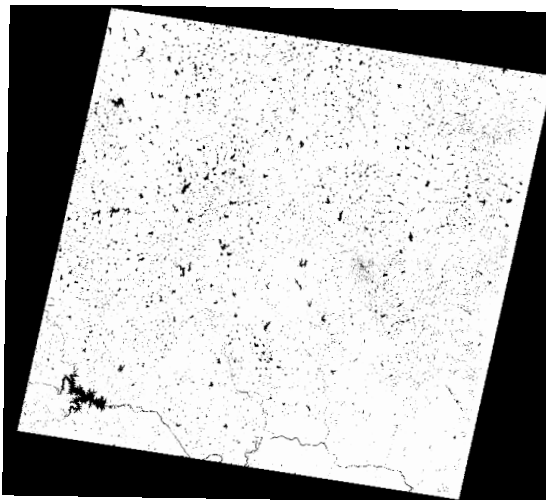


Figure 5: Image shows all built-up features within the area.

Testing and evaluation of these images was done at ISRO, Bangalore using ERDAS Imagine software. When we stacked up these images in RGB format (which is readable with naked eyes), we obtain the image shown in figure 6.

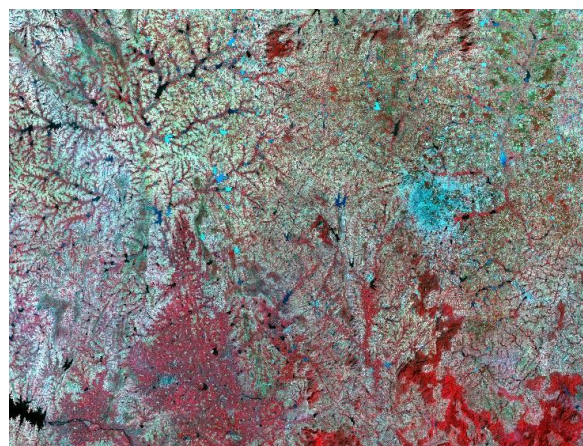


Figure 6: This is what the satellite image looks like in RGB format after all the bands have been stacked up and evaluated using ERDAS Imagine software.



## VIII. CONCLUSION

The original eight bands of a Landsat ETM+ image can be reduced into two thematic-oriented bands to produce a new image. The reducing was not carried out through the conventional methods. Instead, the three new bands were generated directly from three thematic indexes, NDVI, NDWI, and NDBI using the Python scripting language on the Django web-based framework. This dramatically reduced data correlations and redundancy between multispectral bands, significantly avoided the spectral confusion between the classes, and thus largely improved the extraction accuracy. It also increased easy access by making the whole project web-based to collect the images and output them on suitable web browsers.

For future enhancement, our aim is to make an application to accept a color image as input and output a color image using only programming platform, like Java. A few other applications, where satisfactory extraction results can be obtained include snow covered areas, cloud detection, edge inclusion and dense forest mapping using programming languages instead of existing software. All the research made for this project was done in order to create awareness and bring the space sciences and the computer science community together for future projects.

## REFERENCES

- [1]. H. Xu, "Modification of normalize difference water index (NDWI) to enhance open water features in remotely sensed imagery," *Int. J. Remote Sens.*, vol. 27, pp. 3025–3033, 2006.
- [2]. S. KiranKumar, "Advanced Wide Field Sensor (AWiFS)," *SAC Courier*, vol. 28, pp. 3–7, 2003.
- [3]. S. Frazier and K. J. Page, "Water body detection and delineation with landsat TM data," *Photogramm. Eng. Remote Sens.*, vol. 66, pp.1461–1468, 2000.
- [4]. P. Manavalan, P. Sathyanath, and G. L. Rajegowda, "Digital image analysis techniques to estimate water spread for capacity evaluations of reservoirs," *Photogramm. Eng. Remote Sens.*, vol. 59, pp. 1389–1395, 1993.
- [5]. V. Suresh Babu, M. Shanker, V. V. Rao, and V. Bhanumurthy, "Generation of water spread contours for Tungabhadra reservoir during low water levels of the year 2002 using satellite remote sensing technique," in *Proc. GIS India 2003*, Jaipur, India, 2003.
- [6]. B. S. D. Sagar, M. Venu, and B. S. P. Rao, "Distributions of surface water bodies," *Int. J. Remote Sens.*, vol. 16, pp. 3059–3067, 1995.
- [7]. V. Kulkarni, S. K. Singh, P. Mathur, and V. D. Mishra, "Algorithm to monitor snow cover using AWiFS data of ResourceSat-1 for the Himalayan region," *Int. J. Remote Sens.*, vol. 27, pp. 2449–2457, 2006.
- [8]. V. Suresh Babu, P. Roy, and S. Subramaniam, "Automatic extraction of waterbodies from IRS AWiFS data," *Int. J. Geoinform.*, vol. 3, no. 3, pp. 63–71, 2007.
- [9]. S. K. McFeeters, "The use of normalized difference water index (NDWI) in the delineation of open water features," *Int. J. Remote Sens.*, vol. 17, pp. 1425–1432, 1996.
- [10]. B. L. Markham and I. J. Barker, "Landsat MSS and TM post-calibration dynamic ranges, exoatmospheric reflectances and at-satellite temperatures," *EOSAT Technical Notes*, vol. 1, pp. 3–8, 1986.