ALTERNATIVE ACCESS METHODS FOR IMAGE RETRIEVAL

Mark J. Carlotto
TASC, 55 Walkers Brook Dr., Reading, MA 01867
Internet: mjcarlotto@tasc.com

ABSTRACT

Pattern recognition techniques for content-based image retrieval are addressed. Two of the methods are based on indexing images at the signal level (image feature indexing) and are able to find related images in a database that look similar. Next, a more general method is discussed which indexes images via attached textual descriptions (indexing by caption). A technique known as overlap or surrogate coding is used to index images by all of the words in a caption. By graphically comparing the structure of the corresponding feature spaces, indexing by caption is shown to be significantly better than image feature indexing over a sample database. Finally the possibility of indexing images at an intermediate level are explored. Two techniques are demonstrated for searching map databases based on region properties and global statistics. One of these indexes maps by region properties using surrogate codes where the regions in a map are analogous to words in a text. Surrogate coding allows texts to be compared in an amount of time proportional to the number of words, as opposed to quadratic time for direct key word matching. Access methods based on surrogate coding and other feature-based indexing schemes can thus be expected to scale well for larger databases.

1. INTRODUCTION

Previous research in image retrieval and image databases has concentrated largely on representations, query languages, user interfaces, and applications[1,2]. This paper explores the use of pattern recognition techniques for accessing images in large databases based on their content. We describe several content-based image access methods that have been implemented and tested on limited databases, provide preliminary estimates of their performance, and identify promising avenues toward practical near-term systems. Among the issues considered are: the levels in which images can be described and indexed (i.e, signal, semantic, or intermediate), the internal representation and its generality (limited image domain or general-purpose), access modes (e.g., query-by-example), and the potential scalability of the approach.

The organization of the remainder of this paper is as follows: A general architecture for content-based image access is presented in Section 2 which serves as a framework for the techniques discussed in the paper. Section 3 describes two preliminary experiments conducted to assess the feasibility of indexing images at the signal level, specifically in terms of local and global statistical representations. At the other extreme, Section 4 discusses a method based on indexing images by attached textual (semantic) descriptions. Finally in Section 5 we consider two techniques for indexing images, in particular map data, at an intermediate level. Preliminary conclusions of our work and future directions are summarized in Section 6.

Preprint from:
International Space Year Conference on Earth and Space Science Information Systems,
Feb. 10-13, 1992, Pasadena CA

2. ARCHITECTURE FOR CONTENT-BASED IMAGE ACCESS

Fig. 1 depicts a functional architecture for a content-based image database system. The encoding algorithm creates an internal representation of the image or query. Images are indexed by encoding and storing the internal representation of each image in working memory along with a pointer to the corresponding image on disk or some other mass storage device. The similarity measure defines how images are to be compared in terms of their internal representations. Images are retrieved by computing and comparing the internal representation of the query to those of all images in the database. Images are ranked in terms of their similarity to the query and are retrieved and presented to the user in that order.

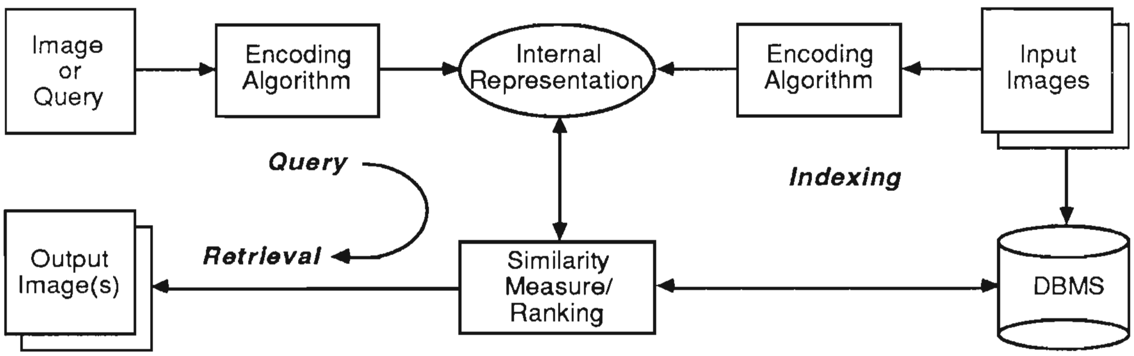

Fig. 1 Functional architecture for content-based image retrieval

3. IMAGE FEATURE INDEXING

Image feature indexing represents images by a small number of numerical features derived from the image. This section considers two approaches for indexing images at the signal level: one based on comparing images by matching local features (e.g., brightness, color, and texture at various positions), the other by comparing images in terms of their global energy distributions.

