International Journal of Scientific & Engineering Research, Volume 4, Issue 9, September-2013
ISSN 2229-5518

Datamining Techniques used for Classification of High Resolution Remote Sensing Images

Bharathi S, P Deepa Shenoy, Venugopal K R, L M Patnaik

Abstract— Data mining is a form of knowledge discovery essential for solving problems in a specific domain. In this paper KDD is used to discover knowledge from remote sensing database. Combined feature of texture and color is used to create a feature databased from different resolution of remote sensing images. The classification is performed using Bayes, SVM and NN. All the three classifiers are compared and give the very good result for the image of different resolutions. Time complexity of each classifier is computed. The time complexity of the neural network is little expensive compared to other two classifiers. The proposed algorithm shows excellent accuracy assessment even if the image resolution changes.

Index Terms— remote sensing, feature extraction, clustering, classification, accuracy assessment.

1. INTRODUCTION

Image classification is an important part of the remote sensing, image analysis and pattern recognition. Satellite imagery, a remote sensing technique, is convenient for large scale surveys, and has been used widely for land cover and habitat mapping using different applications. Remote sensing image classification can be viewed as a joint venture of both image processing and classification techniques. In this paper color and textures are used for feature extraction and different classification algorithms are used for classification model. Digital representation of color images is realized by storage of color intensity values of each pixel. RGB space is a widely used color space for image display. It is composed of three color components red, green and blue. Since color cameras, scanners and displays are most often provided with direct RGB signal input and output, this color space is the basic one, which is, if necessary, transformed into other color spaces. Texture is ubiquitous in natural images and constitutes an important visual cue for a variety of image analysis applications like image segmentation, image retrieval and shape from texture. Texture classification is a fundamental issue in computer vision and image processing, which plays a significant role in a wide range of applications that includes medical image analysis, remote sensing, object recognition, content-based image retrieval, and many more. Due to its importance, texture classification has been an active research topic over several decades. The design of a texture classification system essentially involves two major steps: (1) Feature extraction and (2) Classification. Most research in texture classification focuses on the feature extraction part. Texture analysis refers to a class of mathematical procedures and models that characterize the spatial variations within imagery as a means of extracting information. Data mining is a part of a larger area of recent research in artificial intelligence and information management: knowledge discovery in databases. The quality of a supervised classification depends on the quality of the training sites. The training sites are done with digitized features. Supervised classification is done using neural network (NN), SVM and Bayes classifier.

1.1 Organization

This paper is organized as follows – Section 2 deals with related topics and Section 3 describes the study area. Section 4 presents the architecture model and methodology and Section 5 is about the problem definition. Section 6 gives the implementation of the proposed algorithm and performance analysis. Section 7 contains the conclusion.

2. RELATED WORK

Li Liu and Paul W. Fieguth [1] described a classification method based on representing textures as a small set of compressed, random measurements of local texture patches, leading to results matching or surpassing the state of the art in texture classification, but with significant reductions in time and storage complexity.

Stephen Moysey, Rosemary J. Knight, and Harry M. Jol [2] compared six measures of radar texture that can be used to characterize GPR data. Their analysis indicates that radar texture can be a powerful criterion for discriminating between radar facies.

K Perumal and R Bhaskaran [3] analyze the performance of various classifiers and found that the Mahalanobis classifier outperforms even advanced classifiers for land use/land cover. This is accurate but simple classifier shows importance of considering the data set - classifier relationship for successful image classification. Syaza Putri Abdul Rahman Putra[et al], [4] used texture measures for supervised classification using Maximum Likelihood Classifier (MLC) and K Nearest Neighbor (KNN) classifiers. Accuracy assessment for each measure was carried out using random ground samples. Maximum Likelihood Classifier, the parametric classifier, performed better compared to KNN classifier.

