RESEARCH ARTICLE OPEN ACCESS

Landscape Pattern Mining using Machine Learning Intelligence

Velu.S [1], Baskar.K[2], Kumaresan.A[3], Suruthi.K[4]

[1,2,3,4] *Department of Computer Science and Engineering,* [1,2,3,4] *SKP Engineering College,*
[1,2,3,4] *Tiruvannamalai,* [1,2,3,4]*Tamil Nadu,* [1,2,3,4]*India*

Abstract:

The complexity of landscape pattern mining is well stated due to its non-linear spatial image formation and inhomogeneity of the satellite images. Land Ex tool of the literature work needs several seconds to answer input image pattern query. The time duration of content based image retrieval depends on input query complexity. This paper focuses on designing and implementing a training dataset to train NML (Neural network based Machine Learning) algorithm to reduce the search time to improve the result accuracy. The performance evolution of proposed NML CBIR (Content Based Image Retrieval) method will be used for comparison of satellite and natural images by means of increasing speed and accuracy.

Keywords: Spatial Image, Satellite image, NML, CBIR

I. INTRODUCTION

India is a nation whose backbone is considered to be agriculture since the nation has a huge amount of lands for cultivation along with enough rivers to provide proper irrigational facilities. But this is the scenario of the previous decade since the current technological revolution have taken over most of the agricultural lands and filled those lands with factories, apartments, IT industries and such other buildings leaving no area for agriculture. In addition to that, this have caused enough damage to the rivers and lake nearby these constructions by polluting them with the garbage outlets from these huge structures. Considering the need of agriculture and taking the advantage of satellite imagery, we can use these satellite images and gather various information about the landscape and the water resources in the nearby areas in order to decide whether the landscape can be used for either agricultural purposes or for construction works. Satellite images are to be segmented and classified [1]. A digital image is partitioned into multiple segments in order to represent the image in a different format facilitating the analysis and gathering of useful information from the image.

Classification of images involves differentiating the distinct landcover areas in an image where pixel classification is a problem [2]. Clustering is used for classification where n objects are classified into K distinct grouped based on a similarity metric [3]. In order to gather various information from the satellite images, we are in need of methodologies for image retrieval.

Image retrieval is the process of retrieving images from an enormous database based on the metadata added to the image which could be said as the annotations. But this annotations have some complexities. Manual annotation is a time consuming work to be done and the ambiguous annotation of images would result in a state where the user would never get the required results no matter the number of times he search the image database. Research is being done regarding several methods to be utilized for automatic image annotations due to the advancement in the field of semantic web and social web applications. In spite of the advancements, an effective methodology termed CBIR (Content Based Image Retrieval) is being used, in which feature extraction is basis. These features can be categorized into two, first one representingtext based features corresponding to keywords as well as annotations and the second one represents visual features corresponding to color, texture and faces along with shapes [4]. Since, features plays a major role here, when userinputs a query image, the pixel value of these images are compared with the pixel values of all the images prevailing in the dataset and the results given to the user would contain all the images containing a part of the image queried which is an effective way of avoiding annotations to avoid ambiguity.

II. Related Work:

In order to understand the process of retrieving data from images, we are in need of looking deep into the research done by various people till date.Herbert Daschiel et al have conducted a research on content-based remote sensing [5]. image information mining system is usedto provide users fast access to the content oflarge image databases. The system has done two major processes. The first process is computationally the off-line data ingestion in the archive, image feature extraction, and indexing. The second process consists of a graphical man–machine interface that manages the information fusion for interactive interpretation and the image information mining functions. Due to the complexity of images, the performance of the image retrieval can be varied in content based image retrieval. Bayesian classification is used to increase the accuracy in training data. Based on the stochastic nature of user-defined cover types, the system retrieves images using probabilistic measurements in order to find the performance of manmade machines. Finally, we compare the objective component of the evaluation protocol with the users' degree of satisfaction in order to point out the significance of the computed measurements.

