CrossMark

A unified learning framework for content based medical image retrieval using a statistical model

K. Seetharaman, S. Sathiamoorthy *

Department of Computer Science and Engineering, Annamalai University, Annamalai Nagar 608 002, Tamil Nadu, India

Abstract This paper presents a unified learning framework for heterogeneous medical image retrieval based on a Full Range Autoregressive Model (FRAR) with the Bayesian approach (BA). Using the unified framework, the color autocorrelogram, edge orientation autocorrelogram (EOAC) and micro-texture information of medical images are extracted. The EOAC is constructed in HSV color space, to circumvent the loss of edges due to spectral and chromatic variations. The proposed system employed adaptive binary tree based support vector machine (ABTSVM) for efficient and fast classification of medical images in feature vector space. The Manhattan distance measure of order one is used in the proposed system to perform a similarity measure in the classified and indexed feature vector space. The precision and recall (PR) method is used as a measure of performance in the proposed system. Short-term based relevance feedback (RF) mechanism is also adopted to reduce the semantic gap. The Experimental results reveal that the retrieval performance of the proposed system for heterogeneous medical image database is better than the existing systems at low computational and storage cost.

© 2015 The Authors. Production and hosting by Elsevier B.V. on behalf of King Saud University. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Medical images play a vital role in disease analysis, education, research, etc. Evolution of computer vision and digital imaging modalities has generated a vast amount of digital images in the medical domain. Consequently, the task of retrieving heterogeneous medical images from a large-scale image database becomes more difficult than ever before for a computer vision system due to the noise, variation in size, shape, color, illumination, etc. Hence, it is necessary to build up an appropriate system for medical image retrieval with efficient storage and effective retrieval to assist the physicians.

The conventional text-based image retrieval systems use textual keywords that are manually annotated on images. With the vast and diversity of images, textual keywords hold the disadvantages of laborious, tedious and time-consuming. Moreover, the manual annotation of the images strongly depends on what the users focus on and it may vary between persons, and also vary in time for the same person. Thus, textual keywords are inefficient in providing sufficient and distinctive discriminatory power of the images (Stricker and Orengo, 1995; Rui et al., 1999). The picture archival and communication

* Corresponding author. Mobile: +91 (0) 9994029213.
E-mail address: ks_sathia@yahoo.com (S. Sathiamoorthy).
Peer review under responsibility of King Saud University.

system (PACS) (Müller et al., 2004) compliant with digital imaging and communications in medicine (DICOM) format is used by most hospitals to handle huge collections of medical images. The PACS uses textual information stored in the DICOM header such as patient identity, date, type of examinations, modality, body parts examined, etc. for the image retrieval operations. The literature reveals (Müller et al., 2004) that the DICOM headers include a high rate of errors and storing the DICOM format images in any of the compressed formats such as JPEG, TIFF, etc. leads to loss of DICOM header information. In order to improve the performance of PACS, content based image retrieval (CBIR) techniques have been proposed by several researchers in the PACS environment (Müller et al., 2004). Consequently, a number of researches have been focused on content based medical image retrieval (CBMIR) to facilitate the physicians such as CBMIR system for HRCT images of the lung (Shyu et al., 1999), PET images of the human brain (Cai et al., 2000), X-ray images of the human cervical and lumbar spines (Long and Thoma, 2001), histological images of GI tract (Tang et al., 2003), CT images of chest (Yu and Chiang, 2004), a PathMiner system for pathological images (Chen et al., 2005), X-ray images of spine (Hsu et al., 2009), X-ray images of chest (Avni et al., 2010), mammogram images of breast (Wei et al., 2011), etc. However, the aforesaid literature is distinct to modalities, biological system, body orientation etc.

In the last two decades, only a little effort has been taken to develop a framework for heterogeneous medical image retrieval. For instance, Orphanoudakis et al. (1994) reported I^{2}C system based on object oriented architecture and it uses global level image features for image indexing, storage and retrieval. KMeD (Knowledge based multimedia medical distributed database) system presented by Chu et al. (1998) exploits four semantic layers for information modeling and it uses shape, texture and alphanumeric information with spatial and temporal constructs for image retrieval. El-Kwae et al. (2000) introduced a system called COBRA, which is an open architecture based on widely used healthcare and technology standards and it improves the capability of PACS by exploiting global level color, shape and texture features extracted from the automatic segmented image regions. The MedGIFT/GIFT (GNU Image Finding Tool) (Müller et al., 2005) is also an open source framework for medical image retrieval. It uses textual features and visual features such as Gabor blocks, histogram of Gabor filter responses, color blocks and global color histogram for image retrieval. The IRMA (Image Retrieval for Medical Applications) system presented by Lehmann et al. (2005) and Güld et al. (2007) employs six semantic layers to formalize the local level features and their spatial relationship. Rahman et al. (2007) presented a probabilistic multi-class support vector machine (SVM) learning based image pre-filtering scheme with the combination of statistical similarity measure and relevance feedback (RF) mechanism. Later, a retrieval framework based on feature and similarity level fusion is proposed by Rahman et al. (2008). Subsequently, Rahman et al. (2009) developed a framework in which the images are represented in local visual and semantic concept based feature spaces toward semantic based image retrieval by utilizing the self-organizing map (SOM) and multi-class SVM. Furthermore, an advance in feature representation and similarity matching techniques of the aforesaid work (Rahman et al., 2009) is reported in Rahman et al. (2011).

Recently, a multimodal hierarchical modality classification approach for image filtering system is reported by Rahman et al. (2013) for retrieving heterogeneous medical images and it uses color layout descriptor (CLD), edge histogram descriptor (EHD), color moments, gray level co-occurrence matrix (GLCM), edge frequency, primitive length, Gabour moments, Tamura moments, Scale Invariant Feature Transform (SIFT), Local binary pattern (LBP), LBP-I, color and edge directivity descriptor (CEDD), fuzzy color and texture histogram (FCTH), autocorrelation coefficient as visual features and "Bag of words" as textual feature to represent the image. The "Bag of words" contains image-related text such as title, modality, region of interest, problem, anatomy of the image, etc.

Subsequently, Sudhakar and Bagan (2014) described a heterogeneous medical image retrieval framework, which performs phase congruency process in L*a*b* color space to extract edge co-operative maps and is processed using the SIFT to drive keypoints. The extracted keypoints are quantized to build a codebook using Spherical Self-Organizing Map (SOM) built with a geodesic data structure.

