# Image Mining Using Texture and Shape Feature

Prof.Rupali Sawant
Sardar Patel Institute of technology

**ABSTRACT**

Discovering knowledge from data stored in typical alphanumeric databases, such as relational databases, has been the focal point of most of the work in database mining. However, with advances in secondary and tertiary storage capacity, coupled with a relatively low storage cost, more and more non standard data (in the form of images) is being accumulated. This vast collection of image data can also be mined to discover new and valuable knowledge. During the process of image mining, the concepts in different hierarchies and their relationships are extracted from different hierarchies and granularities, and association rule mining and concept clustering are consequently implemented. The generalization and specialization of concepts are realized in different hierarchies, lower layer concepts can be upgraded to upper layer concepts, and upper layer concepts guide the extraction of lower layer concepts. It is a process from image data to image information, from image information to image knowledge, from lower layer concepts to upper layer concepts. In this paper framework of image mining based on concept lattice and cloud model theory is proposed. The methods of image mining from image texture and shape features are introduced here, which include the following basic steps: firstly pre-process images secondly use cloud model to extract concepts, lastly use concept lattice to extract a series of image knowledge.

**Keywords:** Image mining, texture feature, shape feature image retrieval, image clustering, Concept lattice, Cloud model

## 1. Introduction

Image mining can be define as the nontrivial process to discover valid, novel, potentially useful, and ultimately understandable knowledge from large image sets or image databases. Image mining is not only a branch of data mining and knowledge discovery, but also an interdisciplinary research area which includes digital image processing, image understanding, database, artificial intelligence and so on. In order to use the traditional data mining methods from rational data and database in image mining area, various content features extract the from the images. In the real world, many feature dimensions may be irrelevant or redundant to the image-mining task. In fact, many different research fields are affected by it. The curse of the dimensionality is especially serious in the case of data mining and image mining. If we want to mine the images efficiently and effectively using traditional data mining methods, we must solve this problem firstly, that is to say, we have to reduce the image feature dimensionality. Dimension reduction can be seen as the transformation from a high order dimension to a low order dimension.

Concept is a kind of thinking form in mind which reflects the attributes of objects. The common essential attributes of the perceptual objects are abstracted, and then the concepts are formed. Concept is composed of intension and extension, and will be changed along with the development of subjective and objective worlds. Concept formation is an important character of brain learning, and it is considered as an effective approach of analogy between data mining and concept formation of brain. If a kind of mathematic formal data structure is created to express intension and extension of concepts, including the abstract relationships among concepts in different hierarchies, the process of data mining and knowledge discovery will be analyzed efficiently and formally.

The cloud model is an effective tool in transforming between qualitative concepts and their quantitative expressions. Cloud model is used to calculate support, confidence and relationship in the field of association rules mining. The hasse diagram of concept lattice reveals the concept hierarchy of the context. It also shows the generalization /specialization relationship between the concepts corresponding to the subset relationship between the property and object sets. Therefore the graph can be used to produce hierarchies and association rules, which are consistent with specialization. Cloud models is adopted to control the generalization of a set of qualitative attributes, and a new constructing algorithm of concept lattice based on cloud models is presented for mining association rules in large databases, which is integrating attribute oriented generalization and concept lattice. Basically a large number of rules are directly mined from relational databases, and most of such rules are unnecessary. However, mining information in high hierarchies will generate some rules that are interesting and useful.

This paper present algorithm of cloud model and concept lattice for mining large collection of image data and hence retrieval of image based on texture and shape feature.

The paper is organized as follows. In Section II, the background and related concept lattice are introduced. Section III describes cloud model and clustering methods. Design of basic framework is introduced in section IV. Image Mining Based on Concept Analysis described in section V, and experiments analysis are discuss in section VI. Finally, conclusions are drawn in the final section.

## II. Concept Formation

Concept lattice is proposed by Wille R. [3]. It reflects the process of human's concept formation with mathematical formal language. Based on binary relationship, concept lattice embodies the unification of intension and extension of concepts, reflects the relationships between objects and characteristics and the relationships between generalization and characterization among concepts. With corresponding Hasse graph, concept lattice can implement the visualization of the hierarchies of data concepts, and it is suit to find the latent concepts from image data [4].

