

Automated Feature Extraction and Retrieval of Ultra Sound Kidney Images using Maxi-Min Approach

S. Manikandan
Professor

Department of Computer Applications
R.M.D. Engineering College,
Kavaraipettai, Tamil nadu, India.

V. Rajamani
Principal

Indra Ganesan College of Engg.,
Trichy, Tamil nadu,
India.

ABSTRACT

A general purpose medical image retrieval framework has been proposed with two subsystems namely enrollment and the query subsystem. As an attempt to design a new content based image retrieval methodology following the above framework, MAXI-MIN approach is implemented for the ultra sound kidney images for the retrieval process. Around hundred ultrasound kidney images have been collected from the clinical laboratory and fourteen features have been extracted from the existing literature for database creation. The difference between the feature of query image and features of each image in the database has been calculated. The image which is more similar to the query image has been retrieved as the resultant image based on the maximum number of occurrences of features for the minimum difference. If the query image does not match with the stored database image, the query image is added as a new image in the database. The process is highly automated and the system is capable of working effectively across different issues without human interference.

Keywords: Medical image, query image, image retrieval, image database, features extraction.

1. INTRODUCTION

The growing number of digital image acquisition and storage systems in clinical routine raises demands for new access methods. Still, most picture archiving and communication systems (PACS) only use textual information to access a patient's image data, which has been mainly entered manually. Content-based image retrieval (CBIR) depends on automatically extracted content descriptions (numerical features) for each image as well as their storage and comparison upon a query. Image retrieval has been an extremely active research area over the last 10 years, but first review articles on access methods in image databases appeared already in the early 80s [1]. The following review articles from various years explain the state-of-the-art of the corresponding years and contain references to a large number of systems and descriptions of the technologies implemented. Enser [2] gives an extensive description of image archives, various indexing methods and common searching tasks, using mostly text-based searches on annotated images. In [3], an overview of the research domain in 1997 is given and in [4], the past, present and future of image retrieval is highlighted. In [5] an almost exhaustive overview of published systems is given and an evaluation of a subset of the systems is attempted in [6].

Unfortunately, the evaluation is very limited and only for very few systems. The most complete overview of technologies to date is given by Smeulders et al. in [7]. This article describes common problems such as the semantic gap or the sensory gap and gives links to a large number of articles describing the various techniques used in the domain. For an even deeper introduction into the domain, several theses and books are available [8-11]. The only article reviewing several medical retrieval systems so far, is to our knowledge [12]. It explains using one paragraph per topic a number of medical image retrieval systems. No systematic comparison of the techniques employed and the data/evaluation used has been attempted.

Considering the implementation of a CBIR system in medical applications, there is currently a gap between monolithic CBIR systems for general-purpose image retrieval [13], and programming tools which support the development of image processing algorithms and the automatic distributed execution of them. Existing general-purpose CBIR systems closely couple feature extraction, feature storage, feature comparison, and the query interface [14]. Since changes often affect all system components, this makes it difficult to extend them according to the specific requirements of medical image retrieval. Existing image processing tools like Khoros/Cantata/VisiQuest1, the Insight Toolkit (ITK)2, the Visualization Toolkit (VTK)3, or ImageJ4 provide a huge number of routines useful for feature extraction and comparison out of the box, but they lack support for organizing feature storage as needed by a CBIR system. This also complicates the easy deployment of retrieval algorithms to the end-user, who must not be involved in technical details.

To close this gap and further motivated by the considerable data volume of medical image archives, recent works apply new technologies to medical CBIR [15]. In recent years, much research has been progressed into specific medical image retrieval systems [16]. It provides the distributed storage of a large scale image database and utilizes distributed computing for content-based retrieval on the image data. The main focus lies on the optimization of the queries themselves.

1.1 Techniques used in medical image retrieval

This section describes the various techniques that are currently used or that have been proposed for the use in medical image retrieval applications. Many of the techniques are similar to those used for general content based retrieval but also

techniques that have not yet been used in medical applications are identified. A special focus is put on the data sets that are used to evaluate the image retrieval systems and on the measurements used for evaluation. Unfortunately, the performance evaluation of systems is currently strongly neglected. Machine learning in medical applications also gets increasingly more important and it is essential to research the various possibilities. Specialized workshops exist for this area [17].

