Proceedings of the Fourth IASTED International Conference
VISUALIZATION, IMAGING, AND IMAGE PROCESSING
September 6-8, 2004, Marbella, Spain

LOCAL PATTERN MATCHING METHOD FOR IMAGE RETRIEVAL AND ITS APPLICATION TO DIGITAL MAMMOGRAMS

Toshiaki Nakagawa
Collaborative Center for Academy, Industry and Government, Gifu University
1-1 Yanagido, Gifu-shi 501-1193
Japan
nakagawa@fjt.info.gifu-u.ac.jp

Takeshi Hara, Hiroshi Fujita
Department of Intelligent Image Information, Division of Regeneration and Advanced Medical Sciences,
Graduate School of Medicine, Gifu University, Japan

Takuji Iwase
Department of Breast Surgery, Cancer Institute Hospital, Japan

Tokiko Endo
Department of Radiology, National Hospital Organization Nagoya Medical Center, Japan

ABSTRACT
We proposed a concept of content-based image retrieval and demonstrated the potential usefulness in mammography. The approach incorporated a local-pattern matching method based on Nth-order autocorrelation features with KL expansion (principal components analysis) to retrieve similar mass shadows on digitized mammograms. The method can perform image retrieval without carrying out the image segmentation. We confirmed the tendency that similar mass images were retrieved as the initial studies by using the 30 simulated patterns and the 75 images of mammographic masses. The result showed that the image retrieval method might provide a new CAD system on mammograms.

KEY WORDS
Image retrieval, Pattern matching, Computer-aided diagnosis, mammogram

1. Introduction

In recent years, large-scale databases of medical images were built for effective usage of information, and the technique was required to retrieve images from database as user's needs.
Comparing the previous and similar images with current case is very effective procedure in interpreting by physicians. The purpose of this work is to develop an image retrieval system based on local-pattern matching technique, and to evaluate the performance by employing mammographic masses.
This system searches out the image database for the image resembling to the entered one (query image) and outputs with diagnostic information. The information indicates whether the interpretation result is normal or abnormal, if the diagnosis is abnormal, the degree of malignancy is represented to determine the follow-up treatments.
Recent researches on image retrieval technique are mainly categorized into two major approaches by using languages as a keyword of an image and by using a feature value of an image. Since the data in words or by numerical values are given in advance in the approach based on keywords, it is not so practical when the number of image in the database became larger. Moreover, if features of images are difficult to express by only language, this method is not effective.
Approaches of using image as key well employ the feature-extraction method in many researches [1-3]. Also in the field of a medical image, some researches are reported on retrieval of the similar image of a disease until now [4, 5]. In general studies, the image segmentation technique took an important role to distinguish the abnormal region from the background area. Various features based on the shape and the size were calculated from the result of the segmentation there. Therefore, since the retrieval results depend on the accuracy of the segmentation processing, it is not an effective approach to retrieve images that the object boundaries are difficult to determine strictly. In mammographic masses especially, the segmentation is not suitable because a part of boundary of the shade is not clear, or if the shade is malignant mass, determining the margin is often difficult.
In this research, the technique of extracting the feature from the whole image automatically and performing image retrieval without carrying out the image

segmentation is proposed. To do this, we applied a local-pattern matching technique based on Nth-order autocorrelation features. The simulation experiment using the monochrome figures and the experiment using the mammographic mass images were conducted to examine the validity of this technique.

2. Methods

The flowchart of this method is shown in **Figure 1**.

2.1 Pre-processing

To raise the accuracy of retrieval results needs to perform the feature extraction using many local patterns from images with high resolution and high grey level. However, since those high qualities of image may be too fine in retrieval of similar images, even if shapes and density distributions of two images are slightly different, it is expected that the degree of similar may become low. Therefore, a pre-processing was performed to changes resolution and grey level into the optimal value judged from the amount of a database for the feature extraction.