J. Li et al have conducted a research on remote sensing, image retrieval systems queries which are done based on sensor, location, and date of image capture [6]. It is not about retrieving information from large imagedatabases. This paper follows the method of combination of retrieving spectral and spatial patterns from remotely sensed imagery using state-of-the-art data mining and advanced database technologies. Land cover information related to spectral characteristics is identified by classification based on support vector machines with automatic model selection while textural features characterizing spatial information are retrieved using the method of Gabor wavelet coefficients. In an object-oriented database with related images in an image database that produce identified land cover types, textural features are clustered and acquired in search-efficient space. The evaluation of the study results using coverage and novelty measures validates the effectiveness of the proposed remote sensing image information mining framework, which is potentially useful for applications such as agriculturaland environmental monitoring.C.-R. Shyu et al have searched for relevant knowledge across heterogeneous geo-spatial databases [7]. This processrequires an extensive knowledge of the semantic meaning of images, a keen eye for visual patterns, and efficient strategies for collecting and analyzing data with minimal human intervention. It is mainly focused on recently developed content-based multimodal Geospatial Information Retrieval and Indexing System. It includes automatic feature extraction, visual content mining from image databases and database indexing for fast retrieval of data. They have developed techniques for complex queries which merge information from different geospatial databases. The retrievals of objects is based on shape and visual characteristics along with analysis of multiobject relationships for the retrieval of objects in specific spatial configurations and semantic models to link low-level image features with high-level visual descriptors. Image analysts can rapidly identify relevant images. GeoIRIS is to find answer for the given query that have similar objects and spatial relationship that are within a certain time.

In this paper, L. Gueguen have handled the problem of constructing an index of compressed object databases [8]. This method introduces an informational similarity calculated based on the coding length of two part codes. Then, a methodology for compressing the database with respect to the inter object redundancies and by using the informational similarity. The z measure is presented. The method produces an index included in the code of the data volume. This index is built such that it contains the minimal sufficient information to discriminate the data-volume objects. The authors have presented an optimal two-part coder for compressing spatiotemporal events contained in Satellite Image Time Series (SITS). The two-part coder allows us to measure similarity and then to derive an optimal index of SITS spatiotemporal events. The resulting index is the representation of SITS information content in order to enable queries based on information content.

J. Jasiewicz et al have worked on Query-by-image-content (QBIC) tools which are in demand in geospatial community because they enable exploration and mining of the rapidly increasing database of remotely sensed images. Accompanying the growth of the imagery database leads to the increase in the number of image-derived products such as high-resolution large-spatial-extent maps of land cover/ land use(LCLU). QBIC-like tools for exploration and mining of such products would significantly enhance their value. In this paper, the authors have presented a method for retrieval of similar scenes from a category valued geospatial database of which an LCLU map is a particular example. The similarity between the two scenes is tantamount to similarity between their spatial patterns of class labels. This method works on the principle of query by example whose input is a reference scene and its output is a similarity map indicating a degree of alikeness between a location on the map and the reference. The two core components of the method are as follows: scene signature—an encapsulation of the scene pattern by means of probability distribution of class labels and the sizes of the patches that they form and scene similarity–a mutual-information-based function that assigns a level of similarity between any two scenes based on their signatures. The method is described in detail and applied to the National Land Cover Dataset 2006. Two examples of queries on this data set are presented and discussed. The applicability of the method to other data sets is also discussed in their work.

III. QUERY AND RETRIEVAL OF SPATIAL PATTERNS FROM SATELLITE IMAGES:

A wide range of data regarding the landscape could be extracted from remote sensing. Remote sensing is the process of gathering information about an object without being physically in contact with the object. This could be done by means of sensors with the help of electromagnetic signals emitted from either aircraft or satellites which is referred as active remote sensing whereas the application of signals emitted from sunlight is termed as passive remote sensing. Even though we have got a vast amount of information by means of remote sensing, we lack the machine search capabilities to gather the required information from the vast dataset. To solve this problem, Tomasz.F.Stepinski et al have come up with a tool named as Land Ex which is a real time Geoweb application for exploring and mining useful information regarding landscapes from large datasets [10]. This mining of information is done based on the implementation of pattern recognition algorithm on computerized maps. The authors have experimented this tool on satellite images available in NLCD (national land cover dataset) in United States. In this process, the user can input the required land cover pattern to be searched for after which the tool would consider the 2D class size histogram. The authors have implemented a higher level of optimization due to which the response time for the query would be very limited.

