Steerable Texture Descriptor for an Effective Content-Based Medical Image Retrieval System Using PCA

B. Jyothi, Y. MadhaveeLatha, P.G. Krishna Mohan and V.S.K. Reddy

Abstract Digital images have increased in quantity especially in the medical field used for diagnostics. Content-Based Medical Image Retrieval System will retrieve similar medical images from large database based on their visual features like texture, color, and shape. This paper focuses a novel method to increase the performance using Boundary detection, Steerable filter, and Principal Component Analysis. The content of the image was extracted with the help of region-based texture descriptor using steerable decomposition followed by extracting Principle Component Analysis which has better feature representation capabilities. The similar medical images are retrieved by comparing the extracted feature vector of the given query image with the corresponding database feature vectors using Euclidian distance as a similarity measure. The effectiveness of the proposed method is evaluated and exhibited via various types of medical images. With the experimental results, it is obvious that the region-based feature extraction method outperforms the direct feature extraction-based image retrieval system.

Keywords Content-based image retrieval · Boundary detection · Texture · Shape features · Principle component analysis · Similarity measure

B. Jyothi (✉) · Y. MadhaveeLatha · P.G. Krishna Mohan · V.S.K. Reddy
Department of Electronics and Communication Engineering,
Malla Reddy College of Engineering and Technology, JNTUH, Hyderabad, India
e-mail: bjyothi815@gmail.com

Y. MadhaveeLatha
e-mail: madhaveelatha2009@gmail.com

P.G. Krishna Mohan
e-mail: pgkmohan@yahoo.com

V.S.K. Reddy
e-mail: mrcet2004@gmail.com

1 Introduction

In the medical field, everyday huge volumes of different types of medical images are produced for diagnostics therapy and also to provide great support to the doctors in computer-aided diagnosing centers and research. Nowadays, different types of medical images are produced in various hospitals as well as in various medical diagnostic centers [1]. Content-based image retrieval has many momentous applications especially in medical diagnosis, education, and research fields. These images are also important because the historical images of different patients in medical centers have valuable information for the upcoming diagnosis with a system which retrieves similar cases that make more accurate diagnosis and decide on appropriate treatment [2]. So there is a necessity for efficient and effective retrieval system.

Content-based image retrieval system (CBIRS) is defined as a process of searching a digital image from the large database on the basis of their visual features like shape, color, and texture. Now the main intention of this research work is to retrieve the most similar medical images matching the given query image from large medical databases using feature extraction and similarity measurement techniques. However, due to the nature of medical, they are represented in gray level rather than color.

Medical images are enriched with specific textural patterns. These patterns have specific information. Low resolution, blurring effect, and contrast results are difficult to identify the tissue patterns and structure of organs, and these are some causes of reduction in the retrieval output.

Feature extraction plays a key role in improving the performance of the medical image retrieval system. In this paper, we are concentrating on the development of a novel and effective approach for extracting content information about region. Medical images have poor boundaries. Hence, there are a few practical problems while acquiring the medical images during scanning such less illumination and noise. A boundary is a line that marks the limits or edges of an entity and separates it from the background. It helps us in object recognition and understanding the shape of an image. In the proposed method, we have extracted the boundary of an image followed by extracting texture feature [3]. Krit Somkath has introduced a technique to overcome the problem of boundary detection for ill-defined edges using edge following algorithm defined in [4] based on intensity and texture gradient features which works better in detecting boundaries, and hence we have adapted it. We extend the framework to extract the content by means of steerable filter at various orientations followed by extracting texture features using principle component analysis. Finally, similar medical images are retrieved by matching the region features with the database images.

The rest of this paper is structured as follows. Section 2 describes the CBMIR system. Section 3 explains boundary detection procedure. Section 4 focuses on features extraction algorithm. Section 5 explains the experimental results. Section 6 concludes the paper.
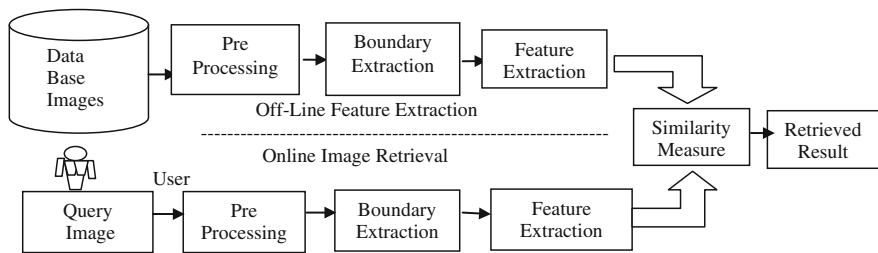
Fig. 1 Content-based medical image retrieval system

