Volume 3, Issue 8, August 2013 ISSN: 2277 128X

International Journal of Advanced Research in Computer Science and Software Engineering

Research Paper

Available online at: www.ijarcsse.com

A Survey on Knowledge Discovery from the Satellite Image Using Association Rule Mining

Jay Narayan Thakre
M. Tech. VIth Semester
Computer Science & Engineering
BUIT, Bhopal (M.P.), India

Divakar Singh
Head of Department
Computer Science & Engineering
BUIT, Bhopal(M.P.), India

Abstract— An image mining is a technique used to mining the knowledge extracted from the image. The satellite images contain an important information for weather forecasting and early prediction of different atmospheric disturbance such as typhoon, hurricanes etc. These Features can be extracted by Content Based Image Retrieval (CBIR). There are several work has been done on image mining and inter transaction association rule which stores the images in the database and retrieves the images by a query image. They used CBIR to retrieve the images and extracts feature from the image. They extracted cloud laver, high pressure, etc. from image and then association rule is applied. Three types of Satellite Images Contain more parameter which can be extracted. Another parameter like humidity extracted from Water Vapor Image with previous parameters like cloud layer, linear cloud, typhoon can be extracted to from a satellite images to get a proper or efficient knowledge and association rule is applied to discover knowledge. New parameter is included so that association rule mining can give more accurate and efficient result. This uses low level feature to extraction feature from satellite image and discovering knowledge from this feature.

Keywords—Satellite Image, CBIR, Knowledge Discovery, Feature Extraction, Association Rule

I. INTRODUCTION

Image mining is extraction of implicit knowledge, image data relationship or other pattern not explicitly stored in images and uses ideas from computer vision, image processing, image retrieval, data mining and machine learning database [2]. Challenge in image mining is to identify low level feature containing in a pixel or group of pixels in an image, or image can be effectively and efficiently processed to identify high level spatial object and relationships. Image mining process involves preprocessing, transformation, feature extraction, Mining, Evaluation and interpretation and obtaining the final knowledge.

Association rule mining is one of the most promising techniques in image mining. There is image system to store weather forecasting images and retrieve them later for research to predict future temperature, relative humidity, rainfall, wind speed and atmospheric pressure.

Image retrieval is based on content based image retrieval. If image volume is large, the content based image retrieval is not so effective. These strategies usually use low level features such as texture, color and shape to calculate the similarity between images. Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem. Content based means that the search will analyze the actual contents of the image rather than the metadata such as keywords, tags, and/or descriptions associated with the image. The term 'content' in this context might refer to shapes, colors, textures or any other information that can be derived from the image itself. CBIR is desirable because most web based image search engines rely purely on metadata and this produces a lot of garbage in the results.

Traditional DBMS does not work well for image data due to the lack of semantic information in the data. In recent years, automatic indexing and retrieval based on image content has become more desirable for developing large volume image retrieval applications.

Color, shape and texture are the main features both humans and computers used to recognize images. Several systems have been proposed in the research community for content-based information retrieval such as QBIC (Query by Image Content) by IBM, and Visual SEEK by Columbia University. Most content-based image retrieval techniques employ the following two steps to retrieve the images. First, each image's feature vector is computed and then stored in the image database. Then secondly, given a query image, the feature vector of the query image is computed and then is taken to be compared with the feature vector of each image stored in the image database. A certain image's feature vector that is close to the feature vector of query image is returned to the user.

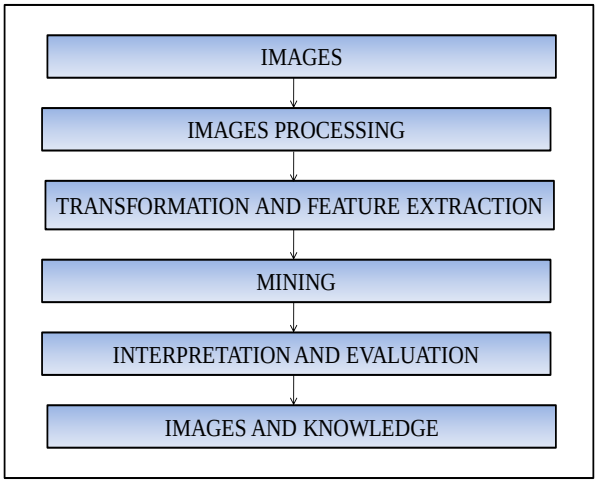

Figure 1 Image Mining System

II. Satellite Image

Satellite Images are of three types:

A. Visible Image

Visible satellite pictures can only be viewed during the day, since clouds reflect the light from the sun. On these images, clouds show up as white, the ground is normally grey, and water is dark. In winter, snow-covered ground will be white, which can make distinguishing clouds more difficult. To help differentiate between clouds and snow, looping pictures can be helpful; clouds will move while the snow won't. Snow-covered ground can also be identified by looking for terrain features, such as rivers or lakes. Rivers will remain dark in the imagery as long as they are not frozen. If the rivers are not visible, they are probably covered with clouds. Visible imagery is also very useful for seeing thunderstorm clouds building. Satellite will see the developing thunderstorms in their earliest stages, before they are detected on radar.