- *Bharathi S is Research Scholar, Department of MCA, Bangalore university, Bangalore, India and working in DR.AIT, Bangalore as Associate professor in the dept of MCA. E-mail: bharathishivu_s@yahoo.co.in*
- *P Deepa Shenoy, Venugopal K R, are Professors in Department of Computer Science and Engineering, UVCE, Bangalore, India*
- *L M Patnaik is Honorary Professor, IISC, Bangalore, India*

K.Angayarkkani, Dr.N.Radhakrishnan [5] discusses novel and efficient approach to detect forest fires from spatial data images. The fuzzy rules derived using the proposed approach, have successfully detected the forest fires in the spatial data. Gang Li, Y ouchuan Wan [6] proposes a new object-oriented classification method based on improved watershed segmentation and fuzzy support vector machine. Wuwei [7] tested a new classification strategy for remote sensing images use neuro-fuzzy model NEFCLASS and texture analysis and found that this proposed algorithm achieve better accuracy.

3. Study area

The area taken for study is Bangalore of different resolution. This area is highly heterogeneous, which has the combination of different features. The study area contains the buildings, water bodies, roads, barren land and green area. The land use/land cover classification is important for management and modification of the city. The remote sensing images considered are a cartosat-1(IRS-P5) of resolution 2.5m and 4 bands (green, red, NIR, SWIR). It is between the longitude 77 35 18.30″ E and latitude 12 58 56.049″ N of spectral bands 0.50-0.85 micron. Images are acquired from Karnataka remote sensing department in 2012.The other image taken is LandSat of 15m resolution and 4 bands (red, green, blue, swri2) in 2006.

4. Methodology

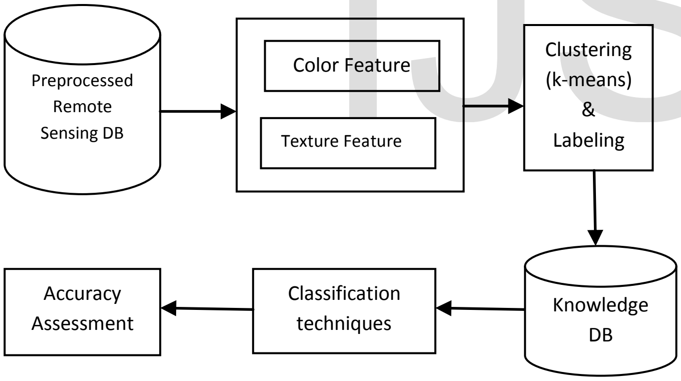

Preprocessed Remote Sensing DB
Color Feature
Texture Feature
Clustering (k-means) & Labeling
Knowledge DB
Classification techniques
Accuracy Assessment

Fig. 1. Architecture for classification and land use mapping

4.1 Preprocessing

After obtaining satellite images in the digital format, images are Geo referenced and mosaiced using ERDAS IMAGINE image processing software. Geo-referencing attachés real world coordinates to the image so that it can be co-registered with any other imagery or spatial data, which overlie the same area. Geo-referencing also enables warping an image to correct the topographic displacement. Real worlds Ground Control Points (GCP), obtained with a GPS, are used for Geo-referencing along with well-distributed points from Geo-coded hard copy of the image. Geo-referencing is done on Lambert Conformal Conic (LCC) projection, Everest spheroid and an undefined datum.

Once the data is acquired, it needs preprocessing. This includes geometric corrections, radiometric correction and image clipping. Geometric corrections and radiometric corrections are already been done. Image is in RGB format, it is converted into gray-scale image and unwanted part is clipped off from the image. Image is passed through the digital filters to remove noise and inconsistency. An image is normalized by scaling its values so that they fall within a small-specified range, such as 0.0 to 1.0. Range [new-mina, new-maxa] Variance is scaled to fit in the range one. This preserves the relationship with the original image. Range [new-min$_a$, new-max$_a$] is computed by

$$\begin{aligned} v' = ((v - \text{min}_a)/ (\text{max}_a - \text{min}_a)) \\ * (\text{new\_max}_a - \text{new\_min}_a) \\ + \text{new\_min}_a \end{aligned} \qquad (1)$$

Variance is scaled to fit in the range one and it is computed by,

$$v' = ((v - \bar{A})/ \sigma_A) \qquad (2)$$

where $\bar{A}$ and σ_A) are the mean and the standard deviation respectively of image A. Both are linear operation.