A.Calculation of Similarity

The authors have considered the comparison of similarity between two maps to be the core technology to be considered in their proposal. In existing measures, two maps are considered to be highly similar if same scene is available in both the maps viewed in the same perspective with minor changes in the assignment of corresponding cells. The existing scenarios would not find the similar contents available in two maps if they are rotated around a particular angle. This is due to the fact that the corresponding cells would not have similar values due to rotation. But we are in need of considering the overall spatial pattern of the maps eventhough the target map is rotated, translated or slightly deformated in pattern. Pattern oriented similarity is considered important in the case of landscape ecology [11] – [14]. Due to these demerits, the authors have used a method based on the concepts of CBIR. The main approach of CBIR is to find an image from an enlarged database considering only the visual information available in the image without considering the ambiguous annotations provided by humans. The comparison of images is made in CBIR based on the extracted feature vectors from the image representing the distinct characteristics of an image. The features to be retrieved from the image could be of any means but the extraction should be automatic [15]. This methodology would retrieve the entire set of queried results and the user is in need of segregating the results needed.

IV. IMPLEMENTATION OF MACHINE LEARNING CONCEPTS IN DIFFERENTIATING LANDSCAPES FROM SATELLITE IMAGES

We propose the use of machine learning technique in order to make the system segregate the areas such as water, lands that can be used for vegetation or building purposes with respect to the query given by the user. Training of the system is dome with the help of several images taken from the satellite image database based on neural networks. The first step involves the creation of a database with a set of images taken from the satellite image dataset. After the machine learning procedures done successfully, the system would be able to retrieve the results by segregating the land, mountain and water in addition to the details of the year, month and other relevant details. This information can be used for making decision based on different climatic conditions experiencedduring different parts of the year whether the area could be used for vegetation or building or any other purposes. Though the work of Tomasz F. Stepinski is good enough for retrieving information, that method would give bulk results for the query of the use from which user is in need of segregating the essential results. The year wise information also would not be available. But our proposed method has undergone learning to identify the different areas to be useful for multiple purposes which drastically reduces the workload of humans.

V. EXPERIMENTAL RESULTS:

The experiment is done with several satellite images taken from several repository websites available online. The first step is shown in figure 1showing the normal satellite image. The second step shown in figure2 showing the clustered index image. The third, fourth and last image that shown as figure 3, figure 4, figure 5 and figure 6 are separated images with color Id 1,2,3 and 4.