Although many research works have been contributed to CBMIR, the retrieval accuracy of the existing CBMIR systems for heterogeneous medical image database is still limited and unsatisfactory due to the lack of techniques used to extract the efficient and effective features of the medical images.

Though the system presented by Rahman et al. (2013) has shown significantly better results when compared to the systems previously reported in the literature, the high dimensionality of the feature vector results in high computational and storage cost. Correspondingly, even if the system reported by Sudhakar and Bagan (2014) uses low feature vector dimension than the system of Rahman et al. (2013), it results in less retrieval accuracy and the feature vector dimension is also not significantly less.

The main objective of the proposed system is to develop a more accurate, and cost effective systems for supporting physicians. Thus, the proposed system employed the color autocorrelogram (Huang et al., 1997; Chun et al., 2008; Penatti et al., 2012) to represent color and its spatial information, edge orientation autocorrelogram (EOAC) to represent shape and its spatial information and micro-textures (Seetharaman and Palanivel, 2013) to represent global and local level texture information of the gray-scale and color medical images respectively. The proposed features are extracted automatically by using a framework based on FRAR model with the BA (Seetharaman and Krishnamoorthi, 2007; Seetharaman and Palanivel, 2013), which avoids the cumbersome process owing to the complementariness of techniques used for extracting various kinds of visual features.

Mahmoudi et al. (2003) constructed EOAC using the edges detected by the Sobel edge detector, which is sensitive to noise and fails to detect very minute and fine edges. Moreover, the Sobel edge detector extracts the edges of color images from its gray-scale version, which leads to loss of some edges due to chromatic changes and also detecting the edges from H or S or V component image loses some edges due to the spectral variations (Liu et al., 2011), which might play a significant role in efficient medical image retrieval.

Medical images are more complex in structure and diverse in the collection. Specifically, the microscopic images are far more complicated and diverse than other types of medical images and contains many different and minute cytological

components. Thus, the proposed system constructs EOAC by extracting the very minute and fine edges using a framework based on FRAR model with BA (Seetharaman and Krishnamoorthi, 2007) in HSV color space (Liu et al., 2011) for color medical images and in gray-scale space for gray-scale images.

Artificial neural networks have been used broadly in the medical domain for classification problems. Many researchers have found that they have more flexibility in modeling and reasonable accuracy in classification problems. The support vector machine (SVM) based on statistical learning theory has been used extensively for non-linear and non-separable problems in various domains, like medicine, engineering, etc. due to its high generalization ability, robustness to high-dimensional data and a well-defined learning theory. Chen et al. (2009) reported that adaptive binary tree based multi-class support vector machine (ABTSVM) is a tree based approach of one-against-one (OAO) method and shows better results when the databases are more unbalanced, achieves a fast classification and maintains higher classification accuracy of SVMs. Since, heterogeneous medical image data set is more unbalanced, we adopted an ABTSVM method for the proposed multi-class classification of medical images in feature vector space. Subsequently, the classified feature vectors are indexed using the kD-tree (White and Jain, 1996) indexing method. The classification and indexing process filters out the irrelevant images and reduces the search space considerably. The RF technique in short-term learning (Squire et al., 1998; Nezamabadi-pour and Kabir, 2009) is adopted to reduce the semantic gap between the low-level visual features and high-level semantic concepts perceived by the physicians. The Manhattan distance measure (Huang et al., 1997) of metric order one is used to measure the similarity between the query and target images in the classified and indexed feature vector space. The experimental results confirm that retrieval performance of the proposed system is significantly better than the existing systems (Rahman et al., 2013; Sudhakar and Bagan, 2014) at low computational and storage cost.

The rest of this paper is organized as follows. FRAR model is described in Section 2. The proposed image retrieval system is discussed in Section 3. Section 4 explains experimental results and discussion. Finally, the conclusion is formulated in Section 5.

2. FRAR model

Recent literature reports that a framework based on FRAR model (Seetharaman and Krishnamoorthi, 2007; Seetharaman and Palanivel, 2013) outperforms the existing methods in terms of capturing the edge and texture features of gray-scale images. Thus, the proposed CBMIR system is designed so as to use the effectiveness of a framework based on FRAR model with the BA for extracting texture information at global and local level and shape and its spatial information of color and gray-scale medical images, and the same framework is used to extract color and its spatial information also.

Let X be a random variable that represents the intensity value of a pixel at location (k, l) in an image of size $L \times L$. The FRAR model (Seetharaman and Krishnamoorthi, 2007; Seetharaman and Palanivel, 2013) is expressed in Eq. (1):

$$X(k,l) = \sum_{\substack{p=-\frac{M}{2} \\ p=q\neq 0}}^{\frac{M}{2}} \sum_{q=-\frac{M}{2}}^{\frac{M}{2}} \Gamma_r X(k+p, l+q) + \varepsilon(k,l) \quad (1)$$

where $\Gamma_r = \dfrac{K \sin(r\theta) \cos(r\phi)}{\alpha^r}$ and $r = |p| + |q| + M(M-1)/2$

The initial assumption about the parameters is $K \epsilon R$; $\alpha > 1$; θ and $\phi \epsilon [0,2\pi]$. In Eq. (1), $X(k+p, l+q)$ is the spatial variation due to image properties and $\varepsilon(k,l)$ is the spatial variation due to additive noise and FRAR model coefficients Γ_r, $(r = 1, 2)$ are the variation among the low-level primitives in the sub-image region of size $M \times M$, $(M < L)$. The model coefficients are interrelated. The interrelationship is established through the model parameters K, α, θ, and ϕ, which are estimated using the BA (Seetharaman and Krishnamoorthi, 2007; Seetharaman and Palanivel, 2013).