2.2 Feature extraction by local pattern matching)

The features were extracted from the query image and the ones within whole database. The images including mass findings were simply cut from original digitized images in a square form. The autocorrelation features were extracted by counting patterns defined as local patterns consisted of some pixels with various grey level distributions. The example of the local pattern used for the feature extraction is shown in **Figure 2**. The number of pixels classified the local patterns into several levels. The number was corresponding to the level indicated in the figure. For instance, the local pattern at level 2 consists of two pixels, the number is considered as the combination of a pixel value of each pixel and the joint direction. Although the kind of these patterns exists innumerably, a pixel value is prepared in the range of the gradation of the target image. If the feature extraction is carried out using N pieces of local patterns on each image, the feature vector with N elements is obtained. The number of those local patterns is considered as the feature values of the image in N-dimensional space. The feature values of all of the images in database were calculated, and the feature was expressed as a vector in N dimensions.

2.3 Creation of the retrieval feature space using KL expansion

After storing the vector in each image, the KL expansion was used to decrease the number of dimension in feature space into from 4 to 10 axes to create the retrieval feature space that expressed effective feature values in retrieval using statistical character. The expanded feature vectors were newly stored to be retrieved. The accumulated rate of contribution of principal components becomes 90% or more was applied to reduction of the number of dimensions using KL expansion.

2.4 Calculation of the degree of similarity

To determine the similar image after the query one was given, the Euclid distance among the feature vectors measured on retrieval feature space evaluates the similarity between the query image and the others within the database. The nearest vector from the query image was employed as the retrieval result.