B. Infra-Red Image

Infrared satellite pictures show clouds in both day and night. Instead of using sunlight to reflect off of clouds, the clouds are identified by satellite sensors that measure heat radiating off of them. The sensors also measure heat radiating off the surface of the earth. Clouds will be colder than land and water, so they are easily identified. Infrared imagery is useful for determining thunderstorm intensity. Strong to severe thunderstorms will normally have very cold tops. Infrared imagery can also be used for identifying fog and low clouds. The fog product combines two different infrared channels to see fog and low clouds at night, which show up as dark areas on the imagery.

C. Water Vapor Image

Water vapor satellite pictures indicate how much moisture is present in the upper atmosphere (approximately from 15,000 ft to 30,000 ft). The highest humanities will be the whitest areas while dry regions will be dark. Water vapor imagery is useful for indicating where heavy rain is possible. Thunderstorms can also erupt under the high moisture plumes.

III. Literature Review

Using fuzzy SOM strategy for satellite image retrieval and information mining projected by yo-ping hung, tsun-wei and li-jen kao. They proposed a model for efficient satellite image retrieval and knowledge discovery. It has two major parts. First, it uses computation algorithm for off-line satellite image feature extraction, image data representation and image retrieval. Important parameter can be extracted from the satellite image by the CBIR (content based image retrieval) technique to discover knowledge about the current whether condition. The extracted features are high pressure, cloud layer, linear cloud and typhoon. A dataset is created by these parameters to apply the association rule. A self organization feature is used to construct a two layer satellite image concept hierarchy. The events are stored in one layer and the corresponding feature vectors are categorized in the other layer. Second, a user friendly interface is developed that retrieves images of interest and mines useful information based on the event in concept hierarchy [1].

A data mining approach for monsoon predication using satellite image data predicts the monsoon on the basis of some parameters which are Sea Surface Temperature (SST), Cloud Top Temperature (CTT), Cloud Density, and Water Vapour or Humidity. The infra Red (IR) spectrum sensor measures the different [2].

A semantics-based approach was proposed to classify satellite images according to the corresponding heterogeneous features. This approach is able to detect multiple semantics classes within one satellite image using a combination of a sliding window and interpolation. The experimental results have shown that the trained semantics classifiers together with the interpolation approach achieved an effective and efficient identification of the semantics classes within the satellite scenes.

As a result not just only the query-by-example approach but also the query-by-terms technique is supported. Both resulted in very satisfactory retrieval results – even for cross-scanner queries, i.e. queries that have to retrieve imagery from one scanner although the semantics was obtained through data from a different scanner. Hence, a high degree of independence was achieved [4].

Satellite Cloud Image Processing and Information Retrieval System proposed a model for the satellite cloud processing and retrieval system. The database construction part is intended to ensure high retrieval efficiency by extracting a feature set for each of the image in database at loading time and storing the feature set along with its corresponding image in the database so that when a query image is presented to the system, the system does not have to perform feature extraction on each database image. A content based image retrieval (CBIR) system has been developed using color, texture and shape as retrieval features from the satellite image database. To extract the grey level/color properties of an image, histogram values have been used. Four function of texture feature have used such as (entropy, energy, correlation and contrast) and shape features (area, perimeter and metrics) have been extracted using the morphological operations [11].

A system is build to store weather forecasting images and retrieve them for weather forecasting. They also used CBIR to extract the feature and retrieve images. Association rule mining is applied in the search for weather forecasting [4].

IV. Image Retrieval and Feature Extraction

An Image Retrieval and Feature Extraction is done by the following techniques.

A. Content Based Image Retrieval

Content based image retrieval (CBIR) is a technique of retrieving images from a large dataset of image. Content based image retrieval is based on the low level visual features of the images. These features are color, texture and shape [11]. Similarities between the images are calculated using the Euclidean distance [12].

$$|Q,T| = \sum |\omega i - ti| \qquad (1)$$

Where Q is query image and qi is low level feature of Q. T is a certain image in database and ti is low level feature of T. ωi is the weight factor [4] [12].