### A. Concept Lattice

It can be defined as;

Given context which is a triple T= (O, D, R), where O is object set, D is attribute set, R is the binary relations between O and D, only one partial ordered set corresponds with R, a kind of lattice structure L will be created, which is called concept lattice. Each node of lattice L is an ordered pair, which is called concept (X, Y); here X ε P (O) is the extension of the concept, Y ε P (D) is the intension of the concept, P (O) stands for power set of object, P (D) stands for power set of attribute set. Each ordered pair is self-contained on relation R. The partial ordered relationships could be built among the nodes of concept lattice. Given H1=(X1, Y1) and H2=(X2, Y2), if H1< H2 ↔ Y1Y2, it means that H1 is the direct generalization of father node of H2. Concept lattice is the representation of concept hierarchy, the Hasse graph could be produced from partial ordered relationships, if H1<H2 and no existing H3 satisfying H1<H3<H2, there exists one edge from H1 to H2. With corresponding Hasse graph, concept lattice can implement the visualization of the hierarchies of data concepts.

Definition 1. A formal context is a triple: K=(O,A,R) ,where O and A are two sets, and R is a relation between O and a.

O = {o1,,om},each oi (i≤ n) is called an object. A = {a1, am), each aj (j ≤ m) is called an attribute. In a formal context K = (O,A, R) ,if (o,a) Є R, we say that the attribute a is an attribute of the object x, or that x verifies a.(o, a) Є R is denoted by 1 ,and (o,a) ∉ R is denoted by 0.Thus ,a formal context can be represented by a matrix only with 0 and 1.

Table 1. A Formal Context Of An Image

| | a | b | c | d | e | f | g | h |
|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
| 2 | 1 | 1 | 0 | 1 | 0 | 0 | 1 | 0 |
| 3 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 |
| 4 | 1 | 0 | 0 | 1 | 1 | 0 | 1 | 0 |
| 5 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 |
| 6 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 7 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

The matrix of the Formal context represented in table 1 is given

$$
\begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array}
\begin{array}{c}
\begin{array}{cccccccc} a & b & c & d & e & f & g & h \end{array} \\
\begin{pmatrix}
1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix}
\end{array}
$$

Fig.1.The Matrix of the Formal Context

The context in Table 1 has 16 concepts. The line diagram in fig 2 represents the concept lattice of fig.1context

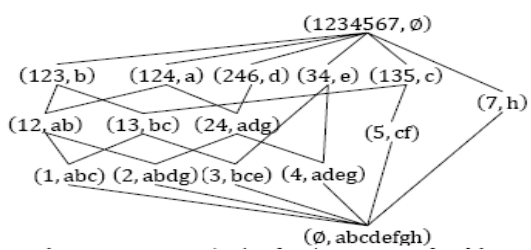



Fig.2 Concept Lattice Of Context

The set of all lower neighbors of a given concept is a subset of the set consisting of all sub concepts of it.

The algorithm for concept lattice is outlined as follows.

1. Create the top node and bottom node of concept lattice.
2. Scan dataset D, build concept lattice, and find the intensions of all frequent concepts.
3. Produce all frequent items with the intensions produced in Step2.
4. For each frequent item in Step3, generate association rules which confidence value is no less than min_conf.

## III. CLOUD MODEL

Cloud model is a kind of uncertain transformation model between qualitative concepts and quantitative numerical values; it has been applied into the fields of digital watermark, network in break, data mining, and image segmentation etc. [6]. Cloud model depicts concepts' randomness, fuzziness, and their association. The methods of forward normal cloud model, cloud transformation, concept synthesis of cloud model theory can be used to implement concept ascending and induction in different granularity worlds.

Definition: If U is a quantity domain expressed with accurate numerical values, and C is a quality concept in U, if the quantity value x ε U, and x is a random realization of concept C, μ (x) is the membership degree of x to C,μ (x) ε [0,1] , it is a random number which has a steady tendency: μ :U→ [0,1] , V x ε U , x→μ (x) . The distribution of x in domain is called cloud; each x is called a cloud drop.