A test database containing 32, 512 by 512 grayscale images (Fig. 2) was used to evaluate these techniques. The database was compiled from several sources and includes aerial, planetary, biomedical and radar images, and photographs of faces, military vehicles, and outdoor scenes.

3.1 LOCAL FEATURE MATCHING

In this approach images are modelled by rectangular patches described by their position (i,j), and gray-level mean $\mu(i,j)$ and standard deviation $\sigma(i,j)$. The "distance" between two patches is defined as

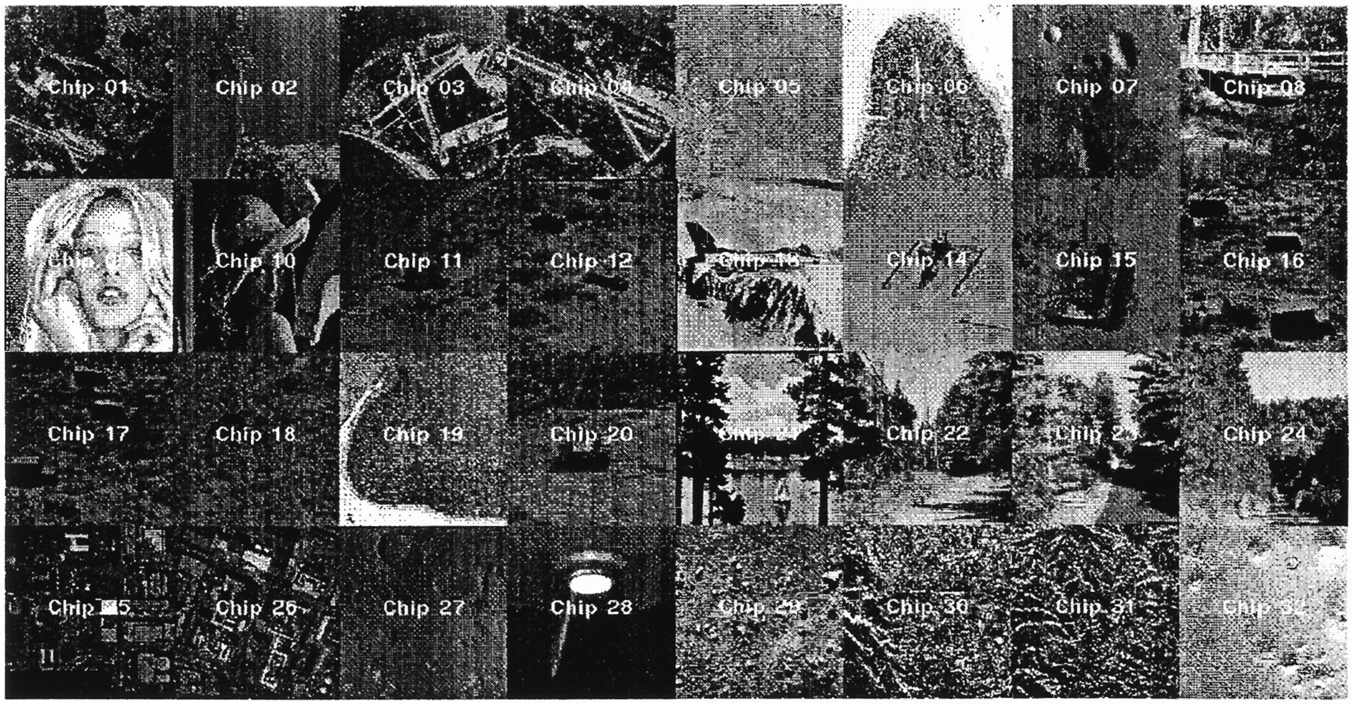

Fig. 2 Image Database

$$d(i,j,p,q,\alpha,\beta,\gamma) = \alpha\,(\,|i-p| + |j-q|\,) + \beta\,|\,\mu(i,j)-\mu(p,q)\,| + \gamma\,|\,\sigma(i,j)-\sigma(p,q)\,| \tag{1}$$

where (i,j) and (p,q) are the positions in the two images, and α, β, and γ are parameters that control the relative importance of differences in position, mean, and standard deviation. The similarity measure attempts to find, for each patch in one image, the most similar patch in the other image, and vice versa:

$$\sum_{p,q} \min_{|i-p| < r,\ |j-q| < r} d(i,j,p,q,\alpha,\beta,\gamma) + \sum_{i,j} \min_{|i-p| < r,\ |j-q| < r} d(i,j,p,q,\alpha,\beta,\gamma) \tag{2}$$

where r is the radius of the search. The above is like template matching except the features of image patches rather than the actual image patches themselves are compared. The computational complexity of this algorithm depends on the search radius and can be as high as quadratic in the number of patches per image.