3. Proposed image retrieval system

3.1. Overall structure of the proposed system

The architecture of the proposed CBMIR system is shown in Fig. 1. The heterogeneous medical image database consists of both gray-scale and color images produced by various modalities for various biological system and body orientations. In the proposed system, the color medical images are in RGB color space and they are converted into HSV color space (Karkanis et al., 2003) then the images are segregated into I_H (Hue), I_S (Saturation) and I_V (Intensity) component images, where I_H and I_S component images have chromatic information and I_V component image contains achromatic information. Then, the color autocorrelogram is computed from the uniformly quantized I_H and I_S images respectively, and is formed as color feature vector F_C of dimension D_C, the micro-texture feature vector F_T of dimension D_T is formed by extracting the micro-textures from the I_V image. Subsequently, the images in the HSV color space are converted from the cylindrical co-ordinate system (H, S, V) to Cartesian co-ordinate system (H′, S′, V′). Then, the image in the Cartesian co-ordinate system is segregated into $I_{H'}$, $I_{S'}$ and $I_{V'}$ component images. The proposed system next extracts the edges from each component image and the detected edges of $I_{H'}$, $I_{S'}$ and $I_{V'}$ are processed together (Liu et al., 2011) to compute the orientation of edges, which is used to construct the proposed EOAC of dimension D_S.

For gray-scale images, the micro-textures and EOAC are computed using the proposed framework. Then, the color autocorrelogram is computed using the proposed framework from the uniformly quantized gray-scale medical images.

The system combines and normalizes the extracted feature vectors to generate a query feature vector Q_F and it is classified using the ABTSVM to predict the class label of query feature vector. Subsequently, the system calculates the similarity between the query feature vector Q_F and each target feature vector Q_T in the corresponding class space. Finally, the system retrieves a list of images based on the similarity ranks. If the physician is not satisfied with the retrieved results due to semantic gap, RF mechanism
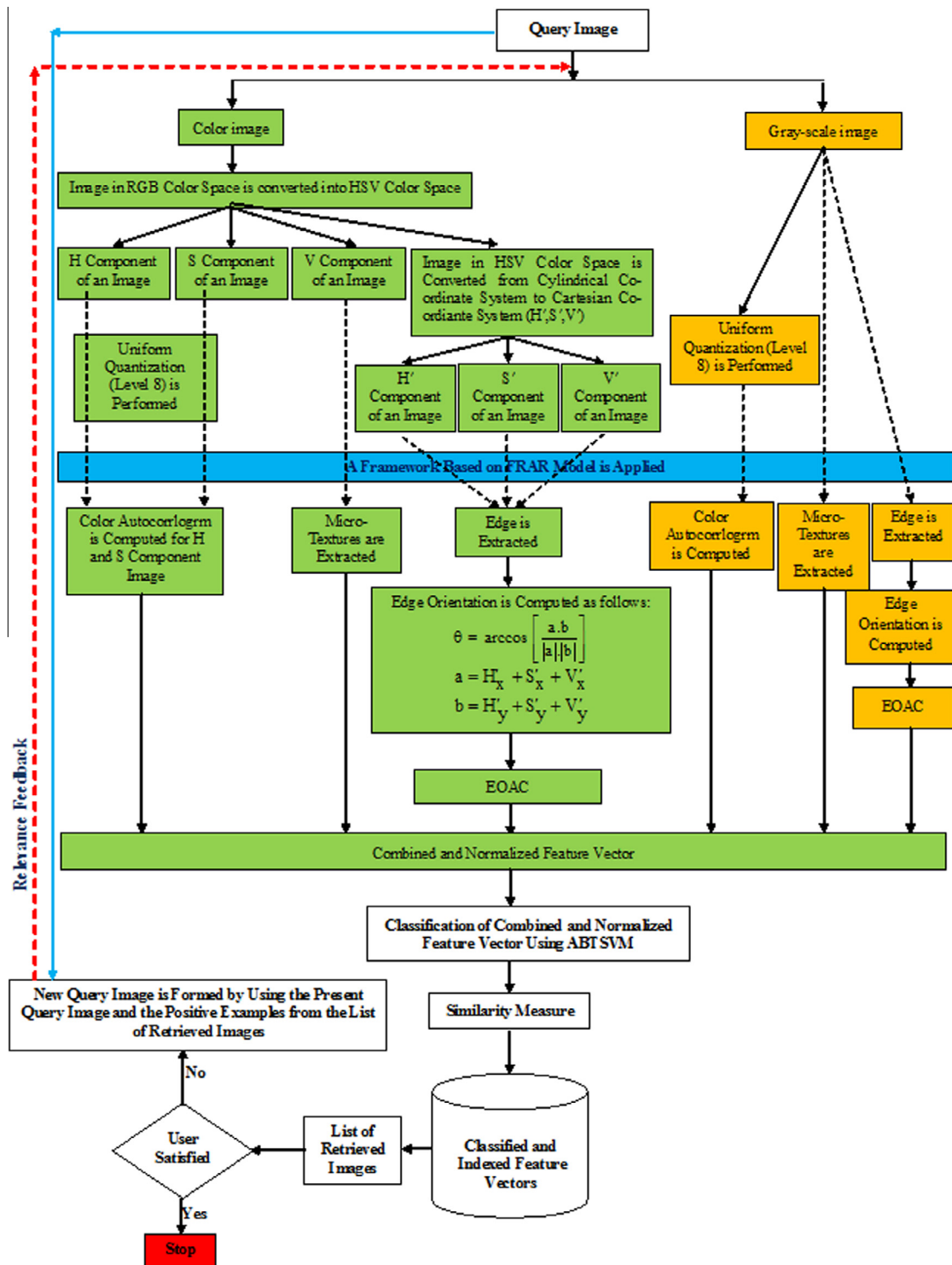
Figure 1 Architecture of the proposed medical CBMIR system.

captures the physician preferences and bridges the semantic gap and it is performed by using a new query image, which is formed by combing the query image used in the last search session and the images that are marked by the users as positive examples in the last retrieved results. The RF processes continuously until the results are closer to the physician perception.

