An Useful Information Extraction using Image Mining Techniques from Remotely Sensed Image (RSI)

Dr. C. Jothi Venkateswaran
Head, P. G. Department of Computer Science
Presidency College, Chennai.

S. Murugan*
Associate Professor of Computer Science
Nehru Memorial College, Puthanampatti.

Dr. N. Radhakrishnan
Director, Geocare Research Foundation, Chennai.

Abstract

Information extraction using mining techniques from remote sensing image (RSI) is rapidly gaining attention among researchers and decision makers because of its potential in application oriented studies. Knowledge discovery from image poses many interesting challenges such as preprocessing the image data set, training the data and discovering useful image patterns applicable to many new application frontiers. In the image rich domain of RSI, image mining implies the synergy of data mining and image processing technology. Such culmination of techniques renders a valuable tool in information extraction. Also, this encompasses the problem of handling a larger data base of varied image data formats representing various levels of information such as pixel, local and regional. In the present paper, various preprocessing corrections and techniques of image mining are discussed.

Key words: image mining, remote sensing image, knowledge discovery, image processing, data format.

1. Introduction

Availability of large quantities of remotely sensed image (RSI) provides an opportunity to extract qualitatively useful information facilitated by advances in computing technology. Extracting interesting patterns and rules from data sets composed of images and associated ground data can be of immense significance in many resources inventory such as precision agriculture, forestry, minerals and so on [1]. This may also be related to study and assess many natural calamities such as forest fire, earth quakes, flood and storms and become a more powerful tool when integrated with other spatial data sets and attributes [2].

RSI in general can be defined as a "two-dimensional" array of pixels. Each pixel of the image data contains various attributes such as bands that control the reflectance value or digital number (DN) of the pixel. In other words, DN values are the reflectance intensities of objects in different spectral band width such as Blue, Green, Red and Infrared. Such intensities under different band width of a pixel represent some particular "character" or aspect of an object and cumulatively giving out certain unique "pattern". This basic concept is well utilized in image mining to extract information and to generated "knowledge data base" to assess natural resources, agricultural yields, soil moisture stress and so on. Knowledge discovery from RSI lies in the inherent attribute, digital numbers (DN) that allows qualitative computational mining methods involving other parameters [3]. The pixel coordinates in raster order constitute the key factor. However, in most cases the image data sizes are too large to be mined in a reasonable amount of time using standard methods and hence require segmentation, partitioning or tiling of images. Moreover, handling of image database necessitates an understanding of the data formats available in the image domain.

The rest of the paper is organized as follows: Section 2 describes the concept of various data formats available in RSI. The different image mining techniques like statistical classification, decision based and association related mining techniques are described in Section 3. Section 4 ends with conclusion.

2. RSI Data Formats

Extraction of information from Remotely Sensed Image (RSI) requires knowledge on data structure so that extraction algorithm may be designed. The sensors on-board remote sensing satellite capture energy at different parts of the electromagnetic spectrum, thus reflected by various objects, converts and store them as pixels to generate digital imagery.

The data may be attenuated and contain noise due to various parameters such as air, water molecules, dust, and any other system aberrations. The gathered information stored as digital values or numbers (DN) and three types of data storage formats (Figure 1) such as band interleaved by pixels (BIP), band interleaved by lines (BIL) and band sequential format (BSQ) are observed wherein array of pixels as discrete patches or raster data.

The numerical value of the pixel, Digital Number (DN), is translated into a shade of gray that ranges somewhere between white and black and when arranged together in the correct order, they form an image of the target in which the varying shades of gray represent the varying energy levels detected on the target. By applying varied algorithms, these images could be processed after correction adapting preprocessing techniques such as radiometric and spectral as well as spatial rectification. Preprocessing corrections include noise reduction, enhancing the spectral values, substituting line drop outs by applying interpolation, eigen values and binomial to retrieve relevant partition of data based on features to be extracted either at pixel level or tile level or region level.

3. Image mining techniques in RSI

Image mining at pixel level involves spectral and textural information about each pixel of the image [5]; similarly, region level features describe groups of pixels; and tile level features present information about whole images using texture, percentages of end members and others. The task of extracting information from image datasets using various methods of approach include statistical classification, decision based and association related mining techniques.