An experiment was performed to assess the ability to query the database in Fig. 2 by example using this technique. Images were selected from the database and used to find other similar images. In each case, all the images in the database were ranked according to Eq. 2 and displayed in that order. We were able to retrieve, with some success, similar images of outdoor scenes, aerial views of water, and military vehicles; however a certain amount of parameter tuning was required in each case. These results are summarized in Table 1. Correct results shown in bold type. In the first two queries all related images were found in correct order. In the last query, seven of eight possible images were found. It is noted that the technique was unable to find similar images of airports, faces, and aircraft.

Query Image	First seven images retrieved from database (ordered left to right)						
22	**23**	**21**	**24**	3	12	8	16
2	**5**	7	27	15	11	14	18
11	**18**	**12**	**17**	**16**	20	**27**	**20**

Table 1 Results for local feature matching.

3.2 COMPARING GLOBAL ENERGY DISTRIBUTIONS

The previous approach models images locally. This technique compares the global energy distributions of images. A Laplacian pyramid[3] is used to compute a set of R bandpass images by averaging and downsampling,

$$i_{r-1}(x,y) = \downarrow\left[\frac{i_r(x,y) + i_r(x+1,y) + i_r(x,y+1) + i_r(x+1,y+1)}{4}\right] \tag{3a}$$

squaring the difference between an image and its local average,

$$e_r(x,y) = \left[i_r(x,y) - \uparrow i_{r-1}(x,y) \right]^2 \quad (3b)$$

and summing the total energy each of these images

$$E(r) = \sum_{x,y} e_r(x,y) \quad (3c)$$

where x and y are the spatial coordinates and r is the resolution. For the similarity measure, a normalized median difference between energy vectors is used to reduce the effect of outlier features:

$$\underset{r}{\text{median}} \left\{ 4^{-r} \mid E_a(r) - E_b(r) \mid \right\} \quad (4)$$

Sample results are shown in Table 2 for queries on airports, brain sections, battlefields, outdoor scenes, and radar imagery. The general performance is comparable to local feature matching, although the computational complexity is significantly less. Both techniques were able to find images that look the same but unable to find related images that look different.

Query Image	First seven images retrieved from database (ordered left to right)						
3	**4**	**1**	16	17	8	21	22
6	**19**	15	7	20	9	27	11
18	**29**	**12**	**11**	**20**	24	19	**15**
22	**23**	13	3	4	16	**24**	**21**
30	**31**	29	18	12	3	17	4

Table 2 Results for indexing based on global energy distributions.

4. INDEXING BY CAPTION

Indexing by caption is a method that allows images to be accessed at the semantic level. It involves creating textual descriptions of images and using full-text document retrieval techniques to index images by all of the words in the attached texts. This allows images to be retrieved by key word(s) or by example (i.e., using the full text of the example image) and is based on the assumption that similar texts (hence images) tend to contain the same words.

4.1 SUROGATE CODING

One of several techniques that have been used in full-text document retrieval systems is known as overlap or surrogate coding[4,5]. Texts are represented by binary vectors and are compared in terms of their Hamming distance or correlation. A graphical overview of surrogate coding is shown in Fig. 3. The process involves 1) selecting content-bearing words from the text, 2) converting words into randomly assigned bit patterns that are stored in a dictionary , and 3) setting those bit positions to one in a binary vector. The number of codes (words) that can be uniquely represented is given by

$$\binom{K}{P} = \frac{K!}{P!\,(K-P)!} \qquad (5)$$

where K is the length of the vector in bits, and P is the number of bits per word. For $K = 1024$ and $P = 4$, this number exceeds 10^{10}! Since these codes may overlap, the probability that two or more codes will interfere and give rise to the code of a word that is not in the text but is in the dictionary is approximately equal to[6]

$$P_{error} = \left(1 - \left(\frac{K-1}{K}\right)^{RP}\right)^{P} \qquad (6)$$

where R is the number of words that are encoded per vector. For example, if $K = 1024$, $P = 4$, and $R = 64$, $P_{error} \approx 0.0024$.