3.2. Feature extraction

3.2.1. Color feature

In the proposed system, the color feature is extracted from the I_H and I_S component images. The I_H and I_S component images are uniformly quantized using the Generalized Lloyd algorithm (Chun et al., 2008). The number of quantization level

(L) is fixed to 8 (Chun et al., 2008). The autocorrelation coefficients (α^k) are derived from the FRAR model coefficients Γ_r (Seetharaman and Palanivel, 2013) as follows:

$$(\alpha^1) = \frac{\Gamma_1}{1 - \Gamma_2};$$

$$(\alpha^2) = \frac{\Gamma_1^2 + \Gamma_1 - \Gamma_2^2}{1 - \Gamma_2}$$

and

$$(\alpha^3) = \frac{\Gamma_1(\Gamma_1^2 + 2\Gamma_1 - \Gamma_2^2)}{1 - \Gamma_2}$$

Similarly, the kth order autocorrelation coefficient can be obtained by solving the following equation using recurrence relation.

$$\alpha^k = \Gamma_1 \alpha^{k-1} + \Gamma_2 \alpha^{k-2}; 1 \leqslant k \leqslant m; \qquad (2)$$

where m is the lag variable and k is the order of autocorrelation coefficient. In the proposed system, the color autocorrelogram, which captures the spatial correlation between the identical colors at a distance 1 is computed for I_H and I_S component images respectively. Since the smallest distance gave the most detailed local properties of the image, the proposed system fixes it to 1. The extracted color feature is formed as color feature vector (F_c) and is defined as:

$$F_c = [(\alpha_c^k(I_H)), (\alpha_c^k(I_S))]$$

$$F_c = [F_c^H, F_c^S]$$

where $\alpha_c^k(I_H)$, $\alpha_c^k(I_S)$ and F_c^H, F_c^S are the color autocorrelogram and color feature vector of I_H and I_S component images respectively. The dimension (D_S) of the extracted color feature vector is 16 (F_c^H : 8 and F_c^S : 8).

3.2.2. Texture feature

The I_V component image is divided into a number of overlapping sub-image regions of size 5×5, to locally characterize the nature of the image. The FRAR model uses linear estimates of pixel's gray tone given the gray tones in a neighborhood containing it, in order to characterize the texture. The FRAR model coefficients Γ_r, ($r = 1, 2$) are computed for each sub-image region by using the FRAR model parameters K, α, θ and ϕ, which are estimated using the BA. The autocorrelation coefficient is used to assess the amount of regularity as well as the fineness/coarseness of the texture present in the image and is computed for each sub-image region from its model coefficients Γ_r, which are to be similar for a coarse texture while there will be a wide variation among the model coefficients for the fine texture.

The micro-textures that are below or above the threshold for humans to understand are identified in each sub-image regions by measuring the homogeneity of variances among the computed autocorrelation coefficients using a statistical test of hypotheses and it is expressed as in (Seetharaman and Palanivel, 2013).

$$\tilde{D}_m = n\left[1 - |\tilde{R}_m|^{1/m}\right] \qquad (3)$$

where autocorrelation matrix $\tilde{R}_m$ is formed using the computed autocorrelation coefficients, n is the number of samples

and m is the lag variable. The statistical test of hypotheses is based on the measure $\pm\alpha(\sigma/\sqrt{n})$, where α is the level of significance and σ is the standard deviation.

The computed autocorrelation coefficient (α^k) of each sub-image region is compared with the measure $\pm\alpha(\sigma/\sqrt{n})$ to find the outcome of the statistical test of hypothesis. If the autocorrelation coefficient is highly significant, micro-texture is present in the corresponding sub-image region. Otherwise (autocorrelation coefficients that fall inside the confidence limit, at the 80% level of significance (Seetharaman and Palanivel, 2013)), untexturedness is present in the corresponding sub-image region. The level of significance (i.e. 80%) is fixed based on trial and error basis. A simple transformation ($\alpha^k * 100$) + 100 is applied to represent each identified micro-texture, which results in between 0 and 200 and these numbers are called *textnums*. The frequency of occurrences of each *texnum* in the entire image is computed and is called *texspectrums*.

The extracted micro-textures in an image are represented by a histogram of *texnums* Vs *texspectrum*s and is formed as a feature vector (F_t) and it is written as

$$F_t = h_t(I_v)$$

where h_t represents the histogram of *texnums* Vs *textspectrums* of an image I_v. The dimension (D_t) of the computed micro-texture feature vector is 201.

3.2.3. Shape feature

In the proposed system, the images in the HSV color space are transformed from a Cylindrical co-ordinates system I(H,S,V) to Cartesian co-ordinates system I(H′,S′,V′) (Liu et al., 2011) as follows:

$$H' = S.\cos(H), \ S' = S.\sin(S) \ \text{and} \ V' = V.$$

Subsequently, the images in the I(H′,S′,V′) space are segregated into $I_{H'}$, $I_{S'}$ and $I_{V'}$ component images. The FRAR model in Eq. (1) is applied on the $I_{H'}$, $I_{S'}$ and $I_{V'}$ images to estimate its surfaces. Then, the difference between $I_{H'}$, $I_{S'}$ and $I_{V'}$ images and their estimated surfaces are computed and the resultant images $I_{H'}^r, I_{S'}^r$ and $I_{V'}^r$ are called residual images. The global confidence limit is measured for each sub-image regions (5×5) of residual images $I_{H'}^r, I_{S'}^r$ and $I_{V'}^r$ with a desired significance level. The global confidence limit is calculated with the use of global mean and standard deviation. If the pixel value is greater than or equal to the global confidence limit, then it is squared; otherwise the pixel value is replaced with zero. The value other than 0 represents the edges; and 0 represents non-edge part in the image. This extracts thick edges. To suppress the pixels around the edge pixel and to obtain thin edges, the non-maxima suppression algorithm is applied with the use of local confidence limit. The local confidence limit is calculated with the use of local mean and standard deviation of the thick edge map. Each value in the thick edge map is compared with the local confidence limit. If the value in the thick edge map is greater than the confidence limit, then it is identified as edge pixels. Otherwise, it is identified as non-edge pixel and is represented as 0. After extracting all the edge pixels of images $I_{H'}^r, I_{S'}^r$ and $I_{V'}^r$, the orientation of edge pixels is computed (Liu et al., 2011) as:

$$\theta = \arccos\left[\frac{a.b}{|a|.|b|}\right] \qquad (4)$$

where a and b represent the gradients along x and y direction respectively, and are expressed as follows:

$$a = H'_x + S'_x + V'_x; b = H'_y + S'_y + V'_y \text{ and } ab$$
$$= H'_x H'_y + S'_x S'_y + V'_x V'_y$$

where H'_x is the gradient in $I^r_{H'}$ along the horizontal direction, S'_x is the gradient in $I^r_{S'}$ along the horizontal direction and so on.