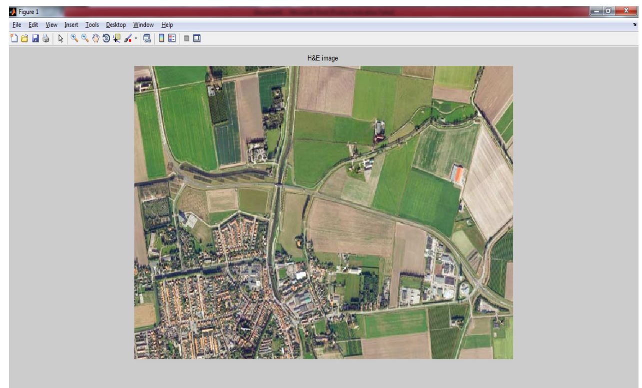

Fig1 A sample : Satellite Image

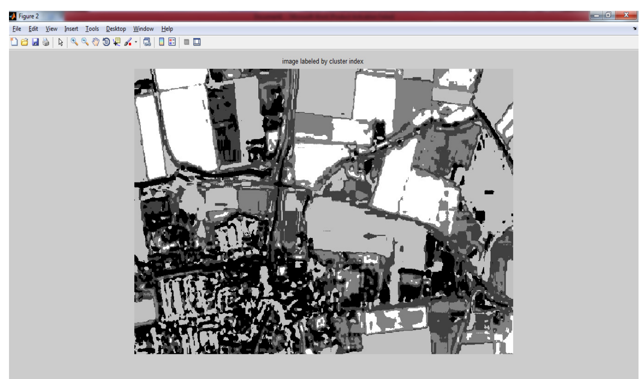

Fig2Clustered Index Image

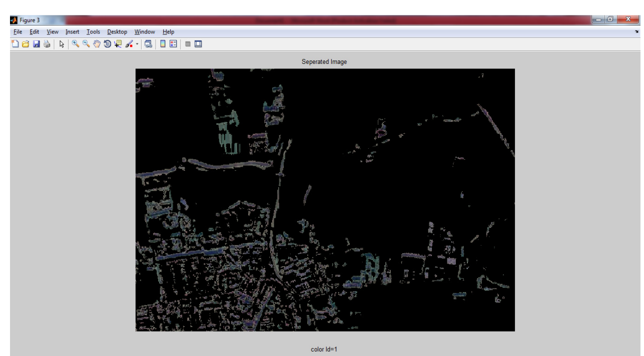

Fig 3Separated image(Color Id 1)

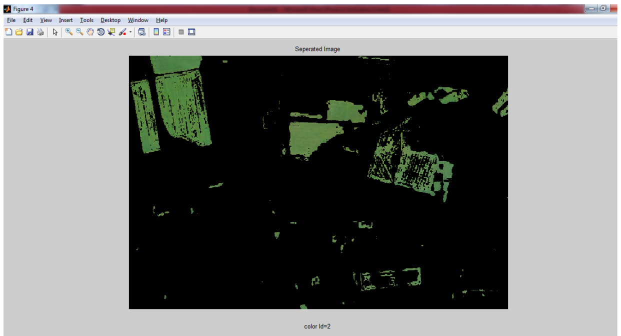

Fig 4Separated Image(Color Id 2)

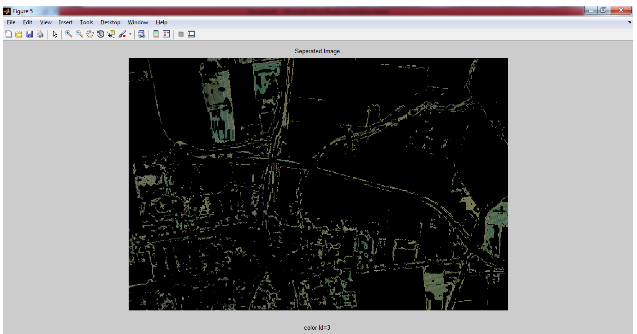

Fig 5 Separated Image (Color Id 3)

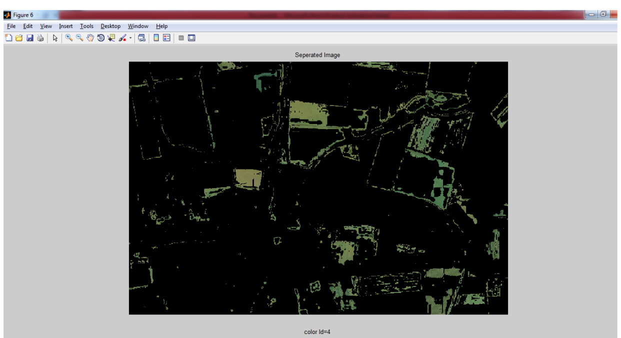

Fig 6Separated Image (Color Id 4)

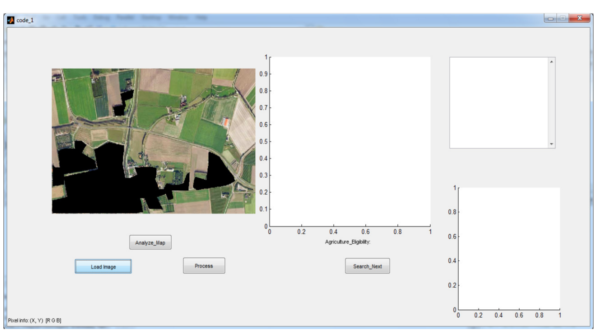

Fig 7. Loading the image for analysis

In figure.7, we have shown the process of loading a landscape image into the system for analysis. This could be done by clicking on the “Load image” push button available in the GUI shown in fig.6 After the image is loaded, the different areas in the landscape are respresented with different colors such as green represents the areas where greeneries are available such as crops, shrebs and bushes and trees. Blue color represents the area covered by water indicating the river, lake, pond or ocean fllowed by Brown color representing the area covered by soil. The area covered by the buildings is represented using black color since we are not considering them.

