

## VISUAL DATA MINING USING UNSUPERVISED CLASSIFICATION OF LULC SATELLITE IMAGERY

KODGE B.G.<sup>1\*</sup> AND HIREMATH P.S.<sup>2</sup>

<sup>1</sup>Department of Computer Science, S. V. College, Udgir, (MH), India

<sup>2</sup>Department of Computer Science, Gulbarga University, Gulbarga, (KA), India

\*Corresponding author. E-mail- [kodgebg@hotmail.com](mailto:kodgebg@hotmail.com)

Received: May 10, 2011; Accepted: June 06, 2011

**Abstract-** This study presents a new visualization tool using unsupervised classification of LULC satellite imagery. Visualization of feature space allows exploration of patterns in the image data and insight into the classification process and related uncertainty. Visual Data Mining provides added value to image classifications as the user can be involved in the classification process providing increased confidence in and understanding of the results. In this study, we present a prototype visualization tool for visual data mining (VDM) of satellite imagery into volume visualization. This volume based representation divides feature space into cubes or voxels. The visualization tool is showcased in an unsupervised classification study of high-resolution imageries of Latur district in Maharashtra state of India.

**Key words** - Visual Data Mining, Classification, LULC, GIS, 3D space feature plot.

### Introduction

Image classification based on satellite imagery is a widely used technique for extracting thematic information on land cover. This image processing step is the translation from spectral reflectance or digital numbers (DN) to thematic information. We classify objects by reducing a multiplicity of phenomena to a relatively small number of general classes. Classification is often performed to generalize a complex image into a relatively simple set of classes. A classified map is then used as input into a geographic information system (GIS) for further processing or analysis. Such inference is most often less than perfect and there is always an element of uncertainty in a classification result. As it can affect further processing steps and even decision making, it is important to understand, quantify and visualize the classification process.

Visual Data Mining (VDM) is a powerful tool which is often overlooked in favour of traditional purely non-visual data mining, defined as the process of (semi-)automatically discovering meaningful patterns in data (Tso B. and Mather P.M, 2002). VDM uses visual interaction to allow a human user to visually extract and explore patterns in data. When conducting a non-visual data mining, no matter how unbiased it may seem, the fact is that by simply choosing to carry out an automated analysis a priori assumptions have been made about what form the important results will take before analysis has actually begun. By visually mining the data this prior bias can be removed. Whilst the bias is removed, subjectivity of the analysis is increased as it is based on a user's perception, a point highlighted by many machine learning purists. However, this increased subjectivity is compensated for by a vastly increased degree of confidence in the

analysis. VDM not only seeks to allow a human user to visually mine data but also to augment the non-visual data mining process. This augmentation usually takes the form of making the automated process more transparent to the user, hence providing increased confidence (GeoEye, 2006).

VDM is not commonly applied in remote sensing applications. A traditional supervised remote sensing classification starts with a selection of training pixels or areas that represent specific land cover classes. The spectral and statistical properties of these pixels are then used to classify all unlabelled pixels in the image with a classification algorithm such as the widely used maximum likelihood classifier (commonly implemented in commercial remote sensing software). The accuracy of the classified map is tested with reference pixels that are not used in the training stage. Accuracy assessment usually takes the form of an error matrix with derived accuracy values such as the overall accuracy and the Kappa statistic. Although the error matrix provides an overall assessment of classification accuracy, it does not provide an indication of the spectral dissimilarity of class clusters, uncertainty related to the attribution of class labels to individual pixels, or the spatial distribution of classification uncertainty. In this study, we argue that VDM is an important tool for visual exploration of the data to improve insight into the classification algorithm and identify sources of spatial and thematic uncertainty. Recent studies showed that exploratory visualization tools can help to improve the image analyst's understanding of uncertainty in a classified image scene. They proposed a combination of static, dynamic and interactive visualizations for exploration of classification uncertainty

in the classification result. Keim D. A. 2002 developed a visualization tool that allowed for visual interaction with the parameters of a fuzzy classification algorithm. The study showed that visualization of a fuzzy classification algorithm in a 3D feature space plot dynamically linked to a satellite image improves a user's understanding of the sources and locations of uncertainty. In this study, we develop and present a new VDM prototype to visualize irregular shapes of class clusters and their spectral overlap in a 3D feature space plot. The tool helps to identify the location and shape of class clusters (showing spectral variance) and the overlap of these class clusters in 3D feature space to highlight sources of uncertainty in the training data for a spectral image classifier. To showcase the visualization prototype we present a classification study based on high-resolution LULC (Land Use/Land Cover) imagery of Latur district to assess the value of VDM in semi-automated image classification as shown in figure 1. This study is limited to a pixel-based classification approach; however, the visualization tool can be used for object-oriented classification as well.