After computing the orientations of the edges, they are quantized into 72 bins of 5° each, which is used to form an EOAC. In the proposed system, the EOAC is constructed as a matrix, which consists of 72 rows and 2 columns. Each element of the matrix is computed by comparing each edge pixel with its neighborhood edge pixels (k pixel distance apart) to determine their similarity based on their magnitude and orientations (Mahmoudi et al., 2003). Since the smallest distance gives a more detailed local property of the image, we fix distance k with {1,3}. The proposed EOAC is formed as a feature vector of dimension D_S and it is described as:

$$F_s = \alpha^k_\theta(I)$$

The dimension (D_S) of the proposed EOAC is 144.

3.3. Formation of combined feature vector

The proposed CBMIR system combines the automatically extracted color, micro-textures and shape feature vectors into one feature vector to represent each image in the experimental database. The combined feature vector is given as follows:

$$F = (F_c, F_t, F_s)$$

$$F = [[(\alpha^k_c(I_H)), (\alpha^k_c(I_S))], h_t(I_V), \alpha^k_\theta(I)]$$

where F denotes the combined feature vector, F_c, F_t and F_s denote the color, micro-textures and shape feature vectors respectively. The dimension (D) of the combined feature vector is 361 [color autocorrelogram: 16 (H component image: 8; S component image:8), EOAC: 144, Micro-textures: 201]. The combined feature vector is normalized based on the expression given in the Eq. (5) to bring all the feature vectors to a common range.

$$F = \left[\left(\frac{F^H_c - \mu^H_c}{\sigma^H_c}, \frac{F^S_c - \mu^S_c}{\sigma^S_c} \right), \frac{F_t - \mu_t}{\sigma_t}, \frac{F_S - \mu_S}{\sigma_S} \right] \quad (5)$$

where μ^H_c, μ^S_c, μ_t, μ_S and σ^H_c, σ^S_c, σ_t, σ_S denotes the mean and standard deviation of color feature vectors extracted from I_H and I_S images, texture and shape feature vectors extracted from I_V image respectively. The normalized feature vectors are stored in a separate feature vector database.

3.4. Classification and indexing

The employed ABTSVM performs an image feature vector classification in the feature vector space. For an N-class problem, the ABT approach needs to train $N(N-1)/2$ binary SVMs of one-against-one (OAO), where each binary SVM is trained on data from two information classes. The top of the binary tree includes all the information classes. Each node in the binary tree is divided into two sub-nodes and each internal node in the binary tree, selects the binary SVM with the fewest average number of support vectors (SVs). This process is repeated until the system reaches the leaf node. i.e., when an unlabeled sample reaches the leaf node in the testing phase, an unlabeled sample is assigned to the class.

For the experimental databases, the ABTSVM classifies the feature vectors in the feature vector space into 83 classes such as X-ray-Foot-Coronal, X-ray-Elbow-Axial, X-ray-Chest-Bone, Microscopy-Liver-Hematoxylin and eosin (H and E) stain, Microscopy-Liver-Masson's Trichrome, Microscopy-Liver-Oil red O, etc. in the final level. Thus, the search space is reduced significantly by filtering the irrelevant classes of images and it improves the retrieval speed.

In the experiment, to attain the best generalization performance and to reduce the overfitting problem, ten-fold cross-validation is used. The N feature vectors in the feature vector database are randomly divided into ten subsets of approximately equal size. Each multi-class model was trained using $10 - 1 = 9$ subsets and tested using the remaining subset. Training is repeated ten times. Each binary SVM classifier requires the selection of regularization parameter C and the kernel parameter σ. The parameters C and σ of each SVM classifier have been tuned by using the grid search algorithm. The best combinations of parameters are used to build the final ABTSVM training model.

Afterward, the classified feature vectors are indexed using the kD-tree (White and Jain, 1996) indexing method, which increases the image retrieval speed.

3.5. Similarity and performance measure

Measuring similarity between the query and target images based on a derived feature vectors is a key component of any CBMIR system. In the literature, various similarity measures from computational geometry, statistics and information theory are used in the CBMIR systems. In this paper, the most frequently used geometrical similarity measure, Manhattan metric of order 1 is used (Huang et al., 1997), because the Manhattan similarity measure of order 1 is more robust to outliers and it saves much computational cost. The Manhattan distance (Huang et al., 1997) between the query and target images is computed as follows:

$$D(F^q, F^t) = \sum_{i=1}^n (F^q_i - F^t_i) \quad (6)$$

where, F^q_i is a query image feature vector, F^t_i is a target image feature vector in the feature vector database, and n is the size of the combined feature vector.

The performance of the proposed CBMIR system is measured in three aspects, namely, efficiency, effectiveness and computational complexity. Efficiency is associated with the storage requirements. The effectiveness of a system is related to the retrieval accuracy of the system and is measured using the most widely used precision (percentage of retrieved images that are also relevant) and recall (percentage of relevant images that are retrieved) methods (Huang et al., 1997; Nezamabadipour and Kabir, 2009) and is defined as follows:

$$\text{Precision} = \frac{R_i}{T_i} \quad (7)$$

$$\text{Re call} = \frac{R_i}{T} \quad (8)$$

where R_i is the number of relevant retrieved images, T is the total number of relevant images in the image database, and T_i is the number of all retrieved images. The effectiveness of the proposed system is also measured in terms of average recognition rate (ARR), which is defined as the percentage of retrieved images in top matches, which belongs to the same class as a query image.

3.6. Relevance feedback

In the proposed system, a user can query the image database and the system will return a list of target images in decreasing order of their computed similarity to the query image in some visual sense. In the proposed CBMIR, the ABTSVM classifier classifies the query image into a known label L using the automatically extracted visual features of query image and the knowledge learned during the training process. Then, Manhattan distance measure of order 1 is applied in the corresponding class of indexed images. This combined strategy (classification and similarity measure) attempts to reduce the semantic gap. However, in some cases the retrieved results of the proposed CBMIR system differ significantly from the expected results of the physicians. Hence, to maintain a balance between the retrieval results and physicians' expectations, query refinement approach of short-term learning based RF technique (Squire et al., 1998; Nezamabadi-pour and Kabir, 2009) is used in the proposed system to improve the retrieval performance as closely as corresponds to what the physician has in his mind in a particular search. Thus, the system allows the physicians to give feedback by selecting the positive examples from the list of retrieved images in each iteration of RF mechanism until the fulfillment of physicians. In the each iteration of RF, query image is updated by means of combining the current query image with the images that are marked as positive examples in the current retrieval results. Thus, the similarity measures can be redefined more effectively in the retrieval process to reduce the semantic gap as less as possible.