### *A. The Numerical Characteristics of Cloud Model*

The numerical characteristics of cloud model are expressed with Expectation Ex, Entropy En and Super-entropy He, and they reflect the whole characteristics of the quality conception C. Expectation Ex of the Cloud drops' distribution in domain, is the point which can best represent the quality concept, reflect the cloud centre of gravity of cloud drops of the concept [3]. Entropy En is the uncertainty measurement of the quality concept, is decided by the random and fuzziness of the concept, it reflects the connection between the fuzziness and the random. Entropy En is a random measurement of the quality concept, reflects the discrete degree which can represent the quality concept; in another aspect, it is the measurement of fuzziness, and it reflects the value range which can be accepted by the concept of the cloud drop [3]. Use Entropy En the same numeric characteristic to reflect fuzziness and random, and it embodies the connection between each other. The super-entropy He is the uncertain measurement of entropy, namely the entropy of the entropy. It reflects the coagulation of uncertainty of all points which representing the concept in the number domain, namely the coagulation degree of cloud drop.

The algorithm for uncertain concept ascending based on cloud model is outlined as follows.

1) Calculate the frequency distribution function f(x) of List
2) Set NULL as cloud concept, h(x) = f(x), While (peak value of h(x) > ε)
3) Set the peak value of h(x) as Expectation Exi of cloud concept
4) Calculate Entropy Eni of cloud model which is fit to h(x) and its expectation value is equal to Exi
5) Calculate Hyper-entropy Hei
6) Find Clouds = Clouds_{C (Exi, Eni, Hei)}

### *B. Feature Clustering*

Han and Kamber [5] define the clustering as a process of grouping data into classes or clusters. The objects in a cluster have high similarity in comparison to each other, but they are dissimilar to the objects in other clusters. The clustering procedure is developed for the images stored in the image database. The clustering is based on the image content that is described using features. Hence, the clusters are formed in the feature space. The selection of the clustering method is an essential point in the clustering procedure. Several methods and algorithms for clustering have been developed. In this paper single-linkage clustering method are describe.

#### 1) *Single-Linkage Clustering*

The single link method is probably the best known of the hierarchical methods and operates by joining, at each step, the two most similar objects, which are not yet in the same cluster. The name single link thus refers to the joining of pairs of clusters by the single shortest link between them.

Given below are the steps of clustering process.

1. Let C1, C2, C3, ….Cn be the clusters, the distance between two clusters is defined to be the distance between the two objects they contain; that is $d_{CiCj}$= dij
2. Let t=1 be an index of the iterative process.Find the smallest distance between any two clusters. Denote these two closest clusters Ci, Cj.
3. Amalgamate clusters Ci and Cj to form a new cluster denoted Cn+1
4. Denote the distance between new cluster Cn+1 and all the remaining clusters Ck as follows
   $d c_{n+1} c_k = \min\{d_{cick}, d_{cjck}\}$
5. Add cluster Cn+1 as a new cluster and remove clusters Ci and Cj Let t=t+1.
6. Return to step 1 until one cluster remains

## IV. DESIGN AND DEVELOPMENT OF BASIC FRAMEWORK

The detailed block diagram is shown in figure 3.