2 Proposed CBMIR System Architecture

The block diagram of a typical conceptual content-based retrieval system is illustrated in Fig. 1. It consists of two phases: offline feature extraction phase and online image retrieval phase. In offline phase, the images stored in the database are pre-processed by image de-noising. Next, the boundary of the images is detected followed by extracting the visual content of the image. The texture features of the database images are extracted and described with feature vectors and stored in the database. The database consists of various classes of medical images characterized by certain objects such as liver, body outline, and spine for CT or MRI images. The same process is repeated for query image in online phase.

In online image retrieval process, the user submits query images as an example to the retrieval system for searching similar images. The system retrieves related images by computing the similarity matching between the feature vectors of the query image and those of the database images. Finally, the system returns the results which are most relevant to the query image.

3 Boundary Detection Approach

Medical images consist of many objects which can be identified with the help of proper boundaries. In this approach, the boundary of the given query image and database images are detected using edge following algorithm which provides complete information of an image by considering both magnitudes and directions [5].

3.1 Gradient Vector Field Model

The magnitude and directions of given image $f(x, y)$ are calculated according to the following equations in x and y directions, respectively:

$$H(i,j) = \frac{1}{H_r} \sum_{(i,j) \in N} \sqrt{H_x^2(i,j) + H_y^2(i,j)} \tag{1}$$

$$A(i,j) = \frac{1}{H_r} \sum_{(i,j) \in N} \tan^{-1}\left(\frac{H_y(i,j)}{H_x(i,j)}\right) \tag{2}$$

$$H_x(i,j) = -G_y x f(x,j) \approx \frac{\partial f(x,y)}{\partial y} \tag{3}$$

$$H_y(i,j) = -G_x x f(x,j) \approx \frac{\partial f(x,y)}{\partial x} \tag{4}$$

where G_x and G_y are the Gaussian weight difference mask in the x and y directions and H_r is the total number of pixels in the 3×3 window neighborhood N. The idea of boundary extraction in unclear images is shown in Fig. 2.

3.2 Edge Following Technique

Applying from Law's texture followed by double gradient detects, edges of objects in an image are derived from Law's texture followed by canny edge detection. The texture feature of images is computed by convolving an input image $F(x, y)$ with
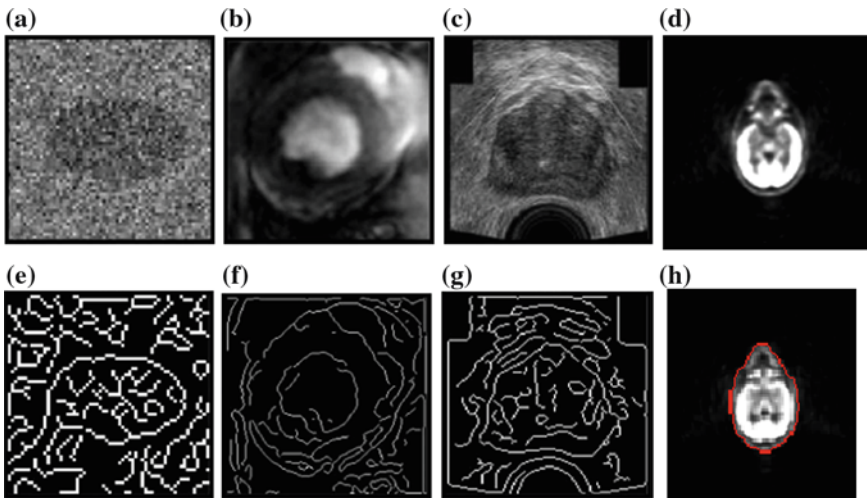