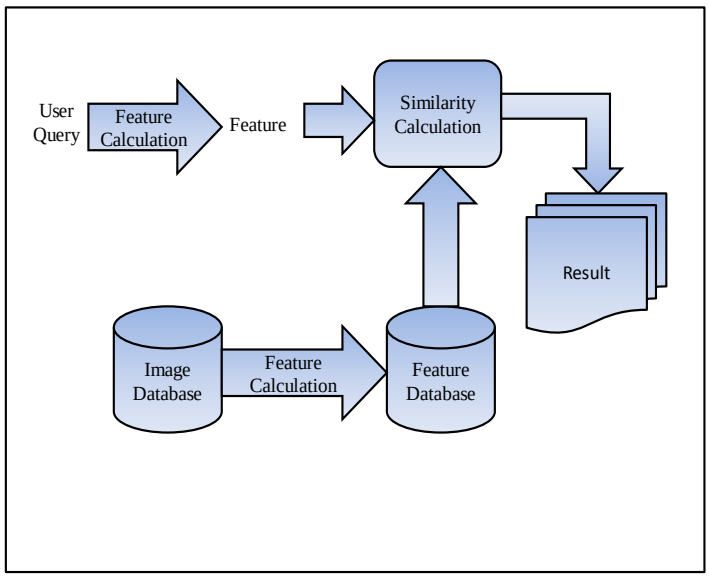

Fig 2: CBIR Process

Figure 2 represents the complete working of Content based image retrieval (CBIR). Image database has already stored an images and their feature along with image ID stored in feature database. Initially user the image retrieval process provides a query image as input, and then the system starts by extracting the feature from the queried image. Afterwards, the system measures the similarity between the feature set of the query image and those of the image stored in the database. The system ranks the relevance based on the similarity and returns the result.

B. Feature Extraction

1) Grey level/Color Feature Extraction: Color property is one of the most widely used visual features in Content Based Image Retrieval (CBIR) systems. Researches in this field can be grouped into three main subareas: (a) definition of adequate

color space for a given target application, (b) proposal of appropriate extraction algorithms, and (c) study/evaluation of similarity measures.

Color information is represented as points in three-dimensional color spaces (such as RGB, HSV, YIQ, L∗ u∗ v∗ , L∗ a∗ b∗ [17]). They allow discrimination between color stimuli and permit similarity judgment and identification [17]. Some of them are hardware-oriented (e.g., RGB, and CMY color space), as they were defined by taking into account properties of the devices used to reproduce colors. Others are user-inspired (e.g., L∗ u∗ v∗ , L∗ a∗ b∗) as they were defined to quantify color differences as perceived by humans.

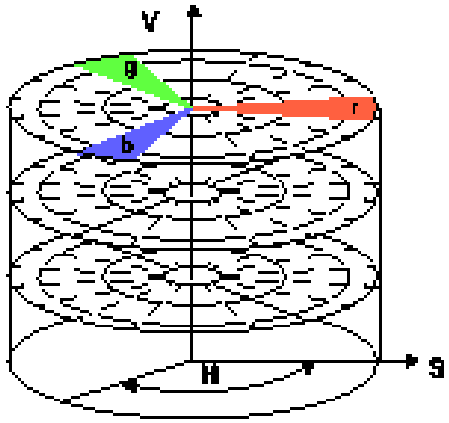

Fig 3: The HSV color space

Examples of descriptors that do not include spatial color distribution include *Color Histogram* and *Color Moments*. The color histogram extraction algorithm can be divided into three steps: partition of the color space into cells, association of each cell to a histogram bin, and counting of the number of image pixels of each cell and storing this count in the corresponding histogram bin. This descriptor is invariant to translation and rotation. The similarity between two color histograms can be performed by computing the L1, L2, or weighted Euclidean distances, as well as by computing their intersection.

Other example of descriptor that does not consider color spatial distribution are the so called **Color Moments** [21]. Usually, the *mean* (first order), *variance* (second), and *skewness* (third) are used to form the feature vector. These moments are defined, respectively, as

$$E_i = \frac{1}{N}\sum_{j=1}^{N} p_{ij}, \sigma_i = \sqrt[2]{\left(\frac{1}{N}\right)\sum_{j=1}^{N}(p_{ij} - E_i)^2} \qquad (2)$$

$$s_i = \sqrt[3]{\left(\frac{1}{N}\right)\sum_{j=1}^{N}(p_{ij} - E_i)^3} \qquad (3)$$

where pij is the value of the i-th color component of the image pixel j, and N is the number of the pixels in the image.