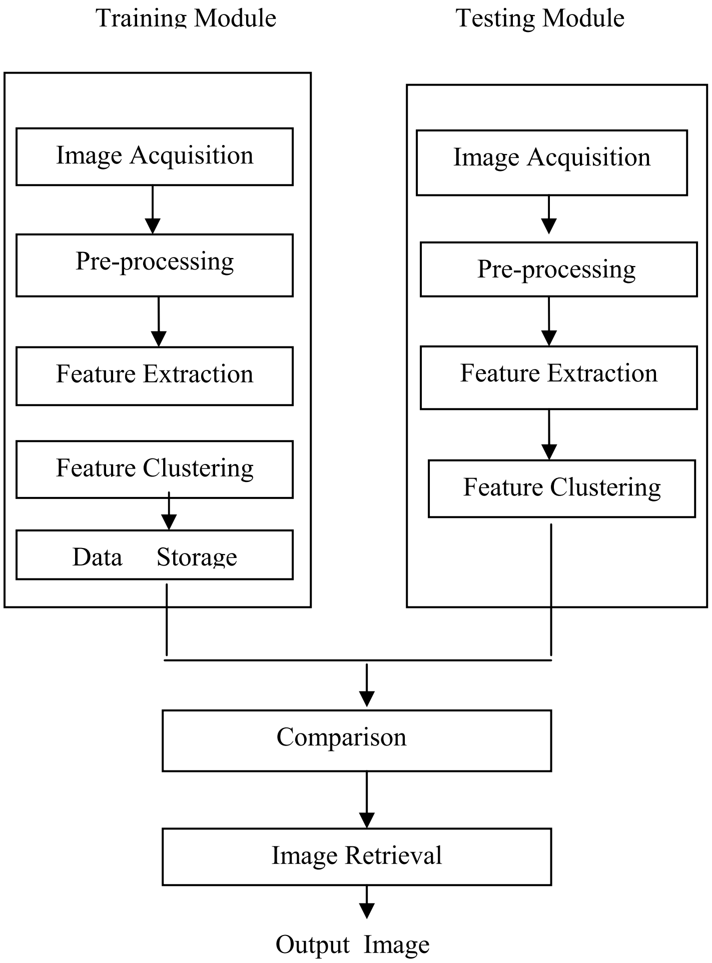



Figure3 Image Mining System

The block diagram is mainly divided into the following stages:

1) Image Acquisition: - An Image is an input to the system. This image can be internal query or the external query.. If it is external query then preprocessing of image is required.

2) Preprocess Image: - The aim of pre-processing is an improvement of the image data that suppresses unwanted distortions or enhances some for further processing. Preprocessing is nothing but the removing of noise, bringing the image in the proper format i.e. size of the image.

3) Feature Extraction: - Transforming the input data into the set of features is called features extraction. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. When performing analysis of complex data one of the major problems stems from the number of variables involved.

4) Feature Clustering: Feature clustering is a powerful alternative to feature selection for reducing the dimensionality of data. Evolve measures of similarity to cluster a collection of documents/terms into groups within which similarity within a cluster is larger than across clusters. Images are clustered such that the mutual information between the clusters and the image content is maximally preserved.

5) Image Retrieval :- An image retrieval is a process of browsing, searching and retrieving images from a large database of image Most traditional and common methods of image retrieval is using shape ,color and texture.

## V. Image Mining Based On Concept Analysis

Concept lattice and cloud model are use to analyze concept formally. Association rule mining is used to identify relationships among attributes in large data sets. Given a set of items and transactions, an association rule miner will determine which items frequently occur together in the same transactions.

### *A. Texture Feature*

Texture is a very general notion that can be attributed to almost everything in nature. For a human, the texture relates mostly to a specific, spatially repetitive (micro) structure of surfaces formed by repeating a particular element or several elements in different relative spatial positions. Generally, the repetition involves local variations of scale, orientation, or other geometric and optical features of the elements.

It is almost impossible to describe textures in words, although each human definition involves various informal qualitative structural features, such as fineness - coarseness, smoothness, granularity, lineation, directionality, roughness, regularity - randomness, and so on. These features, which define a spatial arrangement of texture constituents, help to single out the desired texture types, e.g. fine or coarse, close or loose, plain or twilled or ribbed textile fabrics. It is difficult to use human classifications as a basis for formal definitions of image textures, because there is no obvious ways of associating these features, easily perceived by human vision, with computational models that have the goal to describe the textures [7]

Some key definitions are introduced as follows.

*Root Pixel:* The root pixel of n x n neighborhoods is the centre pixel. Given an image of size N×N, there are (N - n+1)2 root pixels. For a sample image of size 5×5, each grid stands for one pixel; every shading pixel is a root pixel, so the total is 9.

*Item:* Each pixel in the neighborhood of a given root pixel could map to one item. It is specified by a triple (X, Y, I), where X is the column offset, Y is the row offset, and I is the intensity value of the pixel. When X = 0, Y = 0, triple (0, 0, i) specifies the items of root pixel in the neighborhood which is illustrated in Fig.4(b). S is a sample root pixel, the items of S are (–1, 1, 1), (–1, 0, 1), (–1, –1, 2), (0, 1, 2), (0, 0, 1), (0, –1,0), (1, 1, 1), (1, 0, 0), (1, –1, 1).