Figure 8shows the various analysis done in the input image followed by concluding whether the area can be used for agriculture or not. This involves several ateps to be followed. First, we ahould click on the “Analyze map” push button where a category acquisation graph would be generated showing the areas covered by land, water and tree cultivation levels with different colors. This step is followed by the comparison of rainfall in the area under observation during different seasons of the year based on the previous weather reports and other data available about that area. In addition to this, our system also determines the soil fertility which could be either PASS or FAIL based on the fertility and erosion factors of the soil. The system would match the factors observed from the area with varios pre-determined factors specified in order to determine whether the area can be used for cultivation. Due to this factor, sometimes, even though the soil fertility would be high but the area would not have agriculture eligibility since the pre-defined factors fit for agriculture would not have met.

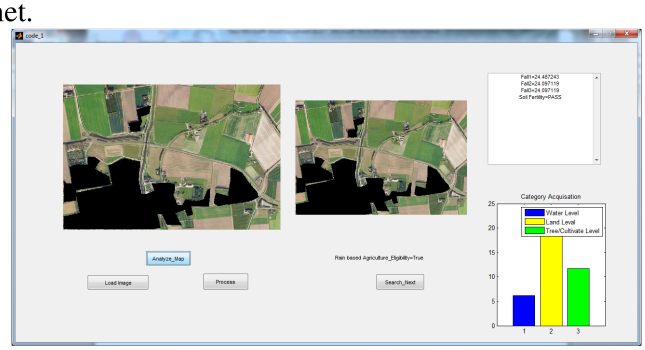

Fig8. Concluding whether the area is fit for cultivation or not

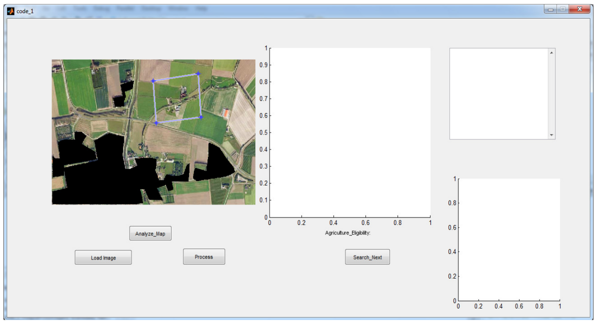

Fig 9. Cropping the desired area to be analyzed in detail

Figure 9 shows the cropping of the user specified area of interest to be analyzed in detail to determine whether the area is fit for ariculture or not. The user can intimate the system that he has finished cropping his area of interest by double clicking on the final point on the landscape image as shown in figure 9.

The system has analyzed the area apecified by the user and the results are shown in figure 10. The area is compared with the previous reports of the same area at different seasons and different years and the soil fertility and agriculture eligibility of the area are decided after comparing the prevailing conditions wih the specified conditions and the final decision is made.

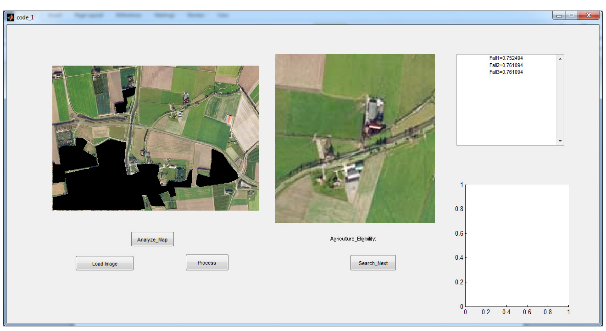

Fig10. Analysed result of the user specified area

VI. CONCLUSION:

In this paper, we have proposed a methodology based on CBIR for segregating the water, land and vegetation from the satellite images for the purpose of segregating land and water covered areas for making a decision whether the area could be used for agriculture or other purposes. This segregation is done by the system after undergoing machine learning technique based on neural networks. Experimental results have shown that our method can reduce the human workloadin terms of segregating areas from bulk results.