### Dimensional Space Feature Plot

When considering a classification problem it is useful to visualize the image data in a 2D or 3D feature space with selected image bands on each of the axes (similar to a scatter plot). This visualization provides important insight into both the patterns in the image data and the operation of classification algorithms. Most commercial remote sensing software offer the tools to visualize a 2D scatter plot. In this study, we extend these common plots to a 3rd dimension to increase the amount of information (image bands) in the visualization. To generalise the large amount of image pixels, 3D feature space can be internally represented by a volume allowing for visualization of the density of class clusters. This volume based representation divides feature space into cubes or voxels. Each voxel is represented by a density value: a count of how many pixels fall in the region of feature space generalised by this voxel. In this way the volume is a 3D frequency histogram with each voxel recording the frequency at which ranges of pixel values occur. The size of the volume, as specified by the user, determines the degree of generalisation and the storage and processing requirements for operations on the volume.

### Visualization of Class Clusters

Creating an isosurface for each ROI and simultaneously visualizing these in a feature space plot offers two insights. Firstly, the user can examine each training cluster at varying levels of density. This is useful for traditional exploratory data analysis (EDA). Traditionally, visual EDA was used in data mining only as a means of checking that data conformed to assumptions prior to analysis (Simoff S. J., 2001). For the maximum likelihood classification algorithm this means checking the training data for a normal distribution. Thematic classes in satellite imagery often do not conform to assumptions made by classification algorithms. Secondly, the user may explore overlap between training clusters. This is another use of

traditional EDA to check underlying assumptions. Many classifiers struggle to deal appropriately with overlapping training data introducing uncertainty in the classification result. It is important to visualize both of these phenomena prior to supervised classification in order to interpret the results such analyses.

### Results

To showcase the visualization prototype we present a classification study based on high-resolution LULC imagery of Latur district area to assess the value of visualization in semi-automated image classification. The study is a simple 5 class problem with training regions as shown in Fig. 1. A random sample of 200 pixels was extracted from each training area for classification and a further 200 independently, randomly sampled pixels extracted for accuracy assessment. Visualization of the training regions is performed using all pixels in the regions. Firstly, bands 3, 2 and 1 are selected to be used for classification, and hence visualization. The tool is configured to display each region as an isosurface. The result is shown in Fig. 2. The shapes are overlap between the Rock (red) and Water (blue) classes. The tool is used to compute this overlap and display it as a new isosurface. The tool is also used to identify the pixels causing this overlap and highlight their location in image space. The new volume produced by the intersection operation is shown as a yellow isosurface. The pixels from the Rock class causing this intersection are highlighted with a purple overlay, and those from the Water class with a yellow overlay.

This visualization tool also has the ability to visualize decision boundaries and parameters. This feature of the prototype is used to visualize the decision boundaries and parameters for the 5 class problem for the minimum distance and maximum likelihood classifiers. The mean is the only property of the data used by the minimum distance classifier.

Pixels are classified according to the closest mean point in feature space. The visualization tool shows that some classification between the Water, constructed area, and vegetation, agriculture and rocky/barren area classes in the form of isosurface. The table 1. showing the classified values of 5 different classes.

The figure 5 showing the percent area covered by individual classes with respect to their RGB(Red, Green and Blue) color representation.

### Conclusion

VDM is a useful technology for image classification. This study has showcased the added value that visualization can provide to analysis of satellite (LULC) imagery. A novel volume based representation was used as a basis for visualization using isosurfaces. Isosurfaces and ellipsoids were used to construct 3D feature space plots showing the relationships between training regions and decision boundaries used during classification. Linkage of feature space visualizations and geographic space image views allowed a thorough investigation of patterns in the image data. This is a very easiest and

fastest method to classify LULC imageries for land assessment, analysis, change detection of area, planning and development, etc.