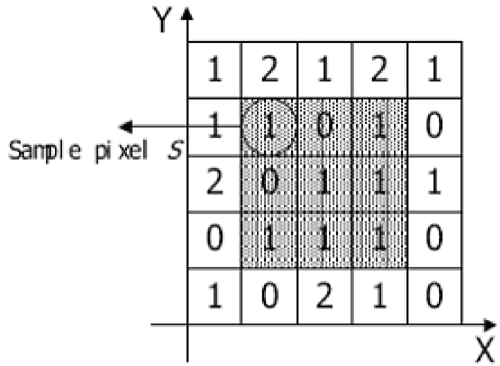



Fig 4(a) Sample Image (5 X 5)

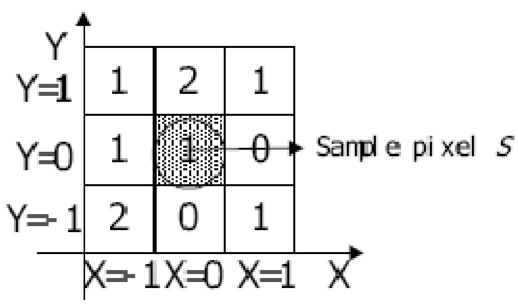



Fig 4 (b) Item Of Sample Pixel S (3 X 3)

*Item set:* A set of items, the number of items in the item set is referred to the cardinality of the item set.
Transaction: A set of items associated with a root pixel .One transaction is only associated with one root pixel. As shown in fig 4. (b), S is a sample root pixel, the transaction is {(–1, 1, 1), (–1, 0, 1), (–1, –1, 2), (0, 1, 2), (0, 0, 1), (0, –1,0), (1, 1, 1), (1, 0, 0), (1, –1, 1)}.
The algorithm of image texture feature association rule mining is outlined as follows.

Input: texture images, which will be mined.
Output: texture feature association rules.

1. Read Image and pre-process it.
2. Use cloud model to extract a series of concepts from the pre-processed image.
3. If the number of concepts generated in Step2 is too large, implement cloud concept ascending; if not, go to Step4.
4. Create transaction database D.
5. Take D as source dataset, build concept lattice C, and draw Hasse graph.
6. According to concept lattice C, generate texture feature association rules.
7. Choose some rules as texture feature knowledge.

The method of texture feature data mining is similar to the following methods of color feature data mining, shape feature data mining and spatial relationship feature mining.

### *B. Shape Feature*

Shape also is an important feature for perceptual object recognition and classification of images. It has been used in conjunction with color and other features for indexing and retrieval. Shape description or representation is an important issue both in object recognition and classification. Many techniques including chain code, polygonal approximations, curvature, Fourier descriptors and moment descriptors have been proposed and used in various applications [1]. Features such as moment invariants and area of region have been used, but do not give perceptual shape similarity.

*Definitions of terminology*

1) Major axis: it is the straight-line segment joining the two points on the boundary farthest away from each other (in case of more than one, select any one).
2)Minor axis: it is perpendicular to the major axis and of such length that a rectangle with sides parallel to major and minor axes that just encloses the boundary can be formed using the lengths of the major and minor axes.
3) Basic rectangle: the above rectangle formed with major and minor axes as its two sides is called basic rectangle.
Eccentricity: the ratio of the major to the minor axis is called eccentricity of the region.
4) Centroid: a single point of an object/region towards which other objects/regions are gravitationally attracted. For 2D shapes, the coordinates (Xc; Yc) of the centroid are defined as:

$$Xc = \sum x \sum y\, f(x;\, y)x / \sum x \sum y\, f(x;\, y) \qquad 4.1$$

$$Yc = \sum x \sum y\, f(x;\, y)y / \sum x \sum y\, f(x;\, y) \qquad 4.2$$

Where (x, y) are pixel coordinates and f(x,y) is set to 1 for points within or on the shape and set to 0 elsewhere.