Fig. 2 a Synthetic noisy image. **b** Left ventricle MR image. **c** Ultrasound image. **d** Brain image. **e–f** Corresponding boundary detection with the help of edge maps derived from Law's texture

each of the texture mask $T(x, y)$ which is defined as law's texture [6] and canny detection on the resultant image $R(x, y)$:

$$T(x, y) = L.L^T \tag{5}$$

$$L = (1, 4, 6, 4, 1)^T \tag{6}$$

$$R(x, y) = F(x, y) * T(x, y) \tag{7}$$

The average gradient vector field and edge map at the position (i, j) of an image are exploited to find the boundary of an image:

$$L_{ij}(r, c) = \alpha H_{ij}(r, c) + \beta A_{ij}(r, c) + \varepsilon R_{ij}(r, c)$$

$$0 \leq r \leq 2, 0 \leq c \leq 2 \tag{8}$$

$$H_{ij}(r, c) = \frac{H(i + r - 1, j + c - 1)}{\text{Max}_{i,j} H_{i,j}};$$

$$A_{ij}(r, c) = 1 - \frac{|A(i, j) - A(i + r - 1, j + c - 1)|}{\pi};$$

$$R_{ij}(r, c) = R(i + r - 1, j + c - 1), 0 \leq r \leq 2, 0 \leq c \leq 2$$

where α, β, and ε are the weights; $H(i, j)$ and $A(i, j)$ are the magnitude and average of gradient edge vector fields; and $R(i, j)$ is the edge map from Law's texture.

4 Features Extraction

A feature is defined to express certain visual properties of an image, either locally for a small group of pixels or globally for the entire image. The contents of an image were described with the help of their features. In the proposed system, the local feature extraction is used in spite of extracting whole image. The input image is $F(x, y)$, where x and y are the pixel coordinates in the image. The image boundary is detected using edge following method, and corresponding region-based texture features are extracted in different orientations followed by extracting principal component analysis. In this section, we will discuss statistical region-based feature extraction in detail.

4.1 Texture Features Extraction

Medical images are mostly represented in gray level, and most medical image surfaces exhibit texture. Texture is a natural surface property, and it has repeated

pixel information regarding the structural arrangement and it also gives the relationship between the surface and external environment. In this paper, we implemented a rotation-invariant and scale-invariant texture representation for image retrieval based on steerable filter decomposition defined in [7].

Steerable filter is a special class of filter which is synthesized at different orientations as a linear combination of basis filters. It allows one to adaptively steer a filter to any orientation and to determine the filter output as a function of orientation.

The steering constraint is

$$F_\theta(m, n) = \sum_{k=1}^{N} b_k(\theta) A_k(m, n) \qquad (9)$$

where $b_k(\theta)$ is the interpolation function based on the arbitrary orientation θ and $A_k(m, n)$ are the basis filter rotated version of impulse response at θ. An image at an arbitrary orientation was synthesized by taking linear combination of the image filtered convolved with the basis filters.

The steerable filter structure is illustrated in Fig. 3.

Texture information can be extracted by applying steerable filter in various oriented sub-bands. In this paper, we extract the texture features from six oriented sub-bands (0°, 60°, 120°, 180°, 240°, 300°), as shown in Fig. 4.

$$B_i(x, y) = \sum_{x_1} \sum_{y_1} I(x_1, y_1) P_i(x - x_1, y - y_1) \qquad (10)$$

where $B_i(x, y)$ and P_i denote the horizontal and directional band pass filters at orientation ($i = 1, 2, 3, 4, 5, 6$).
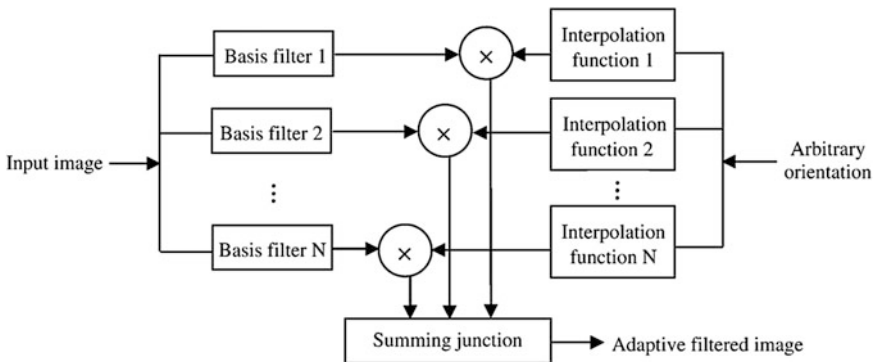