A short textual description (caption) was created for each image in the database in Fig. 2. A typical caption is shown in Fig. 3. Each caption was encoded using the parameter values: $K = 1024$, $P = 4$, and $R = 64$ (i.e., up to 64 words may be encoded per image). An image browser was implemented to retrieve images by key word(s) and by example.

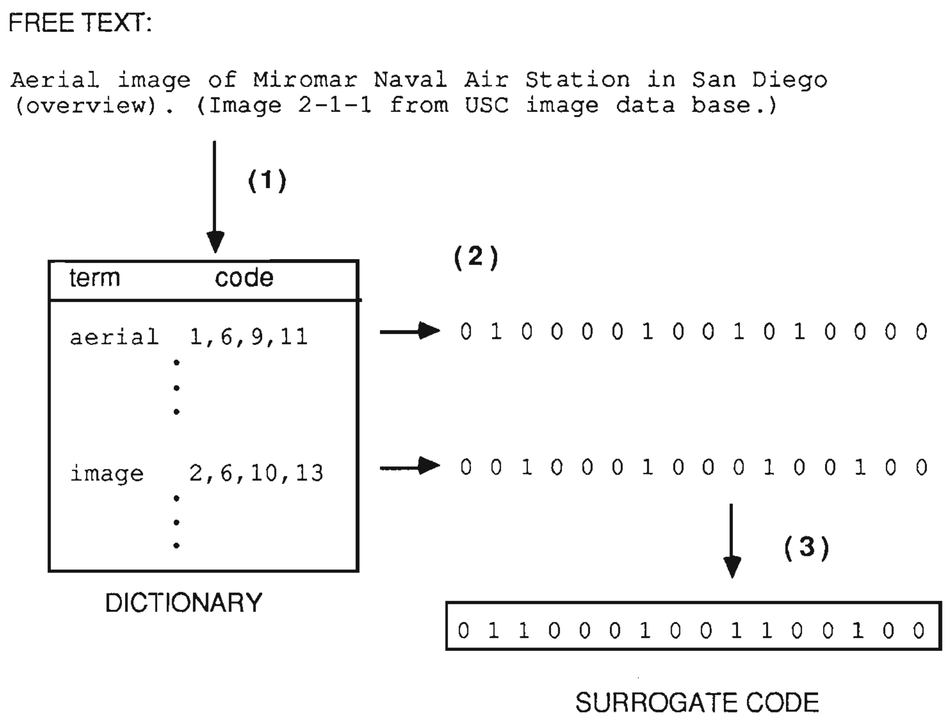

Fig. 3 Surrogate coding process.

5.2 COMPARISON OF INDEXING TECHNIQUES

An effective indexing scheme produces a clustered object space in which classes of similar objects are easily separated from others[7]. In order to visually assess the performance of a particular image indexing technique we project the internal representation of each image in the database to a point in a 2-d graphical space. A nonlinear mapping algorithm[8] is used which preserves distances (similarities) between points in the two spaces.

The 2-d map of the image database (Fig. 2) indexed by caption as described above is shown in Fig. 4. Image classes are distinct with related images appearing in clusters. The distance between images is inversely related to the correlation between surrogate codes which is approximately equal to the number of common words in their captions. Clusters that are related are also close together (e.g., radar and aerial images over California, and images related in some way to NASA).