*Basic idea* - Given a shape region, a grid space consisting of fixed-size square cells is placed over it so as to cover the entire shape region. Assign a "1" to cells with at least 25% of pixels covered and "0" to each of the other cells. A binary sequence of 1's and 0's from left to right is obtained as the shape feature representation. For example, the shape in the above figure can be represented by a binary sequence 11111111 01111111 00110110 00000110 00000010 00000000.The smaller the grid size, the more accurate the shape representation is and more the storage and computation requirements. The representation is compact, easy to obtain and translation invariant. Hence, scale and rotation normalization is carried out to make it invariant to scale and rotation.

*Rotation normalization* - The purpose of rotation normalization is to place shape regions in a unique common orientation. Hence the shape region is rotated such that its major axis is parallel to the x-axis. There are still two possibilities as shown in figure 5 (a) and (b), caused by 1800 rotation. Further, two more orientations are possible due to the horizontal and vertical flips of the original region as shown in figures 5 (c) and (d) respectively. Two binary sequences are needed for representing these two orientations. But only one sequence is stored and at the time of retrieval we can account for these two sequences.

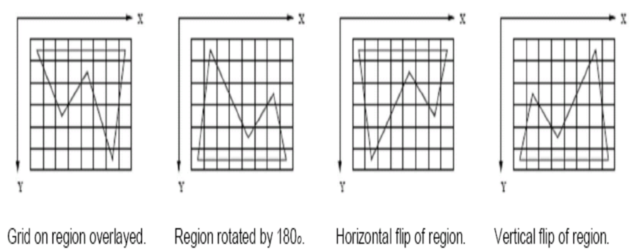



Fig 5(a) Grid on region overlaid (b) Region rotated by 180 $^{o}$. ( c ) Horizontal flip of region. (d) Vertical flip of region.

*Scale normalization*. To achieve scale normalization, proportionally scale all the shape regions so that their major axes have the same length of 96 pixels.

*Shape index*. Once the shape region has been normalized for scale and rotation invariance, using a fixed size of grid cells (say 8x8), we obtain a unique sequence for each shape region. The grid size in our proposed method is kept as 96 x 96 pixels. Each sub-grid cell is of size 12x12 pixels giving a binary sequence of length 64 bits per shape region. Using this sequence, we find both the row and column totals of the 8x8 grids and store them as our shape index, which is more robust and gives a better perceptual representation to the coverage of the shape. A suitable shape similarity measure using this index is employed for matching images at query time.

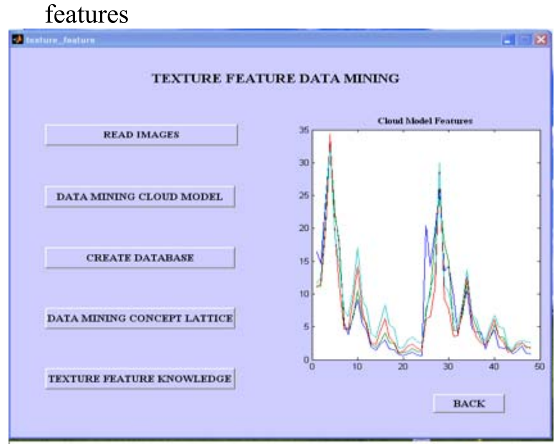
The algorithm for shape feature is outlined as follows.
Input: Images being mined.
Output: Shape feature association rules.
1. Read Image and pre-process it.
2. Set the boundary of sample objects.
3. Calculate the shape feature parameters of these objects.
4. Use cloud model to extract a series of shape feature concepts from the pre-processed images.
5. If the number of concepts generated in Step4 is too large, implement cloud concept ascending; if not, go to Step6.
6. Create database D.
7. Take D as source dataset, build concept lattice C, and draw Hasse graph.
8. According to concept lattice C, generate shape feature association rules.
9. Choose some rules as shape feature knowledge.

### VI. Experiments Analysis

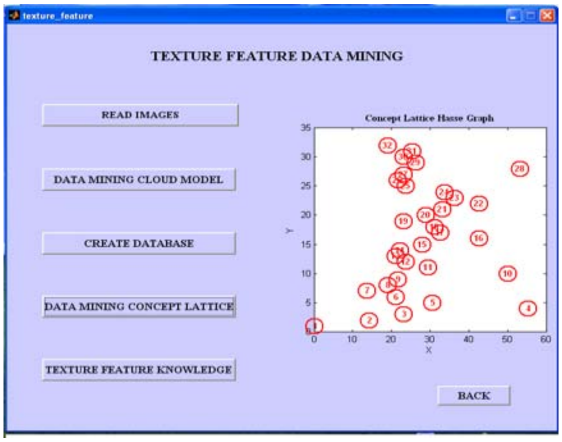
Based on the framework of image mining proposed in the paper, experiments of image texture feature are implemented to confirm the validity of the proposed framework.