2) Texture Feature Extraction: There are varieties of techniques which are used for measuring texture such as co-occurrence matrix, fractals, gabor filters, and variation of wavelet transform.

Co-occurrence matrix is one the most traditional techniques for encoding texture information. It describes spatial relationships among grey-levels in a image. A cell defined by the position (i, j) in this matrix registers the probability at which two pixels of gray levels i and j occur in two relative positions. A set of co-occurrence probabilities (such as, energy, entropy, contrast) has been proposed to characterize textured regions.

Normalized probability density P(i, j) of the co-occurrence Matrices can be defined as follows-

$$\mathbf{P\ (i,\ j)} = \frac{\#\{(x,y),(x+d,x+y)|S|f(x,y)=i,f(x+d,y+d)=j\}}{\#S} \qquad (4)$$

Where,

x, y=0,1,……N-1 are co-ordinates of the pixel

i, j=0,1,…….L-1 are the gray levels

S is the set of pixel which have certain relationship in the image.

#S is the number of elements in S.

P (i, j) is the probability density that the first pixel has intensity value i and the second j, which separated by distance δ= (dx, dy) [6].

3) Shape Feature Extraction: In pattern recognition and related areas, shape is an important characteristic to identify and distinguish objects. Shape descriptors are classified into *boundary-based (or contour-based)* and *region based methods* [18].

This classification takes into account whether shape features are extracted from the contour only or from the whole shape region. These two classes, in turn, can be divided into *structural (local)* and *global descriptors*. This subdivision is based on whether the shape is represented as a whole or represented by segments/sections. Another possible classification categorizes shape description methods into *spatial* and *transform* domain techniques, depending on whether direct measurements of the shape are used or a transformation is applied.

Shape is an important characteristic of an object. The goal of shape descriptors is to uniquely characterize the object shape. Two shape descriptors are used in our experiments [9]:

- Eccentricity is the length ratio between the major and minor axes of the objects, smaller for rounded shapes and greater for distorting ones.

Compactness is the ratio between the length of object's boundary and the object's area.

The spatial information of the region is also considered in the annotation of images. It provides the necessary information in the process of region indexing and semantic definition, like upper left, upper right, center, etc. The spatial information of each region is represented by two parameters, as in Figure 4: the centroid of the region Cx,y=(Xc, Yc) and the minimum bounding rectangle (l,r,t,b), where (l,r) represents the coordinates of the upper left corner and (t,b) represents the coordinates of the bottom right corner.

A region is described in conformity with defined characteristics:

- The color characteristic is represented in the HSV color space quantized at 166 colors. A region is represented by a color index which is, in fact, an integer number between 0...165.
- The spatial coherency represents the region descriptor, which measures the spatial compactness of the pixels of the same color. It represents the spatial homogeneity of a region in an image and is computed for identifying the 8-connected pixels of the same color in a region.
- A seven-dimension vector (maximum probability, energy, entropy, contrast, cluster shade, cluster prominence, correlation) represents the texture characteristic.
- The region dimension descriptor represents the number of pixels from region.
- The spatial information is represented by the centroid coordinates of the region and by minimum bounding rectangle
- A two-dimensional vector (eccentricity and compactness) represents the shape feature.

A descriptor composed by color, texture, dimension, spatial coherency, shape, position is associated to each color region.

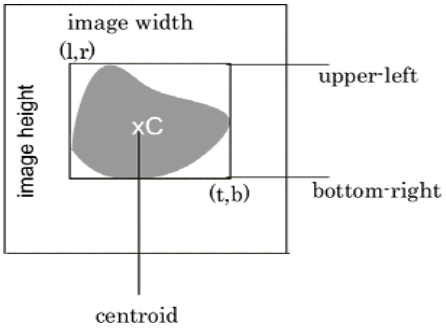

Fig 4: The representation of spatial information of image region

V. Knowledge Discovery by Association Rule

The knowledge discovery in databases, an important part of data mining, is defined as the automated discovery of useful, unknown, non-trivial information [10].

The main component in image discovery is the identification of similar objects from images.

The association rules discover the information about the elements that are frequent. The formal representation of an association rule is the following. Being given I= {i1, i2,.,, im} a set of distinct elements, D is called a transaction set, where each transaction T is a subset of I or TI. A transaction T contains X if and only if X⊆TA and B are called the body, and respectively the head of the rule. The support of a rule is defined as the percent of transactions in which both the body (A) and head (B) of the rule are present. If a rule A=>B has support %s, it means that %s of transactions from D contain A⊆B. The confidence is defined as the ratio between the number of transactions in which both the body (A) and head (B) of the rule are present, and the number of rules in which only the body is present [13].