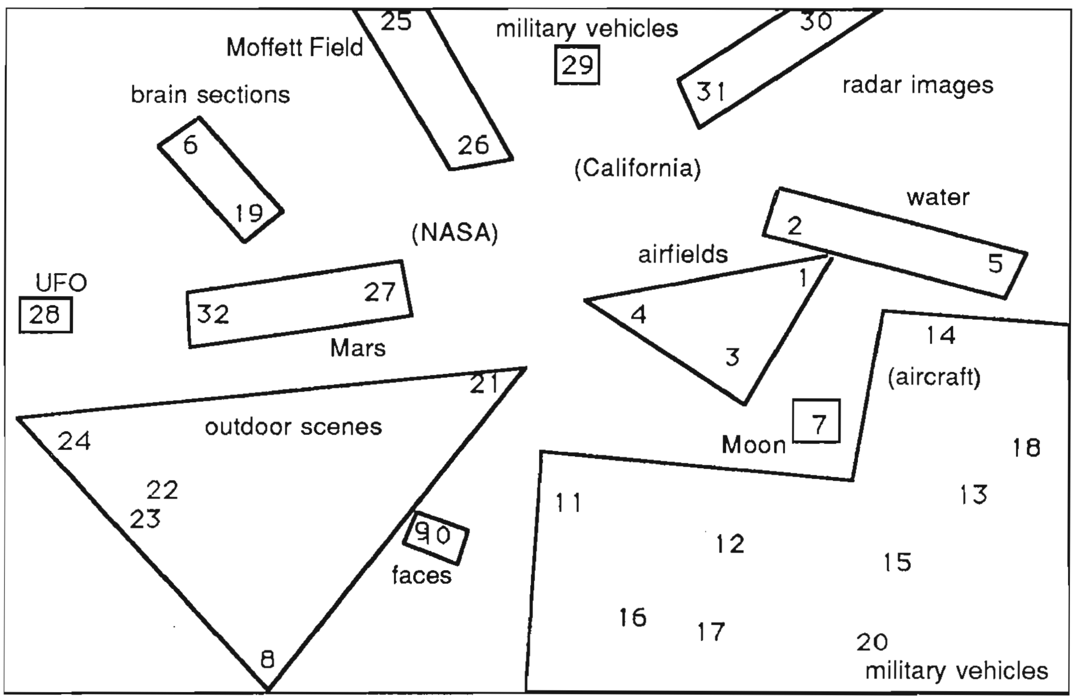

Fig. 4 2-d map depicting contents of image database indexed by caption

This result is contrasted with the map in Fig. 5 for indexing by global energy distribution. Few distinct image classes occur in that map since images which are similar in meaning (semantic level) are often dissimilar in appearance (signal level). Images of aircraft and faces are close together in Fig. 4 because their descriptions are similar, but are far apart in Fig. 5 due to differences in brightness and texture.

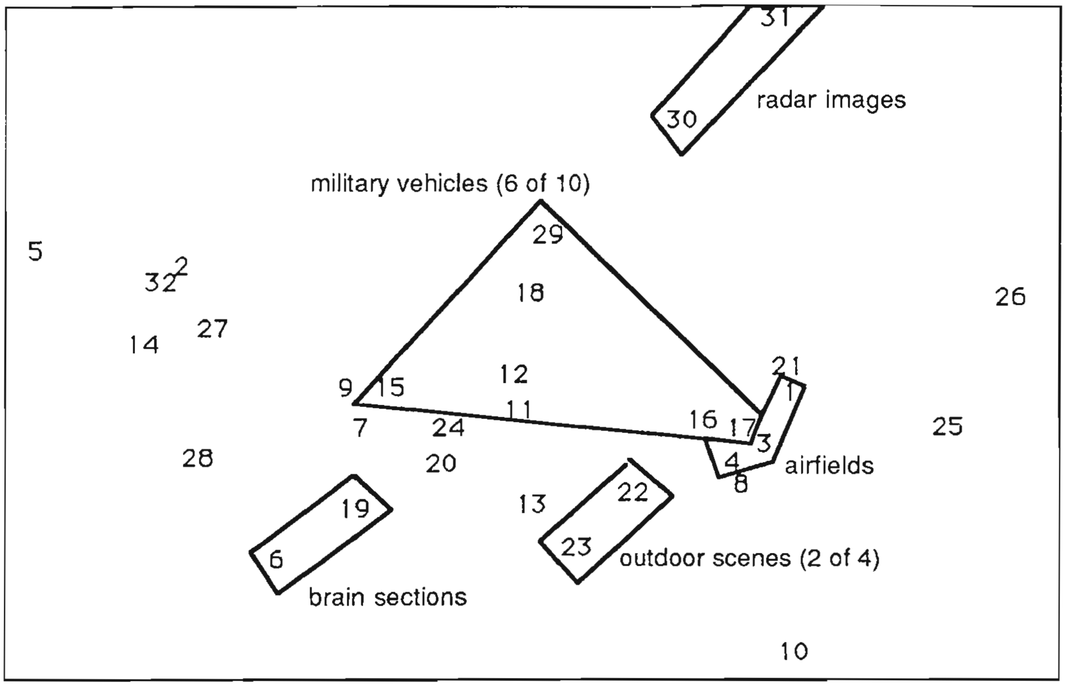

Fig. 5 2-d map depicting contents of image database indexed by energy distribution

5. INTERMEDIATE-LEVEL INDEXING

Images are easily indexed by signal properties; however, it is difficult to query an image database directly at the signal level except by example. In the previous section it was shown that if semantic descriptions are available or can be easily generated, images can be effectively indexed and retrieved at that level. In situations where this is not possible or feasible, indexing at an intermediate-level (i.e., in terms of physical or perceptual descriptions of images) may bridge the gap between signals and semantics. Examples of intermediate-level features that may be useful for indexing specific kinds of imagery include: 2-d shape information (gray-level images), surface material composition (multispectral imagery), 3-d object properties (range sensor), topographic primitives (digital elevation models), and others.