4. Experimental results and discussion

In order to implement the proposed system, we have collected medical images with the ground truth from the Rajah Muthiah Medical college Hospital, Annamalai University, Annamalai Nagar, India. The experimental medical image database contains 6400 images of 83 classes of X-ray, CT, Mammogram, MRI, Microscopy, Ultrasound, and Endoscopy modalities. A number of images in the X-ray, CT, Mammogram, MRI, Microscopy, Ultrasound, and Endoscopy modalities of our experimental database are 1510, 592, 858, 580, 1708, 663 and 489 respectively. The images are in tiff and jpeg format and they vary in size. Some sample images from the experimental database are shown in Fig. 2.

In order to evaluate the retrieval performance of the proposed system in a qualitative and quantitative manner, we have chosen different sub-sets of query images for the entire experimental database and a number of query images in each sub-set varies such that 100, 200, 300,..., 1500. But a number of query images in each sub-set of CT, Mammogram, MRI, Ultrasound and Endoscopy modalities varies such that 50, 100, 150,..., total number of images in the corresponding modality.

4.1. Experiment 1

We have implemented the proposed and conventional EOAC (Mahmoudi et al., 2003) for both gray-scale and color medical images in our experimental database. The retrieval performance of the proposed EOAC is compared to that of conventional EOAC for the gray-scale and color medical images. It is observed from the results that the proposed EOAC performs significantly better than that of the conventional EOAC for both gray-scale and color medical images and the attained average retrieval rate of the proposed EOAC and conventional EOAC for the gray-scale and color medical images are 64.59%, 66.57%, 60.08% and 59.32% respectively. The comparative results are shown in Fig. 3 and it depicts the average precision versus recall of the proposed EOAC and conventional EOAC for the gray-scale and color medical images.

4.2. Experiment 2

We evaluated the individual retrieval performance of each proposed feature for our experimental database and the obtained average retrieval rates are tabulated in Table 1.

4.3. Experiment 3

In our experiment, while the query image is posed to the proposed system, it automatically extracts color and its spatial information using the color autocorrelogram, texture information at global and local level using the micro-textures and then shape and its spatial information using EOAC. The extracted features are combined and normalized to perform the search operation, and the dimension of the proposed feature vector is 361.

We have chosen the medical image retrieval systems presented by Rahman et al. (2013) and Sudhakar and Bagan (2014) for comparative study and they are implemented in our experimental database. While the query image is posed to the existing system described by Rahman et al. (2013), it extracts CLD, EHD, color moments, GLCM moments, SIFT feature descriptor, edge frequency, primitive length, Gabour moments, Tamura moments, CEDD, FCTH, autocorrelation coefficient, LBP and LBP I as a visual feature of dimension 1361 and it extracts "Bag of words" of dimension 2703. After performing the dimensionality reduction, the dimension of feature vector is 3423. Whereas query image is posed to the existing system reported by Sudhakar and Bagan (2014), it extracts phase congruency based keypoints from the L*a*b* color space using the SIFT and the extracted keypoints are clustered using the Spherical Self-Organizing Map (SOM) built with a geodesic data structure. Subsequently, we evaluated the retrieval performance of the proposed and the existing systems for each modality of our experimental database and the results are compared. The attained average retrieval rates of the proposed system for the CT, MRI, Microscopy, Mammogram, Ultrasound, X-ray and Endoscopy images are 86.05%, 86.20%, 81.93%, 86.12%, 85.43%, 85.30%, 82.57%; 83.55%, 84.25%, 79.60%, 83.65%, 82.25%, 82.53%, 79.21% for the system described by Rahman et al. (2013), and 50.54%, 46.21%, 47.18%, 45.31%, 48.57%, 43.46%, 48.73% for the system proposed by Sudhakar and Bagan (2014)

(a)

Figure 2 (a) Sample gray-scale images of various classes taken from the experimental database. (b) Sample color images of various classes taken from the experimental database.

respectively. The obtained results confirm that the proposed system outperforms the existing systems for CT, MRI, Microscopy, Mammogram, Ultrasound, X-ray and Endoscopy modalities. The average precision versus recall of the proposed system and existing systems for each modality is shown in Fig. 4.

4.4. Experiment 4

We evaluated the overall performance of the proposed and existing systems (Rahman et al., 2013; Sudhakar and Bagan, 2014) for our entire experimental database. The feature vector dimension, attained average retrieval rate and
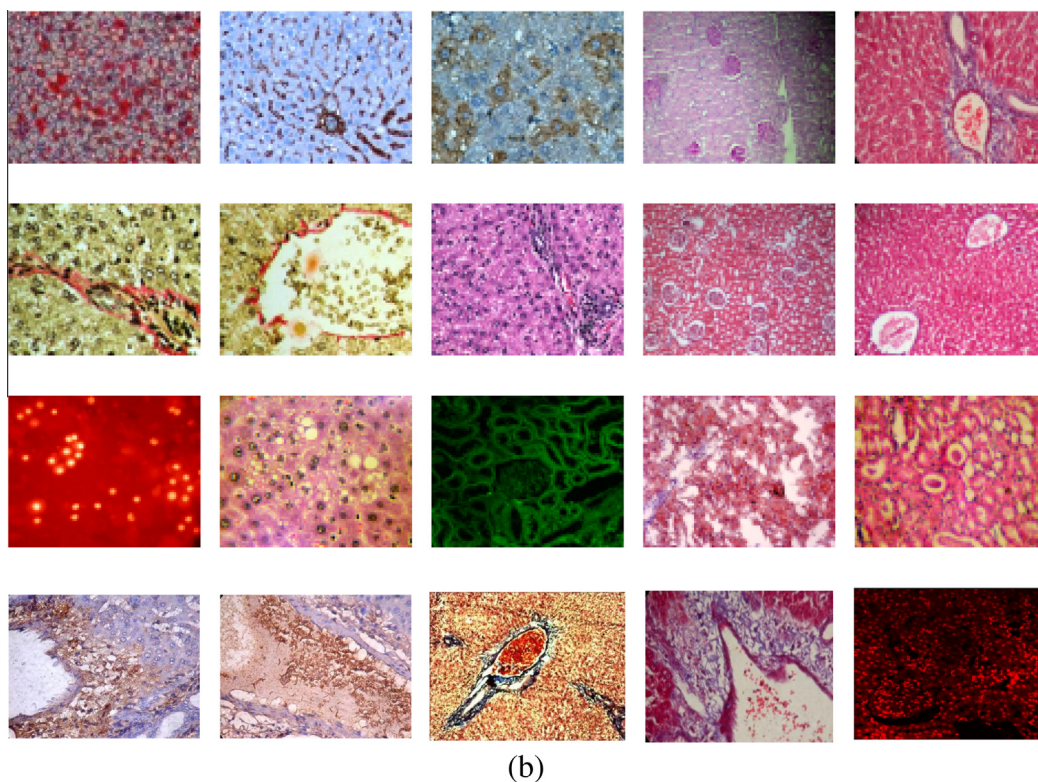
(b)

