

Visualization techniques for data mining of Latur district satellite imagery

Hiremath P.S. and Kodge B.G.*

Department of Computer Science, Gulbarga University, Gulbarga, Karnataka State, India, hiremathps53@yahoo.com

*Department of Computer Science, S. V. College, Udgir, Dist. Latur, Maharashtra, India, kodgebg@hotmail.com

Abstract- This study presents a new visualization tool for classification of 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. The visualization tool is showcased in a classification study of high-resolution imageries of Latur district in Maharashtra state of India.

Keywords: Data Mining, 3D space feature plot, Remote sensing, Visualization

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 (Tso and Mather, 2001). 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 (Witten, 2005). 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 (Simoff, 2002). 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 (Keim, 2002). 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. 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. Lucieer (2004) and Lucieer and Kraak (2004) 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 IKONOS imagery of Latur district to assess the value of VDM in semi-automated image classification. This study is

limited to a pixel-based classification approach; however, the visualization tool can be used for object-oriented classification as well.

METHODS

Study area and imagery.

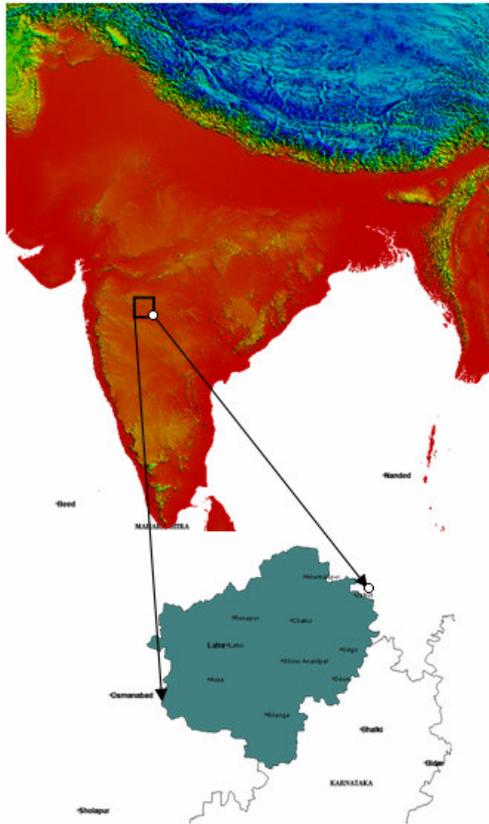


Fig. 1- A false color composite imagery of India, acquired by SPOT & IKONOS, the location of Latur district (Black box).

Latur District is in the south-eastern part of the Maharashtra state in India. It is well known for its Quality of Education, Administration, food grain trade and oil mills. Latur district has an ancient historical background. The King 'Amoghvarsha' of Rashtrakutas developed the Latur city, originally the native place of the Rashtrakutas. The Rashtrakutas who succeeded the Chalukyas of Badami in 753 A.D called themselves the residents of Lattalut. Latur is a major city and district in Maharashtra state of India. It is well known for its quality of education, administration, foodgrain trade and oil mills. The district is divided into three sub-divisions and 10 talukas (sub-districts). The number of village in Latur district according to the 2001 census, is 943. Latur is located at 18°24'N 76°35'E / 18.4°N 76.58°E / 18.4; 76.58. It has an average elevation of 631 meters (2070 feet). It is situated 636 meter above mean sea level. The district is situated on Maharashtra-Karnataka boundary. On the eastern side of the Latur is Bidar district of Karnataka, whereas Nanded is on the Northeast, Parbhani district on the northern side, Beed on the Northwest and Osmanabad on the western

and southern side. The entire district of Latur is situated on the Balaghat plateau, 540 to 638 meters from the mean sea level.

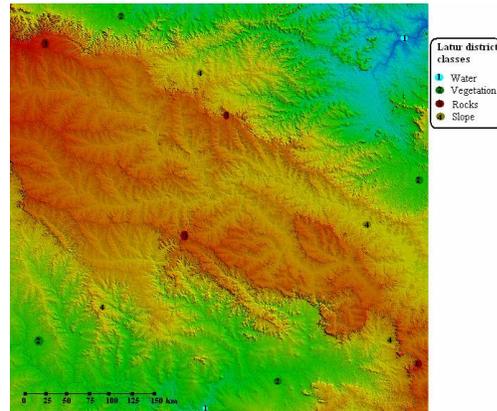


Fig. 2- Subset of IKONOS image of Latur district area showing the locations of ground reference samples as colored circles.