As an example we consider the task of searching a map database by land cover class. A database was constructed from a 768 x 768 portion of a land cover classification map derived from Landsat TM imagery partitioned into 36, 128 x 128 sections (Fig. 6). The area is over central Germany and contains agricultural fields, forested areas, and several built-up areas along a river valley.

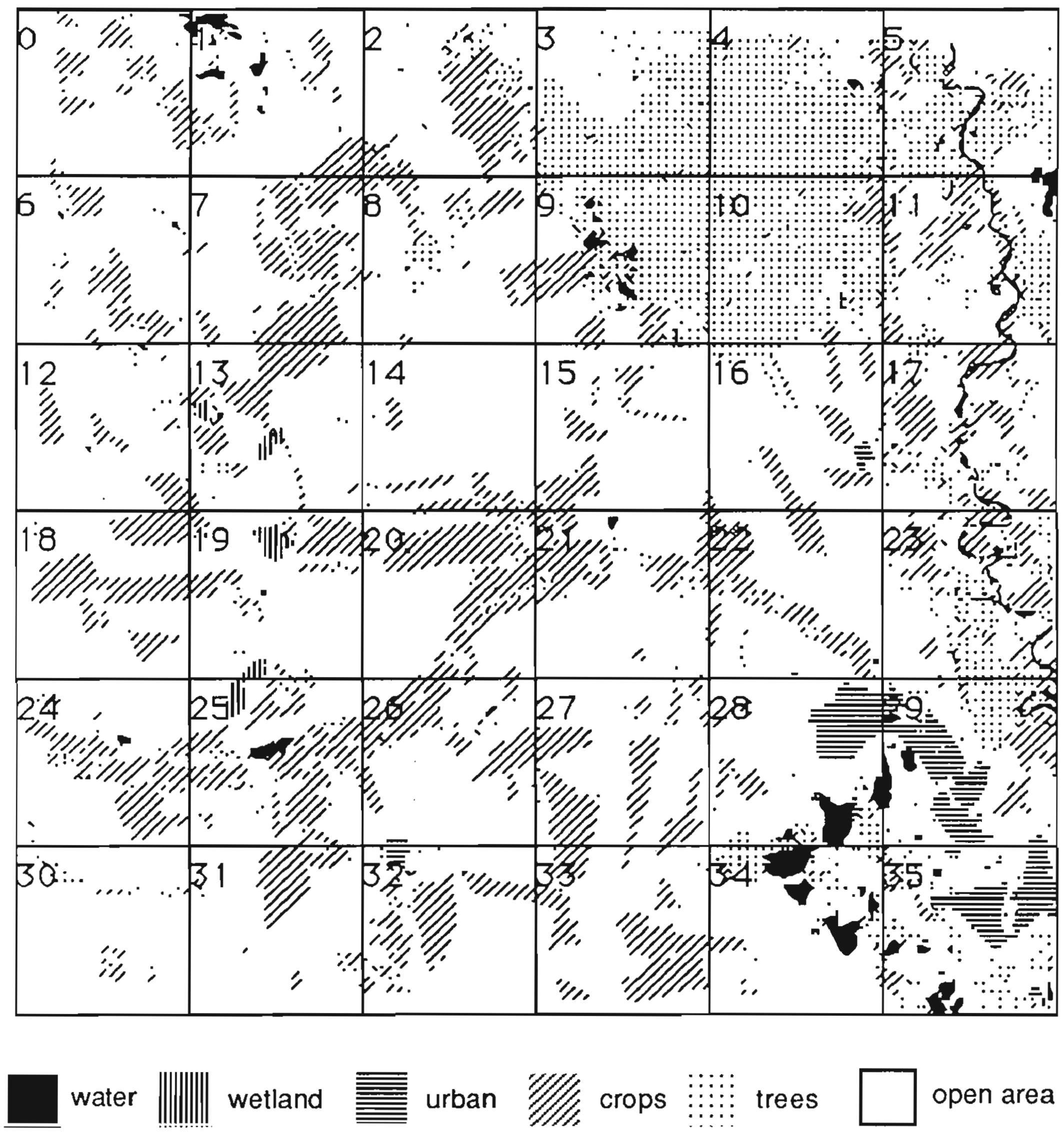

Fig. 6 Map Database

Two methods were implemented. The first encodes regions in maps in the same manner that words in texts were encoded in the previous section. Regions are assigned dictionary terms based on their properties. In the following example, three properties are used: land cover class, area, and perimeter. Classes are assigned numerical labels; area and perimeter are coarsely coded by quantizing $\log_4$ (area) and $\log_2$ (perimeter). Each region in the map is converted into a term; e.g.,

region class = 8, area = 100, and perimeter = 30 $\rightarrow$ "8-3-4"

These terms are converted into assigned bit patterns which are stored in a binary vector (surrogate code).