4.2 Feature Extraction Using Color and Gabor Texture

For color feature extraction technique image color distribution technique is used. The multi-spectral value of pixels in 16x16 neighborhoods is used to extract RGB values. Gabor filter is one of the most effective feature extraction techniques for textures. As the Gabor filters are believed to be rather consistent to the response of Human Vision System (HVS). In the spatial domain, a 2D Gabor filter is a Gaussian kernel function modulated by a sinusoidal plane wave.

$$f(x, y, \omega, \theta, \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma x \sigma y} \exp\left[\frac{-1}{2}\left(\left(\frac{x}{\sigma_x}\right)^2 + \left(\frac{y}{\sigma_y}\right)^2\right) + j\omega(x\cos\theta + y\sin\theta)^2\right] \qquad (3)$$

where σ is the spatial spread, ω is the frequency and θ is the orientation.

4.3 Clustering (K-Means)

Once feature is extracted using texture, it is required to cluster the data for labeling. It is the simplest method in unsupervised technique. This algorithm partitions data into k mutually exclusive clusters. This clustering technique is often more suitable for large amounts of data. It finds a partition in which objects within each cluster is as close to one another as possible, and as far from objects in other clusters as possible. Each cluster in the partition is defined by its member objects and by its centroid or center.

4.4 Classification

Bayes classification (or maximum likelihood classification) is most widely used. It can obtain minimum classification error under the assumption that the spectral data of each class is normally distributed.SVM is an optimal classifier with a maximal margin in feature space, nonlinear and accurate ways to classify the data. "Neural network" (NN), is a mathematical model or computational model that tries to simulate the structure and/or functional aspects of biological neural networks.

4.5 Accuracy Assessment

Accuracy is a measure of how well the model correlates an outcome with attributes in the data that has been provided. There are various measures of accuracy, but all measures of accuracy are dependent on the data that are used. Several accuracy measures are available. But we use confusion matrix, kappa statistics, absolute and relative error, root mean squared error.

5 Algorithms

a. Problem Definition: For a given multispectral satellite image, the main objective of this work is to:

i) To develop an efficient feature extraction algorithm using color and texture.

ii) To build a good land cover classification and prediction model.

Algorithm: The main objective is to obtain better feature from the satellite image for different features in the urban area and build a good classification model using data mining techniques. The algorithm for classification is given in the table1.

Input: m: High resolution image

n: number of classification

Output:

a. Feature databases

a. A set of n number of classifications with very good accuracy

TABLE 1 Satellite Image Classification Using Data Mining Techniques.

```
Cls-mod()
Begin
        Preprocess the image m.
        color_featr (rgb).
        gabor_filter(i, k, gamma, lambda, b, theta, phi, shape)
        K_means(data, k)
        Classification model (dataset)
End.
color_featr(rgb)
        repeat
        extract color features  by taking 16X 16 mask image
        for each row
        find the most common object which those 256 pixels belong to
        add it as a column in the dataset

gabor_filter(i, k, gamma, lambda, b, theta, phi, shape)
        Filter_bank()
        Test orientation separation angle for 300
        Adjust image size to the smallest size if 'valid'
        (Apply threshold and normalize)
        Feature extraction
        Cluster-validation (result, data, param);
        Apply spatial smoothing using Gaussian filter.
        Clustering of pixels in feature space.

cluster-validation (result, data, param)
        Compute partition coefficient PC to measure the amount of overlap.
        Compute the classification entropy CE to measure fuzziness of cluster
        partition.
        Compute Dunn's index for ide notifying  well separated and compact
        cluster set.
K_means(data, n)
        Repeat
        Set initial centers of clusters, $c1,c2…ck$ , to the arbitrarily
        Selected $k$ vectors
        Classify each vector $x1 = [x11,x12…x1d]$ into the closest center $ci$
        recalculate the cluster center $ci = [ci1, ci2…cid]$
        until centroid no longer
        Assign the label.
Classification model (dataset)
        Using rapid miner tool
        Bayes Classifier
        Neural Network Classifier
        SVM classifier
        Accuracy assessment
```