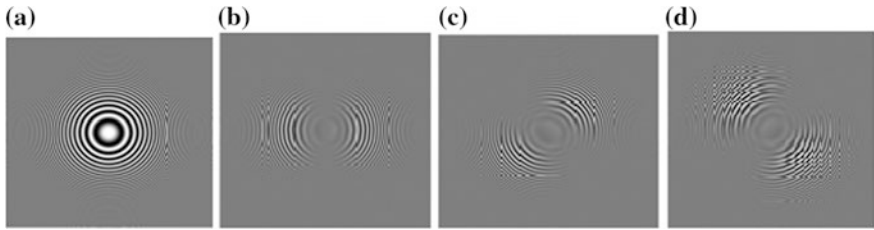**Fig. 3** Structure of steerable filter

Fig. 4 The steerable filter image at different orientations. **a** The original image; **b** the image in horizontal orientation; **c** the image for rotation of 45°; **d** the image for rotation of −45°

4.2 Texture Representation Using PCA

The principal component analysis (PCA) is one of the most unbeaten techniques that have been used in image recognition applications. PCA is a statistical linear transformation method, used for the purpose of large dimensionality reduction of the data space.

PCA is linear transformation technique which is a variable reduction method and generally useful when data have high redundancy. This will result into reduction of information into smaller size which is called principal components. PCA is applied on filtered output at different orientations to reduce the dimensionality of a huge dataset [8]

The following steps give the detailed description of the mathematical calculation of the principal component analysis:

(1) A 2-D image pixels can be represented as 1-D vector by re-arranging each column or row into a single vector and pj's represent the pixel values:

$$X_i = [A_1 . \ldots \ldots \ldots \ldots . A_n], \; i = 1 . \ldots . . M. \tag{11}$$

(2) The mean image was subtracted from each image vector to center the images. Let C represents the mean image:

$$C = \frac{1}{M} \sum_{k=1}^{N} X_i \tag{12}$$

The mean centered image is defined as

$$W_i = X_i - C \tag{13}$$

(3) Covariance matrix of the dimensions m × m for mean centered images is calculated:

$$D = WW^T \tag{14}$$

(4) The eigen values were computed:

$$Det(D - \lambda I) = 0 \tag{15}$$

The nonzero eigenvalues was sorted from high to low where most image data can be represented.

Algorithm for Content-Based Image Retrieval Scheme

Step 1 Input query medical image.
Step 2 Convert image from RGB mode to Gray.
Step 3 Preprocess the image with the help of median filter that suppresses undesired distortions and enhances image features which are very important for further processing
Step 4 Extract the boundary of an image with the help of gradient vector image model and edge map techniques.
Step 5 Apply steerable filter at six orientations and principal component analysis for extracting texture features.
Step 6 For retrieving the similar images, these features were compared with the corresponding features in database.
Step 7 Rank the images according to the sorted virtue distances and display the top 50 images.

5   Experimental Results and Discussion

The proposed method with the input dataset consists of the various images of different modalities. These images were acquired from Frederick National Laboratory for Cancer Research, which is open to all to provide an essential medical imaging research resource for supporting collection of data for private and internal projects and publishing results by project investigators. Our database consists of 1000 medical images of various classes of X-Ray, CT, and MRI scanned images such as chest, lungs, brain, abdomen, etc. We have tested the performance and effectiveness of the proposed approach. In the preprocessing stage, the noise in the images was corrected using median filter followed by segmenting the image for boundary extraction followed by extracting multiple features. The same process was repeated with the database images and query image. Similar medical images have been retrieved by computing the Euclidian distance [9] between the given query image feature vector and corresponding feature of the database images which is defined as

$$D(p_1, p_2) = \sqrt{\sum_{\forall i} (p_1(i) - p_2(i))^2} \qquad (16)$$

The overall performance of a proposed CBMIR system is evaluated by measuring recall rate, mean average precision (MAP), and Error Rate, which have been employed in [10]. The performance of the proposed medical image retrieval system is compared with other systems which are based on various individual features and without extracting boundary of an image. From the experimental results as shown in the Table 1, it is clear that the proposed approach is more prominent for the content-based medical image retrieval as it compares with the existing system. This high performance is due to the combination of the shapes by extracting boundary of an image followed by extracting texture features.

$$\text{Recall Rate} = \frac{\text{Number of Relavant Images Retrieved}}{\text{Total Number of Relavant Images in Database}} \qquad (17)$$

$$\text{Mean Average Precision} = \frac{\sum P(\text{NR})}{N} \qquad (18)$$

where P (NR) is the precision after first NR images are retrieved.

$$\text{Error Rate} = \frac{\text{Number of Non Relavant Images Retrieved}}{\text{Total Number of Images Retrieved}} \qquad (19)$$

The retrieval performance of the proposed CBMIR method was compared with existing feature extraction methods such as GLCM, LLBP, Tumura, and Gobor Features discussed in [11, 12]. Compared to all the above-mentioned techniques, the proposed feature extraction method gives better retrieval performance which works effectively for computer-aided diagnosis retrieval applications which are very effective for various image databases.