A map browser was implemented to search the database for maps that contain specific regions (terms) and to find other maps in the database that are similar to a given map (i.e., using all terms). The 2-d map of the map database indexed by regions is shown in Fig. 7. Maps of forested areas, lakes, urban areas, and along a river valley appear in clusters. The remaining areas are largely agricultural but contain smaller amounts of trees, water, and urban areas.

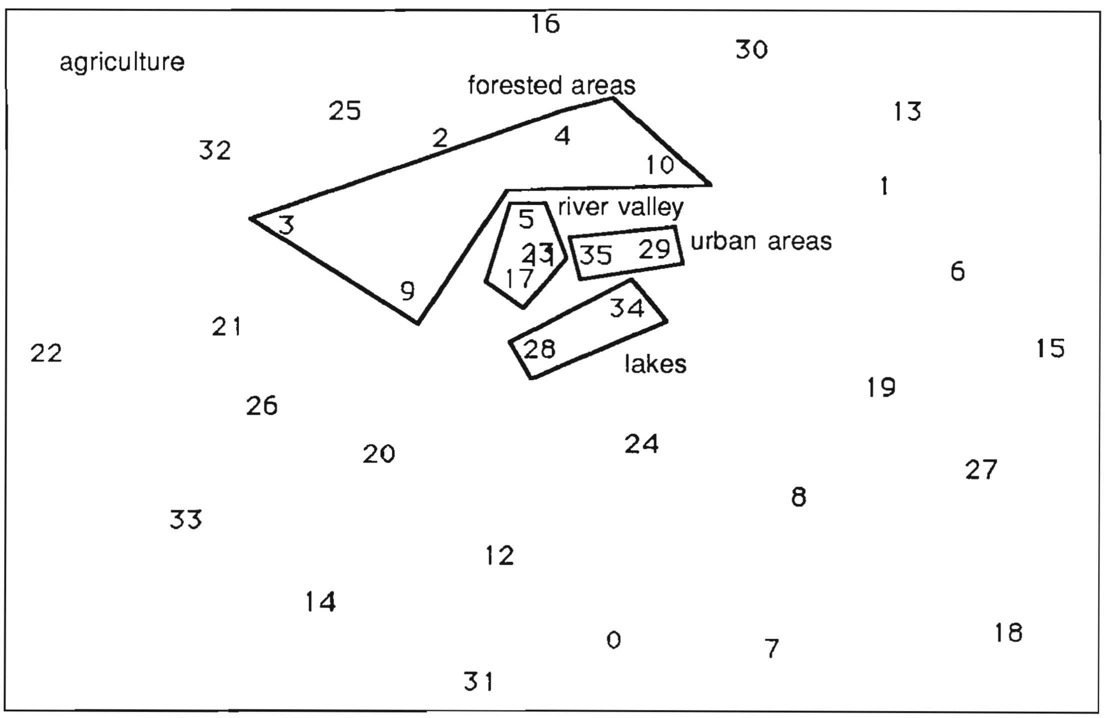

Fig. 7 2-d map depicting contents of map database indexed by region properties

The second method simply computes a histogram of the number of pixels in each land cover category for each image. Another map browser was implemented to search the database for maps by land cover class (e.g., "Find all maps that contain wetlands and no urban areas") and to find other maps in the database that are similar to a given map (based on total absolute difference between histograms). The 2-d map of the map database indexed by land cover frequency is shown in Fig. 8. Clusters of maps of forested areas, urban areas, and along a river valley are not as well-differentiated as in Fig. 7. Agricultural areas organize in a cluster with maps containing the least amount of crops farthest from maps with the most.