REFERENCE:

1. *P. Sathya and L. Malathi , "Classification and Segmentation in Satellite Imagery Using Back Propagation Algorithm of ANN and K-MeansAlgorithm" in International Journal of Machine Learning and Computing, Vol. 1, No. 4, October 2011.*

2. *U.Maulik and S. Bandyopadhyay, "Fuzzy partitioning using a realcoded variable-length genetic algorithm for pixel classification," IEEE Trans. Geoscience and Remote Sensing. vol. 41, no. 5, pp. 1075–1081, May 2003.*

3. *U.Maulik and S. Bandyopadhyay, "Performance evaluation of some clustering algorithms and validity indices," IEEE Transaction on Pattern Analysis Machine Intelligence., vol. 24, no. 12, pp. 1650–1654, Dec. 2002. [4] Yong Rui and Thomas S. Huang, "Image Retrieval: Current Techniques, Promising Directions, and Open Issues" in Journal of Visual Communication and Image Representation **10**, 39–62 (1999)*

4. *Yong Rui and Thomas S. Huang, "Image Retrieval: Current Techniques, Promising Directions, and Open Issues" in Journal of Visual Communication and Image Representation **10**, 39–62 (1999)*

5. *Herbert Daschiel and Mihai Datcu "Information Mining in Remote Sensing Image Archives: System Evaluation"in IEEE Transactions on Geoscience and Remote Sensing, vol. 43, no. 1, january 2005*

6. *J. Li and R. M. Narayanan, "Integrated spectral and spatial informationMining in remote sensing imagery," in IEEE Transactions on Geoscience and Remote Sensing,vol. 42, no. 3, pp. 673–685, March 2004.*

7. *C.-R. Shyu, M. Klaric, G. Scott, A. S. Barb, C. Davis, and K. Palaniappan,"GeoIRIS: Geospatial information retrieval and indexing system Content mining, semantics modeling, and complex queries," IEEETrans. Geosci. Remote Sens., vol. 45, no. 4, pp. 839–852, Apr. 2007.*

8. *L. Gueguen and M. Datcu, "A similarity metric for retrieval of compressed objects: Application to mining sattelite images time series," IEEE Trans. Knowl. Data Eng., vol. 20, no. 4, pp. 562–575, Dec. 2008.*

9. *J. Jasiewicz and T. F. Stepinski, "Example-based retrieval of alike landcover scenes from NLCD2006 database," IEEE Geosci. Remote Sens.Lett., vol. 10, no. 1, pp. 155–159, Jan. 2013.*

10. *Tomasz F. Stepinski, Pawel Netzel, and Jaroslaw Jasiewicz , "LandEx—A GeoWeb Tool for Query and Retrieval of Spatial Patterns in Land Cover Datasets" in IEEE Journal Of Selected Topics In Applied Earth Observations And Remote Sensing, Vol. 7,No. 1, January 2014.*

11. *R. V. O'Neill, J. R. Krummel, R. H. Gardner, G. Sugihara, B. Jackson,D. L.DeAngelis,B. T. Milne,M.G. Turner,B. Zygmunt, S. W. Christensen,V. H. Dale, and R. L. Graham, "Indices of landscape pattern,"Landscape Ecol., vol. 1, no. 3, pp. 153–162, 1988.*

12. *J. D. Wickham and D. J. Norton, "Mapping and analyzing landscape patterns," Landscape Ecol., vol. 9, no. 1, pp. 7–23, 1994.*

13. *T. R. Allen and S. J. Walsh, "Spatial and compositional pattern of alpine treeline, Glacier National Park, Montana," Photogramm. Eng.Remote Sens., vol. 62, no. 11, pp. 1261–1268, 1996.*

14. *D. H. Cain, K. Riitters, and K. Orvis, "A multi-scale analysis of landscape statistics," Landscape Ecol., vol. 12, pp. 199–212, 1997.*

15. *Shu-Ching Chen, Stuart H. Rubin, Mei-Ling Shyu, and Chengcui Zhang," A Dynamic User Concept Pattern Learning Framework for Content-Based Image Retrieval" in 772 IEEE Transactions On Systems, Man, And Cybernetics—Part C: Applications And Reviews, Vol. 36, No. 6, November 2006*

.