Operating Environment

The prototype tool has been built in IDL (the Interactive Data Language) as an extension to ENVI (the Environment for Visualising Images) using the iTools (Intelligent Tools) framework (ITTVIS, 2006). ENVI provides various image analysis and manipulation features, such as Regions of Interest (ROI), which are used by the prototype to extract sample pixels from the image. The iTools framework provides the 3D engine for the feature space plots and the user interface for interaction with the 3D objects.

3D space feature plot

When considering a classification problem it is useful to visualise 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 visualisation 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 visualisation 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 visualising 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 (Wegman, 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 visualise both of these phenomena prior to supervised classification in order to interpret the results such analyses.

Results

To showcase the visualisation prototype we present a classification study based on high-resolution satellite imagery of Latur district area to assess the value of visualisation in semi-automated image classification. The study is a simple 4 class problem with training regions as shown in Fig. 2. 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. Visualisation of the training regions is performed using all pixels in the regions. Firstly, bands 4, 2 and 1 are selected to be used for classification, and hence visualisation. The tool is configured to display each region as an isosurface. The result is shown in Fig. 3. The shapes are coloured according to the legend in Fig. 2. Fig. 3 shows some possible 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 visualisation tool also has the ability to visualise decision boundaries and parameters. This feature of the prototype is used to visualise the decision boundaries and parameters for the 3 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 visualisation tool shows that some misclassification between the Water and Rock classes may be expected.

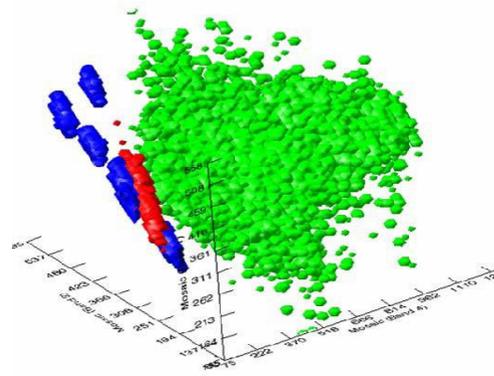


Fig. 3- Feature space plot showing isosurface for 3 training regions (X-axis = band4, Y-axis = band 2, Z-axis = band 1)

Conclusion

VDM is a useful technology for image classification. This study has showcased the added value that visualisation can provide to analysis of satellite imagery. A novel volume based representation was used as a basis for visualisation 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 visualisations and geographic space image views allowed a thorough investigation of patterns in the image data.

References

- [1] Aplin P., Remote sensing: Ecology. Progress in Physical Geography, 29(1): 104-113.
- [2] Bastin L., Fisher P.F. and Wood J. (2002) *Computers & Geosciences*, 28(3): 337-350.
- [3] Blenkinsop S., Fisher P.F., Bastin L. and Wood J. (2000) *Cartographica*, 37(1): 1-13.
- [4] GeoEye (2006) *GeoEye - IKONOS Imagery*. <http://www.geoeye.com>.
- [5] ITTVIS (2006) *ITT Visual Information Solutions-IDL-ENVI*. www.itvis.com
- [6] Keim D.A. (2002) *IEEE Transactions on Visualization and Computer Graphics*, 7(1): 100-107.
- [7] Lillesand T.M. and Kiefer R.W. (2005) *Remote Sensing and Image Interpretation*. Wiley and Sons, New York.
- [8] MacEachren A.M. and Kraak M.J. (2001) *Cartography and Geographic Information Systems*, 28(1): 3-12.
- [9] Robinson S., Turnbull J., Brandner S. and Fyfe S. (2006) *Can remote sensing be used to map vegetation and monitor community change in Antarctica?*, Australian Antarctic Data Centre - CAASM Metadata. http://aadcmaps.aad.gov.au/aadc/metadata/metadata_redirect.cfm?md=AMD/AU/ASAC_2

- [10] Simoff S.J. (2002) *VDM@ECML/PKDD2001: the International Workshop on Visual Data Mining at ECML/PKDD 2001*. *SIGKDD Explor. Newsl.*, 3(2): 78-81.
- [11] Tso B. and Mather P.M. (2001) *Classification Methods for Remotely Sensed Data*. Taylor and Francis.
- [12] Wegman E.J. (2001) *Visual Data Mining*, Center for Computational Statistics, George Mason University, VA. [http://www.galaxy.gmu.edu/stats/syllabi/inf979/- VisualData Mining.pdf](http://www.galaxy.gmu.edu/stats/syllabi/inf979/-VisualDataMining.pdf), Last accessed: 30 Oct.
- [13] Witten I.H. and Frank E. (2005) *Data Mining - Practical Machine Learning Tools and Techniques*. Morgan Kaufmann.