1) Read the image for the Texture Feature data mining


2) Creating the cloud model of the image for texture features

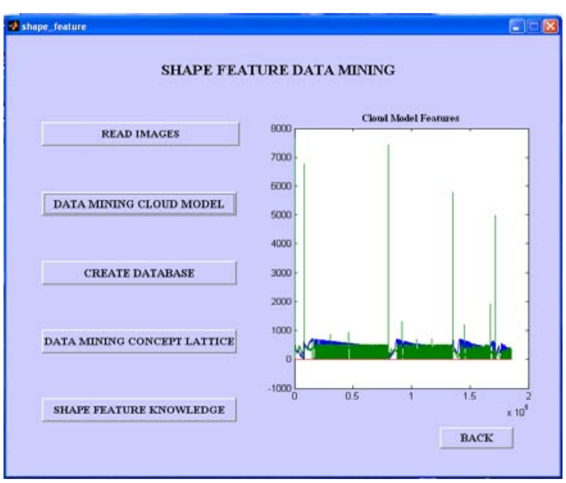


3) Concept lattice of the image has been done.

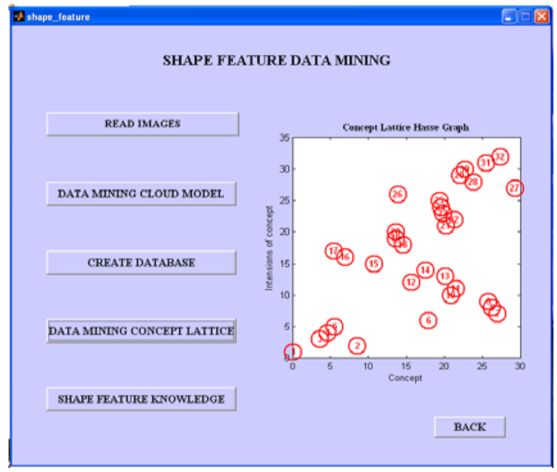



4) Read the image for the Shape Feature data mining.


5) Creating the cloud model of the image for shape feature.

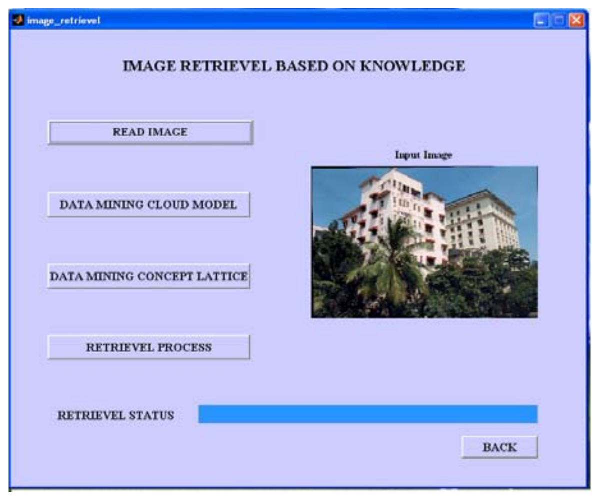



6) Concept lattice of the image has been done.




**A) Results of internal query**

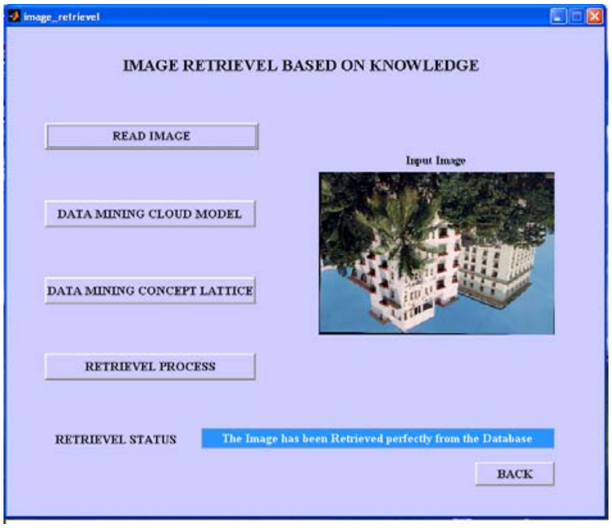
1) Retrieval of image based on Internal image.