Most systems do not give many details on the distance measurements or comparison methods used which most likely implies a Euclidian vector space model using either a simple Euclidean distance (L2) or something close such as city block distance or L1. To efficiently work with these distances even in large databases, the dimensionality is often reduced. This can be done with methods such as Principal Component Analysis (PCA) or Minimum Description Length (MDL) [18] that try to reduce the dimensionality while staying as discriminative as possible. In principle, redundant information is removed but this can also remove small but important changes from the feature space. Techniques such as KD-trees and R-trees are also used in medicine for efficient access to such a large feature spaces. On the other hand, statistical methods are used for the comparison of features that can be trained with existing data and that can then be used on new, incoming cases. This work proposes a new approach using image distortion model for comparing the images.

2. IMAGE RETRIEVAL MODEL

Image retrieval is the process of finding similar images from a large image archive with the help of some key attributes associated with the images or features contained in the images [19][20]. The input images are analyzed to extract the features and these features are stored in the image database along with the original images. In our paper, fourteen features have been extracted. Whenever an image is submitted for search, it's analyzed and its features are also extracted. These features are matched against those in the database. A set of closely matching images are brought out as the result of search output. The architecture of our proposed framework can be divided into two main subsystems namely, the enrollment and the query subsystem. The enrollment subsystem is responsible for acquiring the information that will be stored in the database for later use. On the other side, the query subsystem is responsible for retrieving similar images from the database according to the user's query image.

3. FEATURE EXTRACTION

Medical images are normally distinguished based on their feature characteristics. Feature extraction is a primary step for retrieving image from the image databases [21]. Various features extracted from the medical images are auto correlation, contrast, prominance, shade, dissimilarity, energy, entropy, homogeneity, maximum probability, variance, co-variance, correlation, inverse difference moment and inertia.

The various feature descriptors are generally represented and organized in an index structure or index file. The basic idea of feature indexing is to extract the features from each image in the database and then to map features into points in a multidimensional feature space. CBIR is more challenging in medical domain due to the complex nature of images [22]. The main reason is that, important features in biomedical images are often local features of pathological regions or lesions, rather than global features of entire image. Generating local features is much more complex than global features; however, it can describe fine details of the images and allow efficient retrieval of relevant images based on local object properties. In the feature extraction subsystem, image processing and pattern analysis techniques are used to extract numeric descriptors of various visual features. Each feature may have several representations and generally a weight or value is assigned to each of these features and their descriptors or representations.

4. MAXI-MIN APPROACH

Image retrieval is the process of finding similar images from a large image archive with the help of some key attributes associated with the images or features contained in the images. Here, the input image is given by the user. It is preprocessed to get the feature extraction values. The images given by the user and also the images in the database have been subjected to the absolute difference method. The differences between the feature of query image and features of each image in the database have been calculated. Difference value is calculated for all the extracted features. Then the difference values have been sorted in ascending order. The images with lowest difference which is nearest to zero are returned. The resultant image is displayed based on the lowest difference for maximum number of features. If the differences are very high for all the extracted features, then the query image did not have any match with any image which will be stored in the database. In that case, the query image has been stored in the created image database as a new image. This maxi-min approach has been adopted by the following step by step procedure.

4.1 Steps for calculating maximum of minimum difference

Step 1: The difference measure between the query image features and the database image features have been computed as follows.

General expression to find the difference measure between the feature values is

$$D[j] = \text{abs}[q - y_j] \quad (1)$$

where D is the difference measure, q is the query image and yj is the jth image in the image database.

The difference between the auto correlation of the query image and the database images have been computed

$$D_a[j] = \text{abs}[q_a - y_w] \quad (2)$$

where, D_a is the auto correlation difference, q_a is the auto correlation value of query image and y_{aj} is the auto correlation value of j th image in the database.

By using the above formula, the difference value for all the various extracted features has been calculated.

Step 2: These difference measures are sorted in ascending order.

Step 3: First value is the minimum value. It is implied that the image stored in the database which is correspond to that feature value is closely matched with the query image. But by comparing only one feature, retrieving the correct matching image from the database is not accurate.

Step4: Distance measure has been calculated for all the images by comparing all the feature values.

Step5: More than one feature values are considered for comparison. Different images are having the close match with the query image when many feature values are used for comparison.

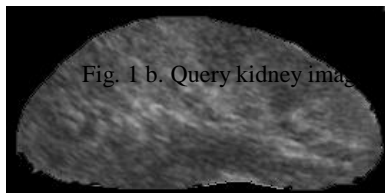
Step6: Then the Maximin approach is implemented. It has been found which image is having the minimum distance measure for maximum number of features. That image has been returned as a resultant image and it has been retrieved from the database.

Step 7: There is no match means the query image is stored in the database as a new image.