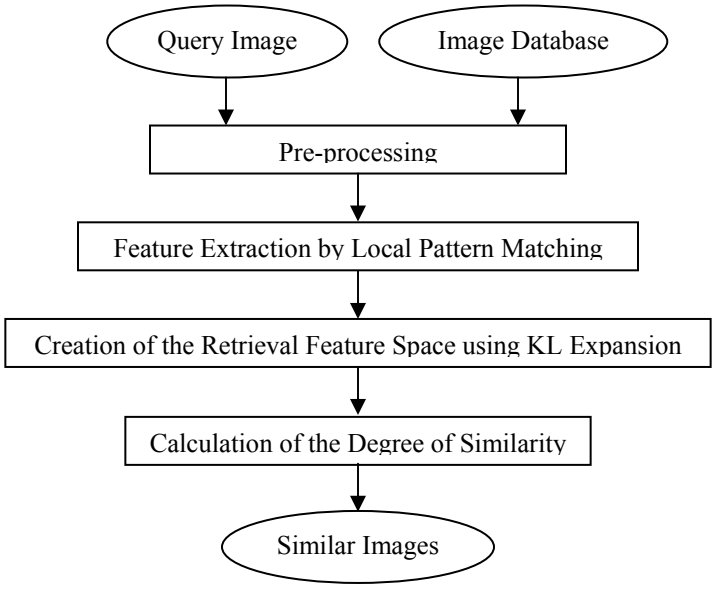

Figure 1 Flowchart of the Image Retrieval Scheme

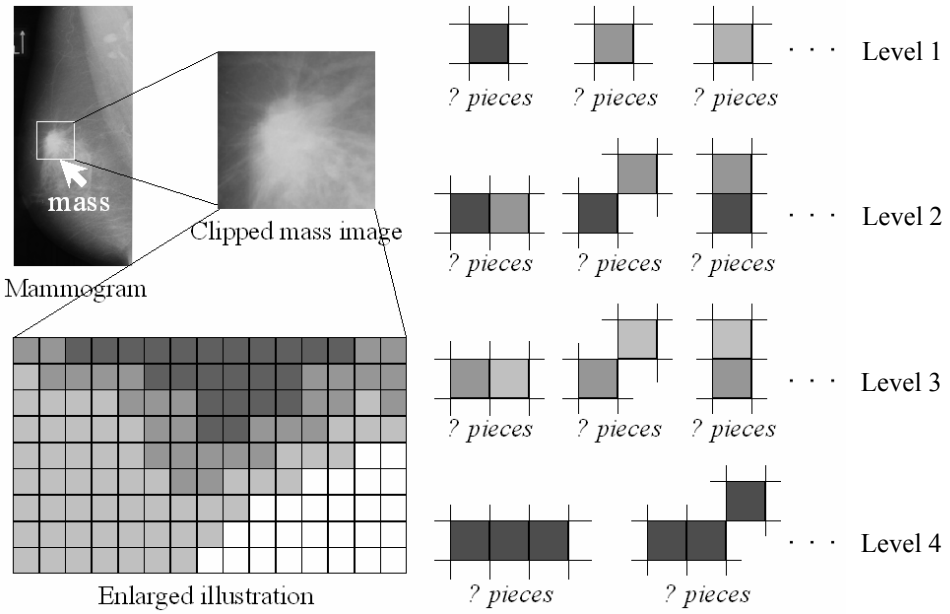

Figure 2 Feature Extraction by Local Pattern Matching

3. Simulation study

We evaluated the retrieval methods by using a database of 30 simulated patterns. Simulated-pattern group was based on handwriting and had 15-kind figure of two sheets each, and the same image did not exist within the data group.
The feature vector based on local-pattern matching was determined after the pre-process, and the local patterns that consisted of up to 2 were employed for the definition of local patterns. That is, sixty-four kinds (8 at level 1, and 56 at level 2) of local patterns were defined to calculate the feature vectors.
KL expansion was performed to the extracted amount of the features, and the N-dimensional retrieval feature space was created from the 1st to Nth principal components of which the rate of accumulation contribution became 90%. One image in which the Euclid distance had the shortest feature in the retrieval feature space was determined the similar image about each image.
As results of simulation study, it tended that the retrieved images were very similar to the query one in subjective opinions. The percentage of the correct answer was 77%.

4. Demonstration of similar image retrieval on mammographic masses

75 mass images digitized and clipped from mammogram taken by the screen/film system were used for similar image retrieval. Original digital data of the images had 4096 grey levels and 100-micrometer sampling pitch. To retrieval in a small-scale database, this quality of image might be too high. Therefore, the pre-process to changes grey level into 8 and resolution into 400 micrometers was performed for the mitigation of retrieval condition. The feature extraction based on local-pattern matching and KL expansion applied same procedure as the simulation study. One image was determined as a similar image to each image by the Euclid distance among the feature vectors measured on retrieval feature space.

5. Results

The retrieval results of the mammographic masses are shown in **Figure 3**. The results indicated in the figure were chosen with degree of similar by subjective observation from 75 sets of retrieval results. The images of four sheets are a query image, a retrieved one, the sketch of the query image drawn by the physician, and the sketch of the retrieved image in an order from the left. Although there were many satisfied cases with the difference of density distribution between those images, there were a few results correspond to physician's opinion on the shape and the condition of margin of mass. The rate of coincidence was 29% on the shape, and was 34% on the margin.

6. Discussion

There was a difference in the classification of mass shade between retrieval results and physician’s opinions. The reason of this was estimated that the signal-feature of an image is not in agreement with a physician's subjectivity measure. However, it can expect that the system to retrieve an image quantitatively by the degree of similar