3.1 Classification in RSI

Classification is a useful approach to mining information from spatial data employing computational and statistical methods. In other words, the tasks of image mining in RSI are mostly concerned with *classification* problems such as “labeling" regions of an image based on presence or absence of some characteristic patterns, and with *image retrieval* problems [7] where “similar" images are identified. In classification, a training (learning) set is identified for the construction of a classifier [8]. Each record in the training set has several attributes. There is one attribute, called goal or class label attribute, which indicates the class to which each record belongs. A test set is used to test the accuracy of the classifier once it has been developed from the learning dataset. The classifier, once certified, is used to predict the class label of unclassified data. This is for the purpose of getting knowledge obtained from training that could be applied on a large number of new datasets or database containing images to fulfill the required tasks, which are especially useful in task such as classifying land cover [9]. Different models have been proposed for classification, such as decision tree induction, neural network, Bayesian, fuzzy set, nearest neighbor and so on. Among these models, decision tree induction is widely used for classification.

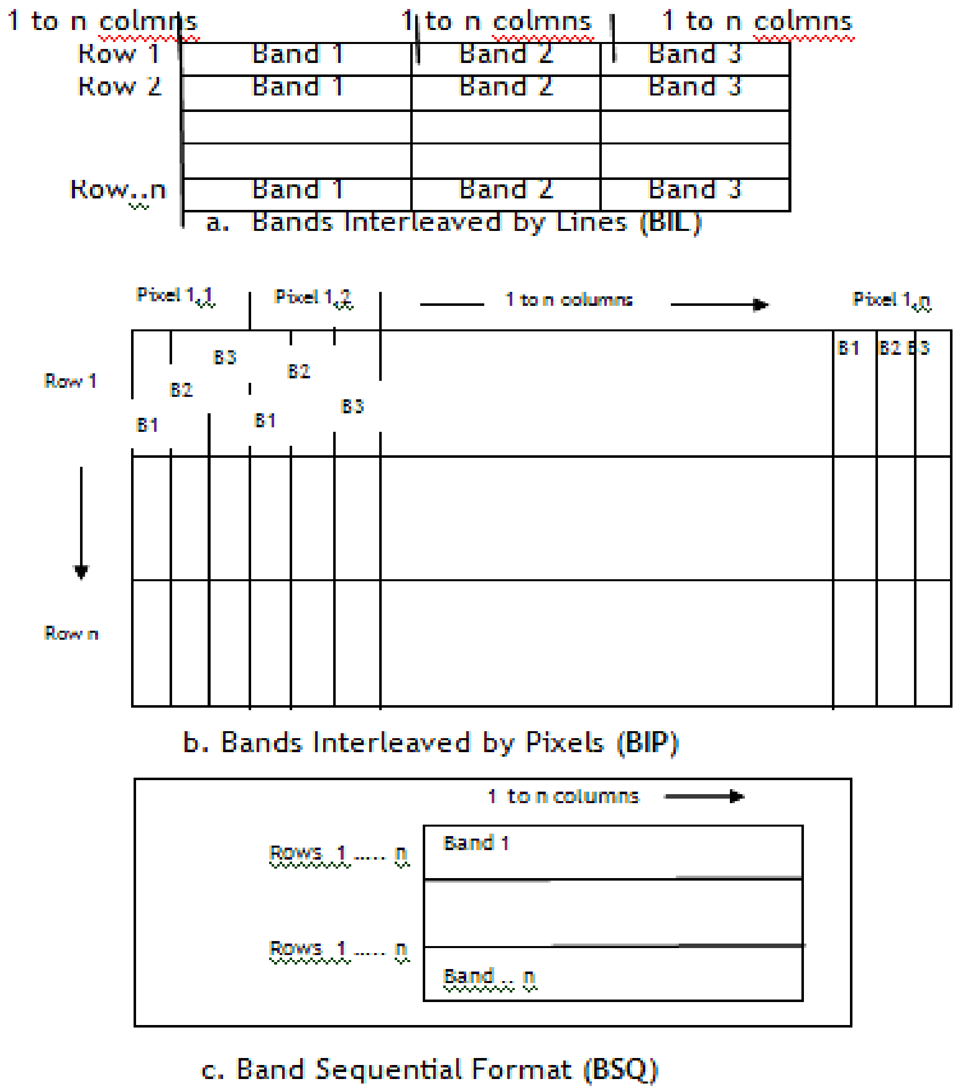

Figure 1. Data formats in Remote sensing Satellite Image

3.2 Association rule mining in RSI

Likewise, in the spatial data domain, association rule mining (ARM) is useful in identifying forest fires, insect and weed infestations, high and low crop yields, flooding and other phenomena as rule consequences. The antecedents of such rules are typically taken from the Remotely Sensed Imagery (RSI) data bands.