6 Implementation And Performance Analysis

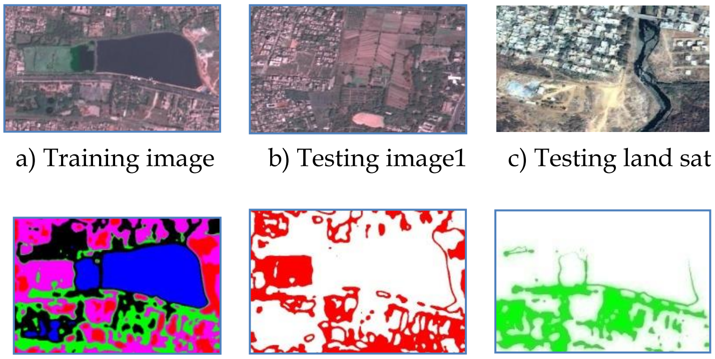
6.1 Simulation Software

Simulation is performed using Matlab7.5. MATLAB has an extensive library of built-in functions for data manipulation. Data preprocessing, image segmentation is done using the tools supported by MATLAB. Rapid miner 5.0 is used for classification and accuracy assessment. Rapid Miner provides data mining and machine learning procedures including: data loading and transformation, data preprocessing and visualization, modeling, evaluation, and deployment.

6.2 Performance Analysis

One of the important requirements in image retrieval, indexing, classification, clustering and etc. is extracting efficient features from images. While we are designing an image classification and retrieval system, the following issues has to be solved. (a) To extract the features resourcefully. (b) To classify the images. To resolve the first issue, the main color and texture features are derived from Gabor filters. For the second issue data mining algorithms are used and compared for better classification algorithm.

Color features are extracted where each row in the dataset corresponds to 16x16 masks in the image. Each pixel is a 8-bit binary word, with 0 corresponding to black and 255 to white. Each line contains the pixel values in the three spectral bands of each of the 256 pixels in the 16x16 neighborhood .Texture features are extracted using Gabor filter and then the segments obtained are validated using partition coefficient PC to measure the amount of overlap, the classification entropy CE to measure the fuzziness of the cluster partition. Dunn's index is computed for identifying well separated and compact cluster set. After these texture boundaries are well localized and smoothened the image to get sharp localization, automatically the segments are labeled as segment1, segment2 and so on. Region merging has been done to merge the neighboring area of same type and handle the occlusions. Feature extraction algorithm is fast, simple and less susceptible to initialization problem. After features are extracted from the image, we cannot apply classification on it since the data is to be labeled. Then we used k-means clustering techniques to cluster the dataset then manually the dataset is labeled. The labeled dataset is ready for classification. The requirement is to classify for five classes. Using Rapid miner tool classification is done. The three data mining classifiers are used to classify the different classes such as water body, farm land, empty land, buildings and trees. The results are validated by giving the same image as the test data. In this case we are getting the result classification accuracy above 94% in all the three classifiers.



a) Training image b) Testing image1 c) Testing land sat

d) Segmented using texture e) Feature extracted images

Fig. 2. Original image and segmented image

All three classifiers considered gives excellent result but Neural network takes 10mins to classify the dataset where Bayes and SVM takes 3mins and 5mins respectively. The computational complexity is not high, and the performance is very good with Bayes. Accuracy assessment is computed using confusion matrix, kappa statistics, absolute error, relative error and root mean squared error. Confusion matrix or matching matrix shows visualization performance of the algorithm, kappa computes the statistical measure. The confusion matrix is shown in the table 2,3 and accuracy measures for different classifiers and different images are tabulated in the table 4 and table 5.