5. RESULTS AND DISCUSSIONS

Table 1 gives the various feature values which have been extracted from a sample image. A Query image which is in the database is given in Fig. 1 and the related features have also been extracted and compared with the features of the images which are in the database by Maxi-Min approach.

Fig. 1 Query kidney image



Ultrasound Image selected

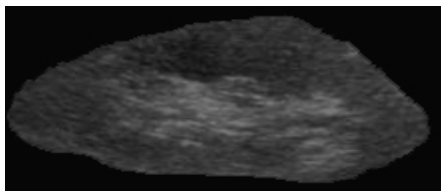


Fig. 2 Resultant image retrieved from database (3rd Image

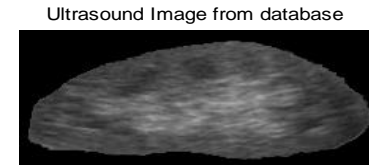


Table 1 Feature values of a sample image

Feature Number	Feature Name	0°	45°
F1	Autocorrelation	7.818436	7.811211
F2	Contrast	0.130795	0.249430
F3	Promenance	51.38861	48.52966
F4	Shade	-2.614942	-3.199996
F5	Dissimilarity	0.112025	0.204540
F6	Energy	0.280524	0.248477
F7	Entropy	1.621969	1.788047
F8	Homogeneity	0.946526	0.903534
F9	Maximum probability	0.406745	0.376467
F10	Variance	7.803828	7.855098
F11	Co-Variance	1.285959	1.218466
F12	Correlation	0.951606	0.907149
F13	IDM	0.945797	0.901966
F14	Inertia	0.130795	0.249430

Our novel Maxi-Min approach, it first calculates the minimum distance between the query image and image database for the entire feature sets. The comparison has been made for every feature value and based on the maximum number of occurrences. Then the resultant image has been retrieved from the database. The detailed calculation for two query images has been given below.

Table 2 shows the difference between the feature values of query image 1 and the retrieved image from the database. The bolded values are the minimum distances obtained compared with other image in the database. In 00 angles, minimum distance has been occurred for four features viz. shade, dissimilarity, homogeneity and idm. But in the case of 450, minimum distance has been obtained for the four features like contrast, shade, maximum probability and inertia.

Table 2 Difference between query image and the resultant image

Feature Number	Feature Name	0°	45°
F1	Autocorrelation	2.692768	2.685788
F2	Contrast	0.073872	0.014069
F3	Promenace	39.559907	42.209973
F4	Shade	0.972114	0.774777
F5	Dissimilarity	0.018047	0.009799
F6	Energy	0.011035	0.0165755
F7	Entropy	0.104546	0.063228
F8	Homogeneity	0.000114	0.008855
F9	Maximum probability	0.018581	0.010107
F10	Variance	2.715166	2.678137
F11	Co-Variance	0.620421	0.655637
F12	Correlation	0.042081	0.033161
F13	IDM	0.003441	0.007286
F14	Inertia	0.073872	0.014069

For query image 1, by comparing the auto correlation feature (F1) 10th image has the minimum distance and 7th image has the minimum distance by comparing contrast feature value (F2) in 00 angle. For 450 angles, 10th and 3rd image are having the minimum distance by comparing the features auto correlation and contrast. Like this the images which are having the minimum distances are taken by comparing all the extracted features. The image which has minimum distance for maximum number of features has been retrieved. In 00 angles 3rd image is having the minimum distance for maximum number of occurrences (i.e. 4 times). Finally, 3rd image has been retrieved as a resultant image. But in the case of 450 angles also 3rd image is having the minimum distance for maximum number of features (i.e. 4 times). It is concluded that in any orientation, 3rd image has been retrieved from the image database and is depicted in Fig.2.

6. CONCLUSION

A novel MAXIMIN absolute difference method for CBIR system has been implemented and tested. The system is composed of database for storing nearly hundred images and its fourteen features. Features are extracted and stored in a database for comparison with the feature value of query image. If the query image is matched with the database image, then that image has been retrieved otherwise it has been added as a new image in the database. This can improve the system performance a lot. The present work has been tested for some

other types of images like MRI, CT scan also. But it has not been matched with the images in the database. This retrieval process may be used for diagnosis purpose. The future work can be concentrated for more number of features for comparison and retrieval.