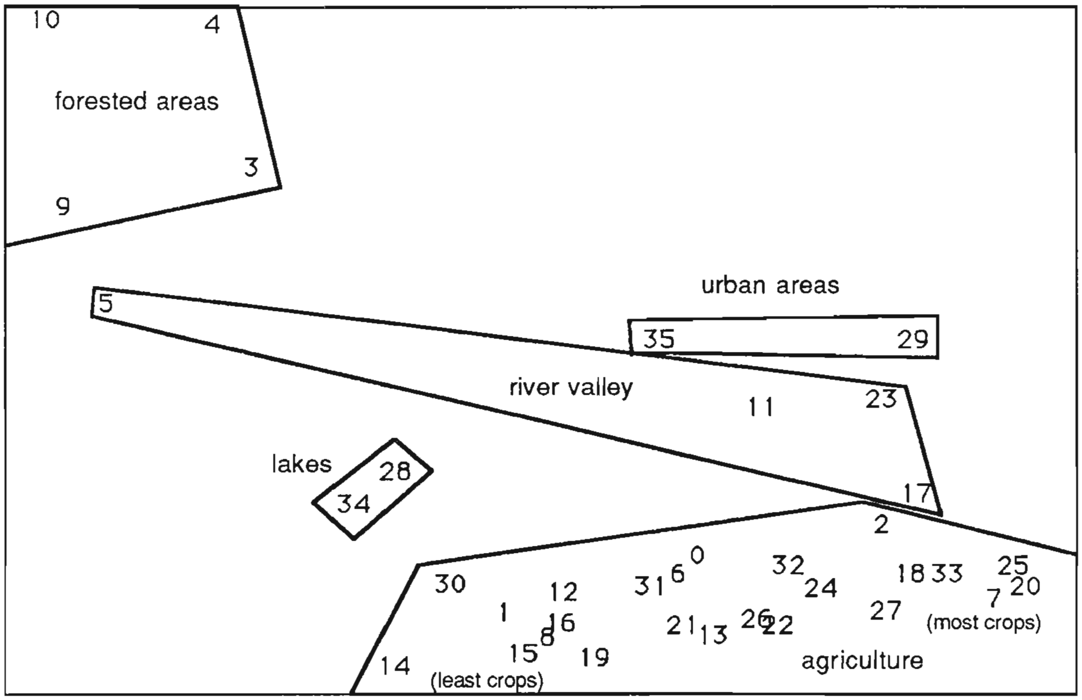

Fig. 8 2-d map depicting contents of map database indexed by land cover frequency

6. SUMMARY

This paper has explored the use of pattern recognition techniques for content-based image retrieval. Five techniques were described. Two techniques based on indexing images at the signal level (image feature indexing) were able to find images in a database that look the same but were unable to find related images that look different. A method for indexing images via attached textual descriptions (indexing by caption) was described next where texts are represented by surrogate (binary) codes. By graphically comparing the structure of the corresponding feature spaces, indexing by caption was shown to be significantly better than image feature indexing over a sample database. Finally the possibility of indexing images at an intermediate level was explored. Two techniques were demonstrated for searching map databases based on region properties and global statistics. One of these indexed maps by region properties using surrogate codes where the regions in a map are analogous to words in a text.

Surrogate coding allows texts to be compared in an amount of time proportional to the number of words, as opposed to quadratic time for direct key word matching. Access methods based on surrogate coding and other feature-based indexing schemes can thus be expected to scale well for larger databases. An area of future work is to explore the possibility of using surrogate codes to index images at multiple levels of description, e.g., by textual descriptions, by physical/perceptual information derived from the image, and by collateral data such as image acquisition conditions, geographic coordinates, and time.

REFERENCES

1. Special Issue on Image Databases, IEEE Trans. Software Engineering, Vol. 14, No. 5, May 1988

2. Special Issue on Image Database Management, IEEE Computer, December, 1989.

3. P. Burt, "Algorithms and architectures for smart sensing," Proc. DARPA Image Understanding Workshop, 1988, Cambridge MA.

4. D. Knuth, The Art of Computer Programming: Vol 3: Sorting and Searching, Addison-Wesley, Reading MA, 1973.

5. C. Stanfill and B. Kahle, "Parallel free-text search on the Connection Machine system," Comm. ACM, Vol. 29, No. 12, Dec. 1986.

6. C. Roberts, "Partial-match retrieval via the method of superimposed codes," Proc. IEEE, Vol. 67, No. 12, Dec. 1979.

7. G. Salton, A Theory of Indexing, Society for Industrial and Applied Mathematics, Philadelphia PA, 1975

8. J. Sammon, "A nonlinear mapping algorithm for data structure analysis," IEEE Trans. Computers, Vol. C-18, No. 5, May 1969.