#### Acknowledgement

Authors are grateful to National Remote Sensing Centre (NRSC) of Indian Space Research Organization (ISRO), Hyderabad, India for providing the LULC of Latur district.

#### References

- [1] Hiremath P.S., Kodge B.G. (2010) *Journal of Advances in computational research*, Vol 2, Issue 1, pp. 21-24.
- [2] Simoff S.J. (2001) *SIGKDD Explor. Newsl.*, 3(2), pp. 78-81.
- [3] Stephen J chapman (2004) *MATLAB programming for engineers, 3rd ed.*, Cengage learning, New Delhi.
- [4] Thomas M. Lillesand and Ralph W. Kiefer (2000) *Remote Sensing and Image Interpretation, 4th ed.*, John Wiley & Sons, New York.
- [5] Tso B. and Mather P.M. (2002) *Classification Methods for Remotely Sensed Data. Geoinformetica, Taylor and Francis*, 2, 4.
- [6] William D. Stanley (2008) *Technical Analysis and Applications with MATLAB, India ed.*, Cengage learning, New Delhi.
- [7] Keim D.A. (2002) *IEEE Transactions on Visualization and Computer Graphics*, 7(1), 100-107
- [8] GeoEye (2006) *GeoEye - IKONOS Imagery*. <http://www.geoeye.com>.

Table 1- LULC classified values of 5 area classes

Classes	Area (%)	Red	Green	Blue	Frequency (pixels)
Constructed	0.379638	239	71	62	7932
Water	1.313144	115	146	206	27489
Rocky/Barren	12.63076	245	240	172	264383
Vegetation	28.1574	170	195	58	589333
Agriculture	57.55275	248	250	129	1204596

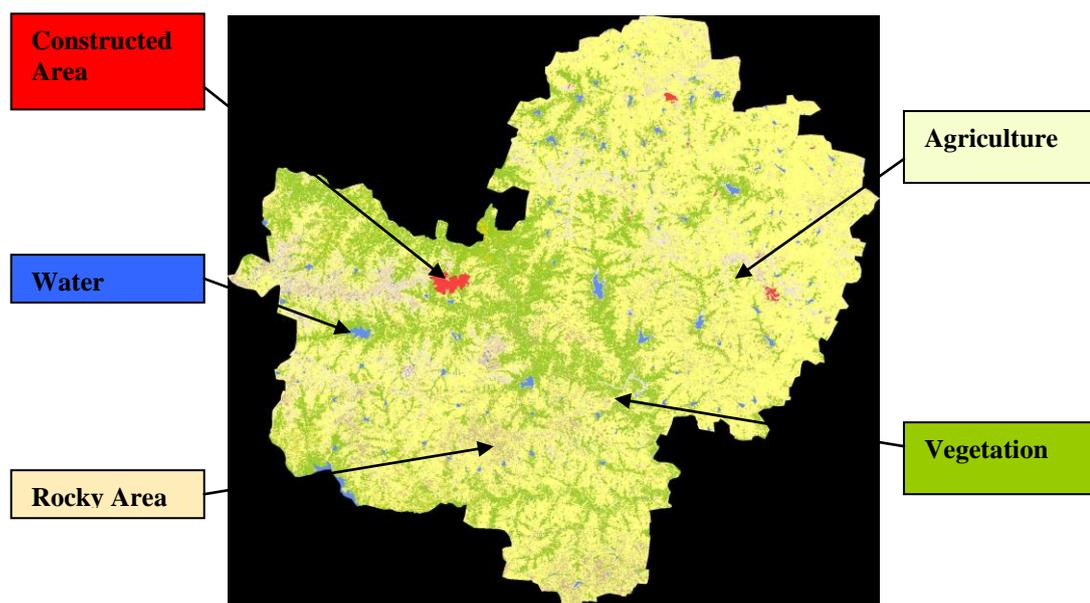


Fig. 1- LULC image of Latur district area (Courtesy NRSC (ISRO), Hyd.)