7. REFERENCES

- [1] S.-K. Chang, T. Kunii, 1981. Pictorial data-base applications, IEEE Computer 14 (11) 13-21.
- [2] P. G. B. Enser, 1995. Pictorial information retrieval, Journal of Documentation 51 (2) 126-170.
- [3] A. Gupta, R. Jain, 1997. Visual information retrieval, Communications of the ACM 40 (5) 70-79.
- [4] Y. Rui, T. S. Huang, S.-F. Chang, 1997. Image retrieval: Past, present and future, in: M.Liao (Ed.), Proceedings of the International Symposium on Multimedia Information Processing, Taipei, Taiwan.
- [5] J. P. Eakins, M. E. Graham, 2000. content-based image retrieval, Tech. Rep. JTAP-039, JISC Technology Application Program, Newcastle upon Tyne.
- [6] C. C. Venters, M. Cooper, 2000. content-based image retrieval, Tech. Rep. JTAP-054, JISC Technology Application Program.
- [7] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, 2000. Content-based image retrieval at the end of the early years, IEEE Transactions on Pattern Analysis and Machine Intelligence 22 No 12 1349-1380.
- [8] H. Muller, User interaction and performance evaluation in content-based visual Information retrieval, Ph.D. thesis, Computer Vision and Multimedia Laboratory, University of Geneva, Geneva, Switzerland (June 2002).
- [9] J. R. Smith, 1997. Integrated spacial and feature image systems: Retrieval, compression and analysis, Ph.D. thesis, Graduate School of Arts and Sciences, Columbia University, 2960 Broadway, New York, NY, USA.
- [10] A. del Bimbo, Visual Information Retrieval, Academic Press, 1999.
- [11] S. M. Rahman, 2001. Design & Management of Multimedia Information Systems: Opportunities & Challenges, Idea Group Publishing, London.
- [12] L. H. Y. Tang, R. Hanka, H. H. S. Ip, 1999. A review of intelligent content-based indexing and browsing of medical images, Health Informatics Journal 5 40-49.
- [13] Carson C, Belongie S, Greenspan H, Malik J: Blobworld 2002. "Image segmentation using expectation- maximization and its application to image querying" IEEE Transactions on Pattern Analysis and Machine Intelligence; 24(8): 1026-1038.
- [14] Montagnat J, Breton V, Magnin IE, 2003. "Using grid technologies to face medical image analysis challenges" Proceedings of the Third IEEE ACM International Symposium on Cluster Computing and the Grid; 588-93

- [15]. Lehmann, T., Guld, M., Thies, C., Fischer, B., Spitzer, K., Keyzers, D., Ney, H., Kohlen, M., Schubert, H., Wein, B, 2004. "Content-based image retrieval in medical Applications" *Methods of Information in Medicine* 43 pp. 354–361.
- [16]. Mattiea, M., Staib, L., Stratmann, E., Tagare, H., Duncan, J., Miller, P.: Pathmaster, 2000. "Content-based cell image retrieval using automated feature extraction" *Journal of the American Medical Informatics Association* 7 pp. 404–415
- [17]. G. D. Magoulas, A. Prentza, Machine learning in medical applications, in: G. Paliouras, V. Karkaletsis, C. D. Spyropoulos (Eds.), 2001. *Machine Learning and its Applications, Lecture Notes in Computer Science*, Springer-Verlag, Berlin, pp. 300-307.
- [18]. C. Brodley, A. Kak, C. Shyu, J. Dy, L. Broderick, A. M. Aisen, 1999. Content-based retrieval from medical image databases: A synergy of human interaction, machine learning and computer vision, in: *Proceedings of the 10th National Conference on Artificial Intelligence*, Orlando, FL, USA, pp. 760-767.
- [19]. M^uller, H., Rosset, A., Vall^ee, J., Geissbuhler, A, 2004. "Comparing feature sets for content-based medical information retrieval" In: *SPIE Medical Imaging*, San Diego, CA, USA.
- [20]. Sameer Antani, Rangachar kasturi and Ramesh Jain, 2002. "A survey on the use of pattern recognition methods for abstraction, indexing and retrieval of images and video" *Pattern Recognition*, Volume 35, Issue4, pp. 945-965
- [21]. H.D.Tagare, C. Jafe, J. Duncan, 1997. "Medical image databases: A content-based retrieval approach", *Journal of the American Medical Informatics Asssociation*,4 (3) pp. 184-198.
- [22]. Paul Miki Willy, Karl-Heinz Kufer, 2004. "Content-based Medical Image Retrieval (CBMIR): An Intelligent Retrieval System for Handling Multiple Organs of Interest", *proceedings of the 17th IEEE symposium on Computer-Based Medical Systems (CBMS'04)*.