- If the rule A=>B has the confidence %c, it means that %c transactions from D contain A∩B.
- The support estimates the probability P(A⊆B), and the confidence estimates the conditional probability P(B|A).
- In image discovery, to label each object, which appears in an image, is a complex task.

In this paper, the knowledge mining from images is used for the definition of rules, which convert the low-level primitives of images into semantic high level concepts. The methods used in this study bring important improvements related to the detailed descriptions of images, which are necessary for defining relationships between:

- Objects/regions,
- Classes of visual characteristics,
- Objects/regions and classes of visual characteristics.

VI. Proposed Work

In previous study most of the work used a database of image and then an image is given as query to retrieve the feature matched images. The features are extracted from the query image and database image and features of both the image are matched. On the basis of this matching image retrieval is done.

In the proposed work there three types of Satellite Images taken as an input. On the basis of their color, texture, and shape the features will be extracted. Three types of Satellite Images are Visible Satellite Image, Infra-Red Satellite Image, and Water Vapor Satellite Image. All this images contains unique information. This information will be extracted in the proposed work with the help of CBIR method as discussed above. The Satellite Images may contain High pressure, Cloud Layer, Snow, Typhoon etc.

Once the feature will have been extracted then these features will be stored in a table and that will be called transaction table for association mining. To discover the knowledge about the weather, association rule is applied to the transition table, created by feature extracted from the Satellite Images.

VII. Conclusion

The proposed system will take some images as input and by using CBIR Features will be extracted. On these feature inter transaction association rule is applied to discover the knowledge. There is no need to store image in database and there in no query image is applied.

REFERENCES

[1] Yo-Ping Huang, Tsun-Wei Chang and Li-Jen Kao "Using Fuzzy SOM Strategy for Satellite Image Retrieval and Information Mining" CITSA 2007, Florida, USA. ISBN 1-934272-10-8.
[2] Dinu John, Dr. B.B. Meshram "A Data Mining Approach for Monsoon Predication using Satellite image data" International Journal of Computer Science & Communication Networks, Vol 2(3), 421-424 2012.
[3] Y. Li, T. Bretschneider "Semantics-Based Satellite Image Retrieval Using Low-Level Features" IEEE 2004.
[4] Senduru Srinivasulu, P. Sakthivel "Extracting Spatial Semantics in Association Rules For Weather Forecasting Image" IEEE 2010.
[5] Image Mining: Trends and Developments http://www.comp.nus.edu.sg/~whsu/publication/2002/JIIS.pdf.
[6] Mahendra Kumar Gurve, Jyoti Sarup "Satellite Clout Image Processing And Information Retrieval System" IEEE 2012.
[7] J. R. Smith, S.-F. Chang. VisualSEEk: a fully automated content-based image query system. Proceedingsof the Fourth ACM International Multimedia Conference and Exhibition, Boston, MA, USA, 1996, 87-98.
[8] R.M. Haralick, K. Shanmugam, I. Dinstein. Textural features for image classification. IEEE Transactions on Systems, Man, and Cybernetics, Vol.3, No. 6, 1973, 610–621.
[9] D. Zhang. Image retrieval based on shape. PhD Thesis, Monash University, March, 2002.
[10] W.J. Frawley, G. Piatetsky-Shapiro, C.J. Matheus. Knowledge Discovery in Databases: an overview. Knowledge Discovery in Databases, MIT Press, 1991, 1–27.
[11] Mahendra Kumar Gurve, Jyoti Sarup. "Satellite Cloud Image Processing And Information Retrieval System" 978-1-4673-4805-8 IEEE 2012.
[12] Chaur-Chin Chen and Hsueh-Ting Chu, "Similarity Measurement between images" 2005.
[13] Data Mining Techniques by Arun K Pujari.
[14] Yo-Ping Huang and Tsun-Wei Chang, "A Fuzzy Inference Model for Image Segmentation", 0-7803-7810-5/03/ IEEE 2003.
[15] Getachew Berhan, Tsegaye Tadesse, Solomon Atnafu, Shawndra Hill, "Drought Information Mining from Satellite Images for Improved Climate Mitigation" 987-1-4673-4805-8/12 IEEE 2012.
[16] Yikun Li and Timo R. Bretschneider, "Semantic-Sensitive Satellite Image Retrieval" 0196-2892 IEEE 2007.
[17] A. del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann Publishers, San Francisco, CA, USA, 1999.
[18] D. Zhang and G. Lu. Review of Shape Representation and Description. *Pattern Recognition*, 37(1):1–19, Jan 2004.