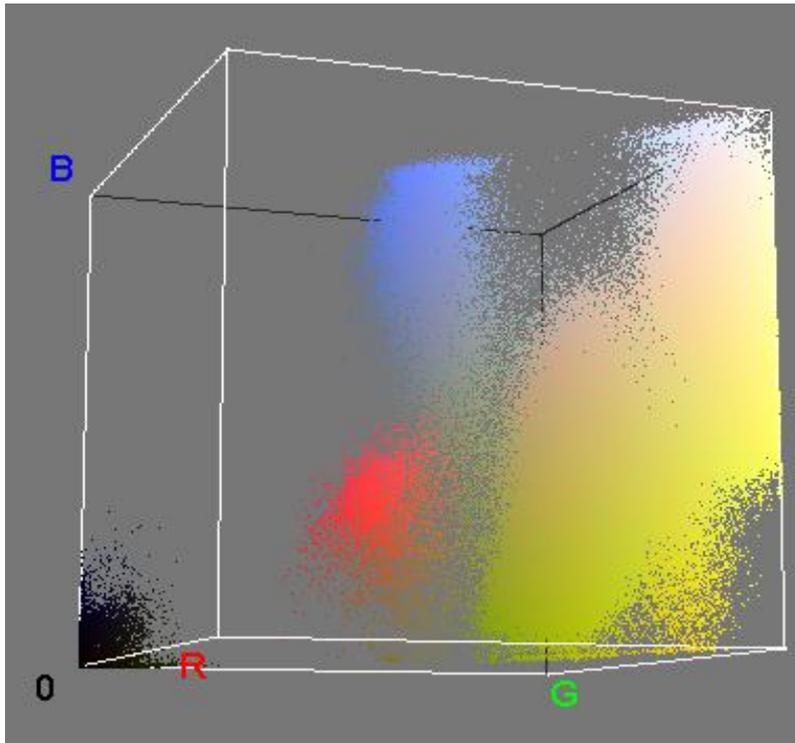


Fig. 2- Feature space plot showing isosurface for 5 training regions

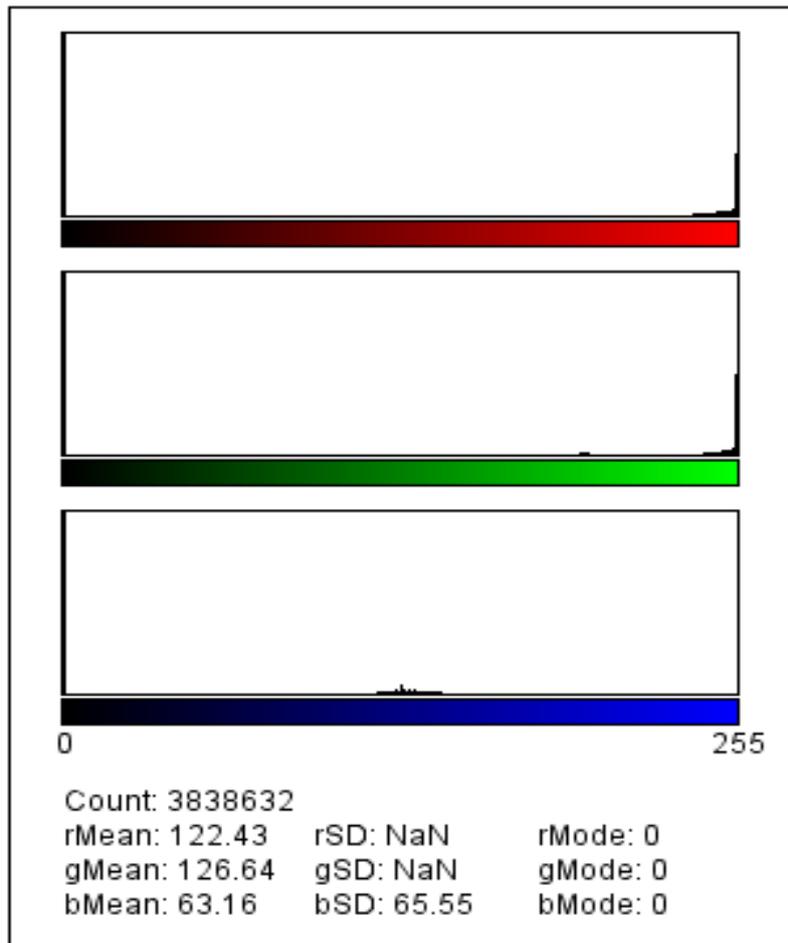


Fig. 3- Color histogram of LULC