An association rule is a relationship of the form X=>Y, where X and Y are sets of items. X is called antecedent and Y is called the consequence. There are two primary measures, support and confidence, used in assessing the quality of the rules. The goal of association rule mining is to find all the rules with support and confidence exceeding user specified thresholds. For example, in association rule mining, it is attempted to derive rules with yield as consequent. A rule like "Green [192,255] ^ NIR [0, 63] => Yield [128, 255]" is expected, where Green [192,255] indicates an interval with value ranged from 192 to 255 in Green band. Similarly, in classification the "yield" is specified as the goal.

To perform association rule mining on RSI data, data partition is required since RSI data are quantitative data [10]. There are various kinds of partition approaches, including Equilength partition, Equi-depth partition and user customized partition. As it has been mentioned, frequent item sets generation is the key step in association rule mining. Usually a step-wise procedure is used to generate frequent item sets [5]. To determine if a candidate item set is frequent, the support is calculated then compared to the

threshold. In Apriori and most other ARM algorithms, the entire transaction database needs to be scanned to calculate the support for each candidate item set. When the transaction set is large, a large image with tens of millions of pixels, the cost will be extremely high. To circumvent this, each pixel may be considered as transaction, and after performing data partition, items are all the intervals in all the bands. An item set is in the form of Int1 x Int2 x ... x Intn = Ω
i=1...n Int..i, where Int..i is an interval of Values in Band'I' (some of which may be the full value range 0~255). A 1-itemset is an item set with (n-1) full value range intervals, while a 2-itemset is an item set with (n-2) full value range intervals, where n is the total number of bands. This is applied on the entire image database. Both association rule mining and classification have been applied in many fields.

4. Conclusion

Remotely Sensed image data is one of the promising application areas since there are huge amount of image data. Large size of image data necessitates multi-oriented mining techniques that depend upon the "goal". Object oriented "yield" provides a fast way to calculate some measurements for mining task, such as support and confidence in association rule mining task and information gain in classification task either from individual image or from temporal images. Mining from image also aid in generating predictive models for detecting spatial processes in a very cost effective manner. Thus, combination of various techniques such as statistical, decision, parametric and association rules help in extracting information effectively from RSI rather than applying any single method.

References

[1] Angayakkani, K. and Radhakrishnan, N. (2009), "Efficient Forest Fire Detection System: A spatial mining and Image Processing Based Approach", Intl. Jou. of Comp. Sci. and network Security, Vol.9, No.3, pp 100 - 107.

[2] Han, J., Fu, Y., Wang, W., Koperski, K., and Zaïane, O. R. (1996), "DMQL: A Data Mining Query Language for Relational Databases", In Proc. of the Workshop on Research Issues on Data Mining and Knowledge Discovery, Montreal, QB, pp. 27 – 34, 1996.

[3] Tadjudin, S. and Landgrebe, D. A. (2000), "A Robust parameter estimation for mixture model", IEEE Trans. Geoscience and Remote Sensing, Vol 38, No. 1, pp. 439-445.

[4] Kumar, S. Ghosh, J. and Crawford, M. M. (2001), "Best basis feature extraction algorithms for classification of hyper spectral data", IEEE Trans. Geosciences and Remote Sensing, Vol 29, No. 7, pp. 1368-1379.

[5] Zhong, Y., Jain, A.K., (2000), "Object location using color texture and shape". Pattern Recognition, 33 (4), pp. 671- 684.

[6] Hand.D, Mannila.H, and Smyth.P (2001), "Principles of Data Mining". A Bradford Book The MIT press, 2001.

[7] Schröder, M., H. Rehrauer, K. Seidel, and M. Datcu (2000), "Interactive Learning and Probabilistic Retrieval in Remote Sensing Image Archives". In IEEE Trans. on Geosciences and Remote Sensing, Vol. 38, No. 5 pp. 2288 - 2298.

[8] Dash. M and Liu. H, (1997)," Feature selection methods for classifications", Intelligent Data Analysis: An International Journal, Vol.1-3, pp 131-156.

[9] Morgan, J. T, Henneguelle, A. Ham, J. Ghosh, J. and Crawford, M. M. (2004), "Adaptive feature spaces for land cover classification with limited ground truth data, Intl. J. Pattern Recognition and Artificial Intelligence, Vol 18, No. 5, pp. 777-800.

[10] Zhang, C., and Baltsavias, E., (2002), "Edge matching and 3D road reconstruction using knowledge based methods". Institute of Geodesy and Photogrammetry. ETH Hönggerberg, CH-8093 Zurich. Switzerland.