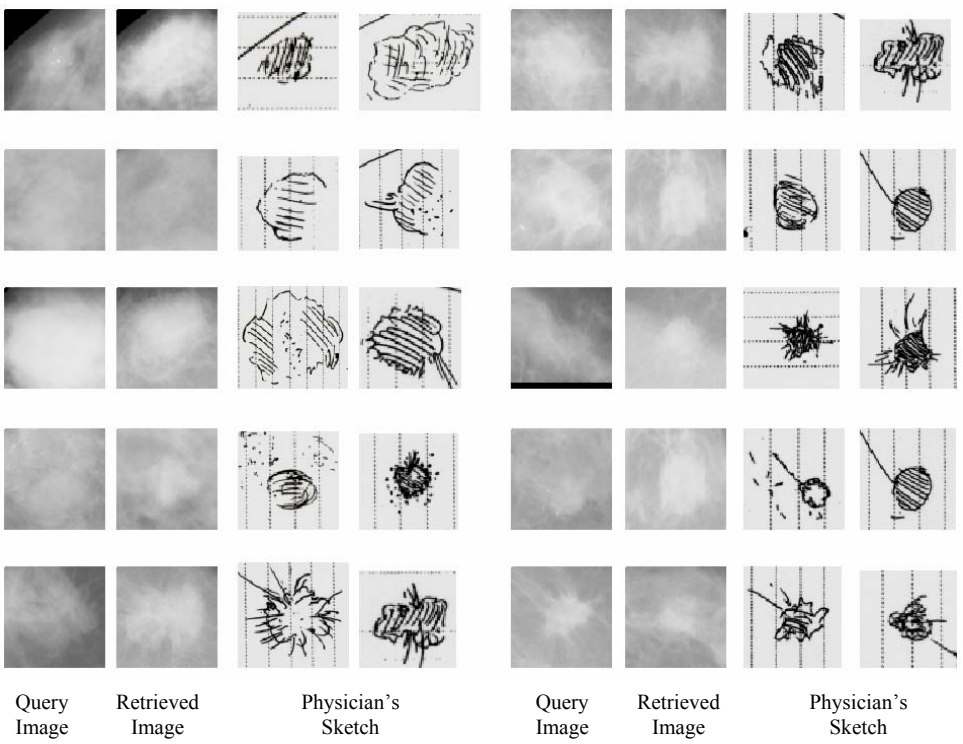

Figure 3 10 retrieval results of mammographic mass images. The images of four sheets are a query image, a retrieved one, the sketch of the query image drawn by the physician, and the sketch of the retrieved image in an order from the left.

based on the feature analysis, and can reduce the variation in diagnosis of physicians, and it is estimated that it may be able to use as diagnostic support. Even when it was an inaccurate solution, the retrieved image has a similar impression in many cases. It is estimated that the feature vectors in the retrieval feature space were reflecting the shape of an object image well as this reason. The feature extracted by our method is shift-in-variant in the object position, but it depends on the density (pixel value) variance and the rotation of the images.

7. Conclusion

The retrieval technique based on local patterns existing in images was proposed. The degree of similarity was estimated by the difference among the Euclid distance of the feature vectors in the retrieval feature space created by KL expansion analysis. We confirmed the tendency that similar mass images were retrieved as the initial studies by using the simulated patterns and the image of mammographic masses. It was concluded that our method was effective to retrieve grey-level patterns such as mammographic masses because image segmentation technique was not required and it did not depend on the accuracy of the determination of the margin.

Acknowledgements

This study was supported in part by the Grant-in-Aid for Scientific Research from the Japan Society for the Promotion of Science, the Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare, Japan, and a grant for the "Intellectual Cluster Creation Project" from the Ministry of Education, Culture, Sports, Science and Technology, Japan.

References:

[1] H. K. Lee, S. I. Yoo, Intelligent image retrieval using neural network, *IEICE Trans. Inf. & Syst.,12*, 2001, 1810-1819.

[2] D. L. Swets, J. J. Weng, Using discriminant eigenfeatures for image retrieval, *IEEE Trans. Pattern Analysis and Machine Intelligence, 18*(8), 1996, 831-836.

[3] T. Kurita, N. Otsu, T. Sato, A face recognition method using highter order local autocorrelation and multivariate analysis, *Proc. of 11th International Conference of Pattern Recognition (ICPR)*, 1 (II), Hague, 1992, 213-216.

[4] H. Qi, W. E. Snyder, Content-based image retrieval in picture archiving and communications systems, *Journal of Digital Imaging, 12*(2), 1999, 81-83

[5] M. L. Giger, Z. Huo, L. Lan, C. J. Vyborny, Intelligent search workstation for computer-aided diagnosis, *Proc. of the 14th International Symposium and Exhibition on Computer Assisted Radiology and Surgery (CARS2000)*, San Francisco, 2000, 822-827.