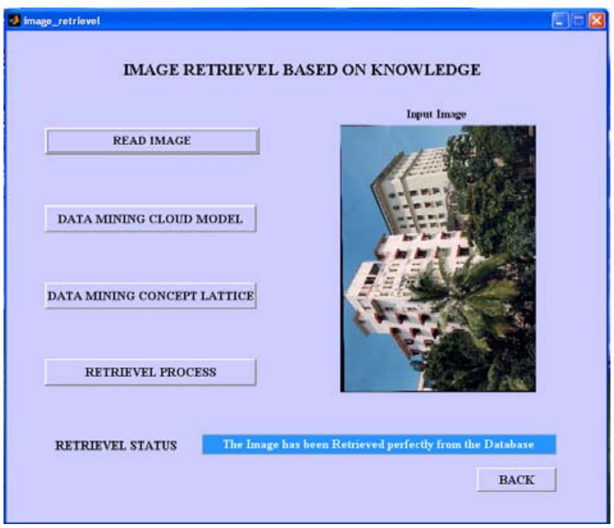

2) Result of image retrieval based on knowledge

**B.Results of internal distorted query**

1) Horizontally flip image.




2) Vertically flip images.




## VII. Conclusion

Image mining is a new research direction, and there is a need to research the basic theories and methods. Both concept lattice and cloud model provide the tools of formal concept analysis, which can be used to discover image knowledge from image data. Some experiment results confirm the validity of the methods of image mining, and shows that the framework of concept driven image mining could preferably mine the knowledge rules of texture and shape features.

The algorithm which is present in this paper has the following advantages.

1) The cloud concept provides a means of both qualitative and quantitative characterization of linguistic terms. This method can reflect the distribution of data in that domain while keeping the soft boundaries.

2) We can find the succinctness association rule from the hasse diagram, and can also choose the concept hierarchy to find the association rules.

If user has the interesting of the relation between part objects, there are two methods to deal with it

A. Directly constructing concept lattice from part objects using above algorithm. The shortcoming of this method does not have the complete relation.

B. Constructing concept lattice from all objects using above algorithm, choose the relevance concept hierarchies for the analysis of the relation between attributes.

The methods of image mining based on concept lattice and cloud model offer a novel idea for the research of image mining.

## VIII. References

[1] Ji Zhang, Wynne Hsu, Mong Li Lee. Image Mining: issues, frameworks and techniques,Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD 2001)

[2] Ordonez C. and Omiecinski E.," Discovering association rules based on image content" .Proceedings of the IEEE (ADL'99), 1999.

[3] Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis Mathematical Foundations*. *Springer-Verlag* Berlin Heidelberg. 1999.

[4] Hu Keyun. "Research on Concept Lattice and Rough Set based on Data Mining Methods". Beijing: Tsinghua University.2001. (In Chinese with English abstract){ ref given in paper keep at same no)

[5] Li Deyi, Du Yi. "Artificial Intelligence with Uncertainty". Beijing: The press of national defense industry. 2005.

[6] John A. Rushing, etc. "Using Association Rules as Texture Features", IEEE Trans, Pattern Analysis and Machine Intelligence, vol. 23, no. 8, pp. 845-858,2001.

[7] Bernhard Ganter and Rudolf Wille. "Formal Concept Analysis Mathematical Foundations". Pringer-Verlag Berlin Heidelberg. 1999.

[8] Carlos Ordonez and Edward Omiecinski. "Image Mining: A New Approach for Data mining" College of Computing Georgia Institute of Technology Atlanta, Georgia 30332-0280

[9] V. Castelli and L. D. Bergman (Eds.). Image Databases: Search and Retrieval of Digital Imagery. Wiley: New York, 2002.