TABLE 2. Confusion Matrix for Bayes for image c

	Water body	Empty Land	Form Land	Buildings	Trees	Classification accuracy in %
Water body	3	0	0	0	0	100
Empty Land	0	15	3	0	0	83
Form Land	0	3	38	0	2	88
Buildings	0	0	0	20	0	100
Trees	0	0	1	0	8	88

TABLE 3. Confusion Matrix For SVM For Image c

	Water body	Empty Land	Form Land	Buildings	Trees	Classificatio accuracy in %
Water body	3	0	0	0	0	100
Empty Land	0	14	1	0	2	82
Form Land	0	0	39	1	2	92
Buildings	0	0	3	39	2	88
Trees	8	0	0	0	0	100

TABLE 4. Accuracy Assessment For The Training Image b

Classifier	Accuracy	Classification error	Kappa	Absolute error	Relative error	Normalized absolute error	Root mean squared error
Bayes	93.5	6.5	0.92	0.65 ±0.12	6.5 ±15	0.119	0.125
SVM	94.8	5.2	0.94	0.52 ±0.10	5.2 ±13	0.115	0.120
NN	95.00	5.0	0.95	0.95 ±0.10	08.1 ±25	0.112	0.117

TABLE 5. Accuracy assessment for the training image c

Classifier	Accuracy	Classification error	Kappa	Absolute error	Relative error	Normalized absolute error	Root mean squared error
Bayes	91.8	8.2	0.90	0.82 ±0.163	8.2 ±23	0.157	0.139 ±0.000
SVM	92.4	7.6	0.91	0.76 ±0.14	7.6 ±22	0.134	0.120 ±0.00
NN	92.8	7.2	0.92	0.72 ±0.12	7.2 ±19	0.112	0.110 ±0.00

7. Conclusion

Classification of remotely sensed data is one of the primary steps for information extraction. High resolution remote sensing images are classified using datamining techniques. In this paper different resolution of images are considered. Accurate and up-to-date features are required for efficient classification. The complex features are extracted using the combination of color and texture. K_means cluster is used to cluster the features and labeled the dataset. Simple Bayes, SVM and neural network classifiers are used. The classification is simple one using only five categories; water body, empty land, farm land, buildings, trees. The validation is done by field survey. The proposed method gives more than 90% of accuracy when we compared with the field survey. These classifiers are very fast even if the data size very huge.

References

[1] Li Liu and Paul W. Fieguth, “Texture Classification from Random Features”, *IEEE Transactions on Pattern Analysis And Machine Intelligence*, Vol. 34, NO. 3, pp.574-585, 2012.

[2] Stephen Moysey, Rosemary J Knight, and Harry M. Jol, ”Texture-based classification of ground-penetrating radar images”, Journal of Geophysics, Vol. 71, No. 6 Nov-Dec pp.K111–K118,2006.

[3] K Perumal and R Bhaskaran,” Supervised Classification Performance Of Multispectral Images”, Journal of Computing, Volume 2, No 2, ISSN 2151-9617, February 2010.

[4] Syaza Putri Abdul Rahman Putra, Sim Chong Keat, Khiruddin Abdullah, Lim Hwee San, and M Nawawi Mohd Nordin, “Texture Analysis of AIRSAR Images for Land Cover Classification”, Proceeding of the *IEEE International Conference on Space Science and Communication (IconSpace)* 12-13 July 2011, Penang, Malaysia.

[5] K Angayarkkani, Dr. N Radhakrishnan, “Efficient Forest Fire Detection System: A Spatial Data Mining and Image Processing Based Approach”, IJCSNS International Journal of Computer Science and Network Security, vol .9 No.3, March 2009.

[6] Gang Li, Y ouchuan Wan, ”Remote Sensing Image Classification Based on Improved Watershed Segmentation and Fuzzy Support Vector Machine”, International Conference On Computer Design And Applications (ICCDA 2010),pp.306-313,2010.

[7] Wuwei, ”Research on Remote Sensing Image Classification Based on Neuro-Fuzzy and Texture Analysis”, Proceedings of the 29th Chinese Control Conference,pp.2900-2094, July 29-31, 2010, Beijing, China.