Fig. 2 (*continued*)
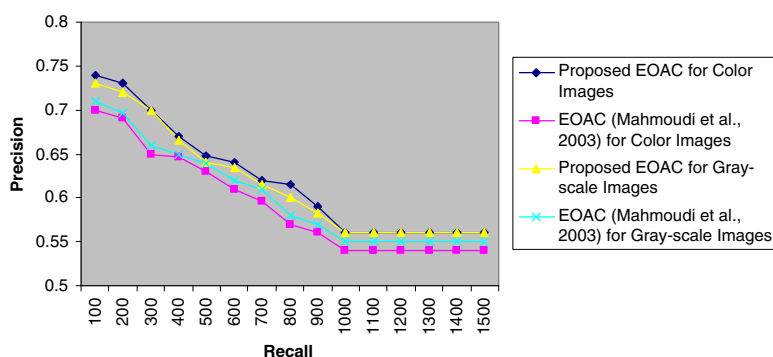
Figure 3   Average precision versus recall of the proposed and conventional EOAC for color and gray-scale medical images.

retrieval time of the proposed and existing systems for our experimental database are tabulated in the Table 2. The average precision versus recall for the proposed and existing systems is depicted in Fig. 5. The results demonstrate that the performance of the proposed system is significantly better than the existing systems and both the proposed and existing systems are not influenced by the size of the medical image database. The points indicated by the arrow marks in Fig. 5 represent the threshold value of precision for the proposed and existing systems for our experimental database. It is observed from the results that the precision is ideal if the query images are more than 600 for the proposed system; 800 query images for the existing system described by Rahman et al. (2013) and 900 query images for the existing

Table 1   Role, dimension and average retrieval rate of the proposed features.

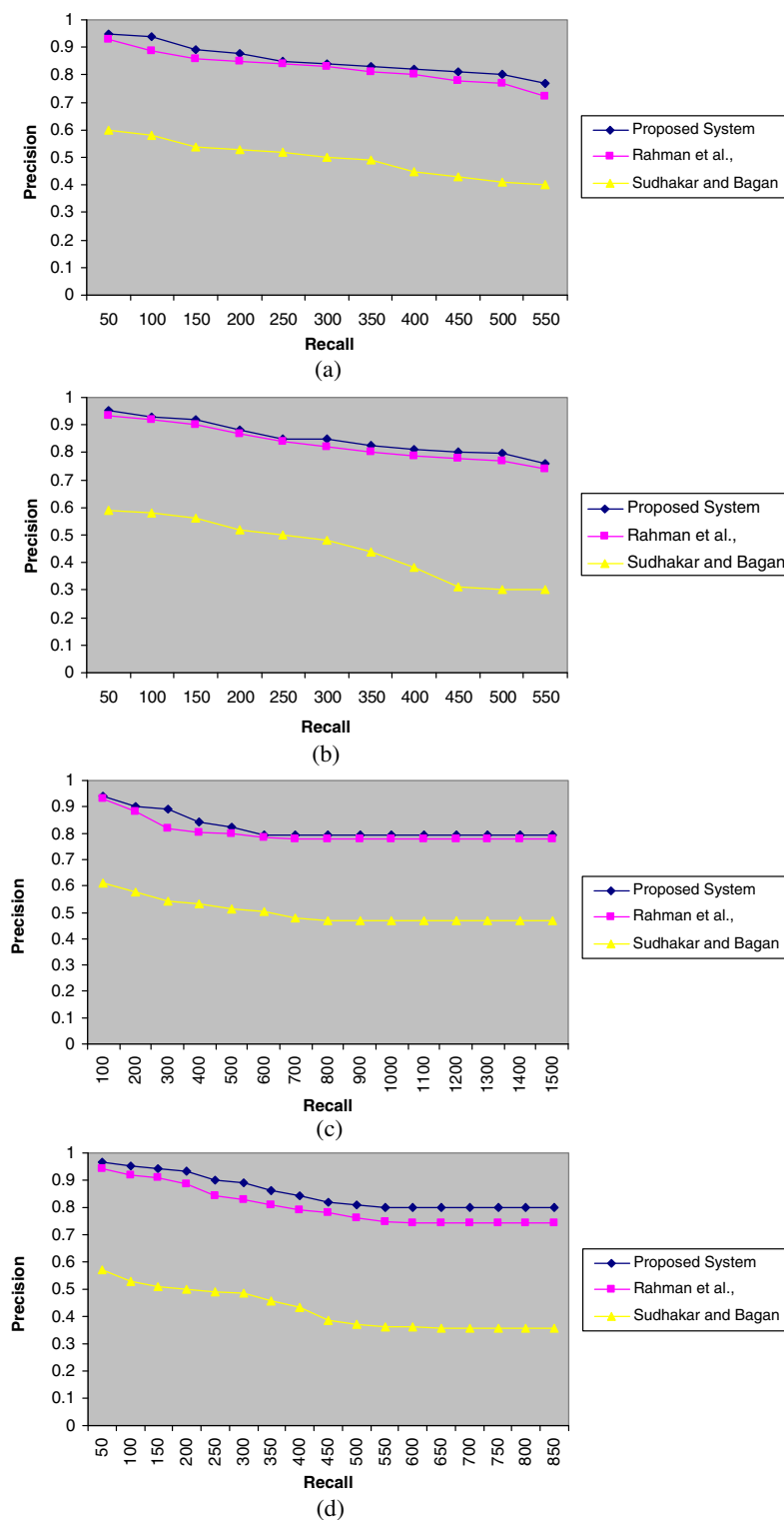| Feature Representation	Role	Dimension	Average retrieval rate (%)
Color autocorrelogram	Color and its spatial information	16	61.62
EOAC	Shape and its spatial information	144	64.59 (gray-scale image)
			66.57 (color image)
Micro-textures	Texture information at global and local level	201	83.49

Figure 4 The average precision versus recall of the proposed and existing systems for various modalities (a) CT; (b) MRI; (c) Microscopy; (d) Mammogram; (e) Ultrasound; (f) X-ray (g) Endoscopy.

system proposed by Sudhakar and Bagan (2014). Figs. 6 and 7 show a test query image and its top ten retrieval results (ascending order of obtained distances between the query image and the target images) of the proposed system for the gray-scale and color medical images respectively.