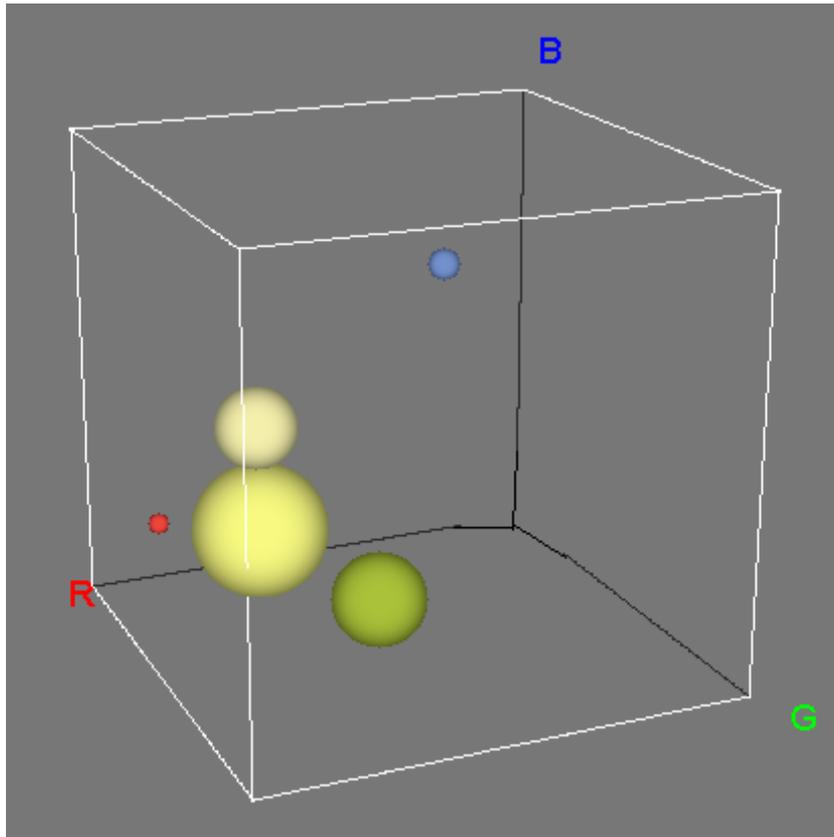


Fig. 4- Isosurface of classified regions of Latur LULC (5 classes)

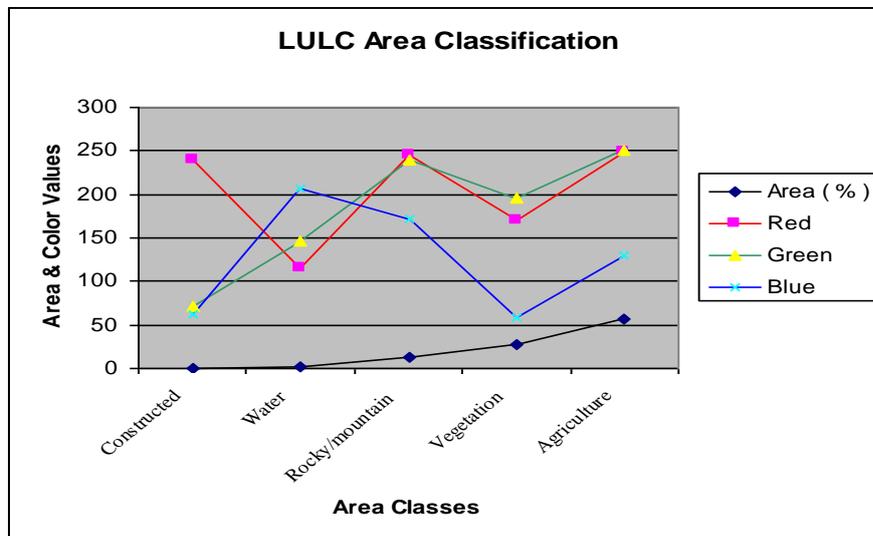


Fig. 5- LULC classified area