Table 1 Comparison of the recall rate, error rate, and mean average precision measures for various methods

Feature extraction technique	Performance measure using without boundary extracted features			Performance measure using with boundary extracted features		
	Recall rate (%)	Error rate (%)	MAP (%)	Recall rate (%)	Error rate (%)	MAP (%)
1. GLCM	42.12	58.88	56.11	43.58	57.42	58.56
2. Tamura	41.23	59.77	57.81	43.78	57.22	59.23
4. LLBP features	43.65	57.35	64.93	50.34	50.66	72.67
5. Gobor features	68.43	32.57	73.45	73.36	26.64	76.36
9. Proposed method Steerable + PCA	73.22	26.78	78.47	76.54	23.46	81.58

6 Discussion and Conclusion

Feature extraction is a significant task in efficient medical image retrieval. We have proposed a novel approach for extracting region features from the object of an image. It can easily be implemented and is computationally more effective than the traditional methods. In this paper, we have projected a new algorithm in which boundary of an image can be detected with the help of edge following algorithm followed by extracting region-based texture features such features with the help of steerable filter followed by extracting principal component analysis. The results have shown that the proposed approach improves the retrieval performance compared to the existing system. In future, the proposed scheme can further be improved by incorporating multiple features and relevance feedback.

References

1. Müller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications-clinical benefits and future directions. Med. Inform. **1**, 73 (2004)
2. Khoo, L.A., Taylor, P., Given-Wilson, R.M.: Computer-aided detection in the United Kingdom national breast screening programme: prospective study. Radiology **237**, 444–449 (2005)
3. Chun, Y.D., Kim, N.C., Jang, I.H.: Content-based image retrieval using multiresolution color and texture features. IEEE Trans. Multimedia **10**(6), 1073–1084 (2008)
4. Somkantha, K., Theera-Umpon, N.: Boundary detection in medical images using edge following algorithm based on intensity gradient and texture gradient features. Proc. IEEE Trans. Biomed. Eng. **58**(3), 567–573 (2011)
5. Veeralakshmi, S., Sivagami, S.V., Devi, V.V., Udhaya, R.: Boundary exposure using intensity and texture gradient features. IOSR J. Comput. Eng. (IOSRJCE) ISSN: 2278-0661, **8**(1), 28–33 (Nov–Dec 2012). ISBN: 2278-8727, www.iosrjournals.org
6. Jacob, M., Unser, M.: Design of steerable filters for feature detection using canny like criteria. IEEE Trans. Pattern Anal. Mach. Intell. **26**(8), 1007–1019 (2004)
7. Wang, X.-Y., Yu, Y.-J., Yang, H.-Y.: An effective image retrieval scheme using color, texture and shape features. Comput. Stan. Interfaces CSI-02706 **33**(1), 59–68 (2011)
8. Navaz1, A.S.S., Dhevi sri, T., Mazumder, P.: Face recognition using principal component analysis and neural networks. Int. J. Comput. Netw. Wirel. Mobile Commun. **3**(1), 245–256 (Mar 2013). ISSN: 2250-1568
9. El-Naga, I., Yang, Y., Galatsanos, N.P., Nishikawa, R.M., Wernick, M.N.: A similarity learning approach to content-based image retrieval: application to digital mammography. IEEE Trans. Medical Imaging **23**(10), 1233–1244 (2004)
10. Arevalillo-Herráez, M., Domingo, J., Ferri, F.J.: Combining similarity measures in content-based image retrieval. Pattern Recognit. Lett. **29**, 2174–2181 (2008)
11. Rasli, R.M, Muda, T.Z.T., Yusof, Y.: Comparative analysis of content based image retrieval techniques using color histogram: a case study of GLCM and K-means clustering. In; 3rd International Conference on Intelligent Systems, Modelling and Simulation (ISMS), pp. 283–286 (2012). ISBN: 978-1-4673-0886-1
12. Dubey, R.S., Choubey, R., Bhattacharjee, J.: Multi feature content based image retrieval. Int. J. Comput. Sci. Eng. **02**(06) (2010). ISSN : 0975-3397 2145 2145-2149