4.5. Experiment 5

In order to ensure the accuracy of the existing image retrieval systems (Rahman et al., 2013; Sudhakar and Bagan, 2014), we have conducted experiments on the heterogeneous medical
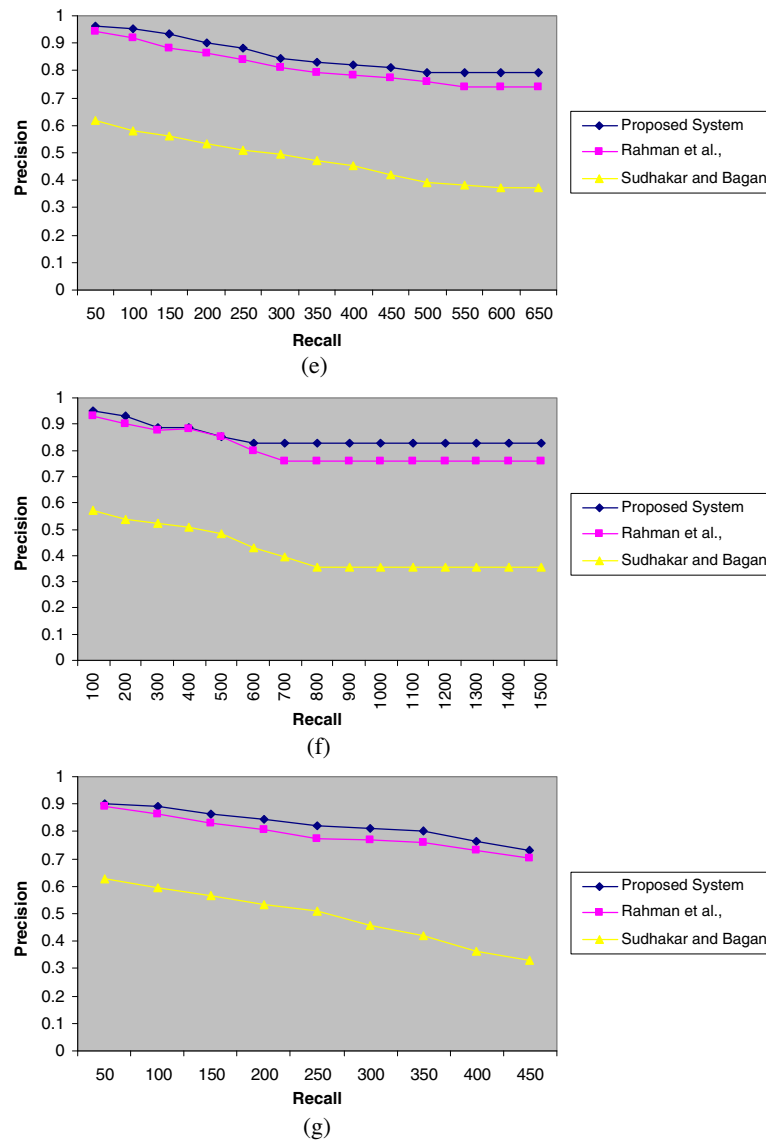
(e)

(f)

(g)

Fig. 4 (*continued*)

image databases used in the existing systems. The heterogeneous medical image database used by Rahman et al. (2013) and Sudhakar and Bagan (2014) consists of more than 300,000 and 6200 images respectively. The experimental results demonstrate that the proposed system performs better than the existing systems. The attained average retrieval rate and retrieval time in seconds for the proposed system and existing systems in the databases used by Rahman et al. (2013) and Sudhakar and Bagan (2014) are 85.01%, 7.42; 83.06%,

26.13; 48.42%, 15.27 and 83.63%, 2.61; 80.85%, 9.02; 49.72%, 5.94, respectively. The comparative results are shown in Fig. 8 and it depicts the average precision versus recall of the proposed and existing systems.

The experimental results show that high dimensional feature vector does not improve the performance of CBMIR and the low dimensional feature vector with distinct and high discriminative power is sufficient to distinguish the wide range of visual contents in the medical images. The feature vector

Table 2 Feature vector dimension, average retrieval rate and retrieval time of the proposed and existing retrieval systems for our experimental databases.

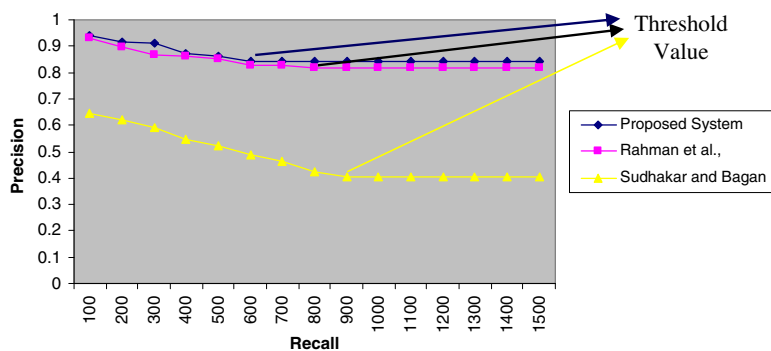| Method	Dimension	Average retrieval rate (%)	Retrieval time in seconds
Proposed system	361	84.87	2.29
Rahman et al.	3423 (after dimensionality reduction)	82.14	8.59
Sudhakar and Bagan	21 keypoints (after clustering) * 128 feature vector = 2688	47.43	5.41

Figure 5   Average precision versus recall of the proposed and existing systems for our experimental database.
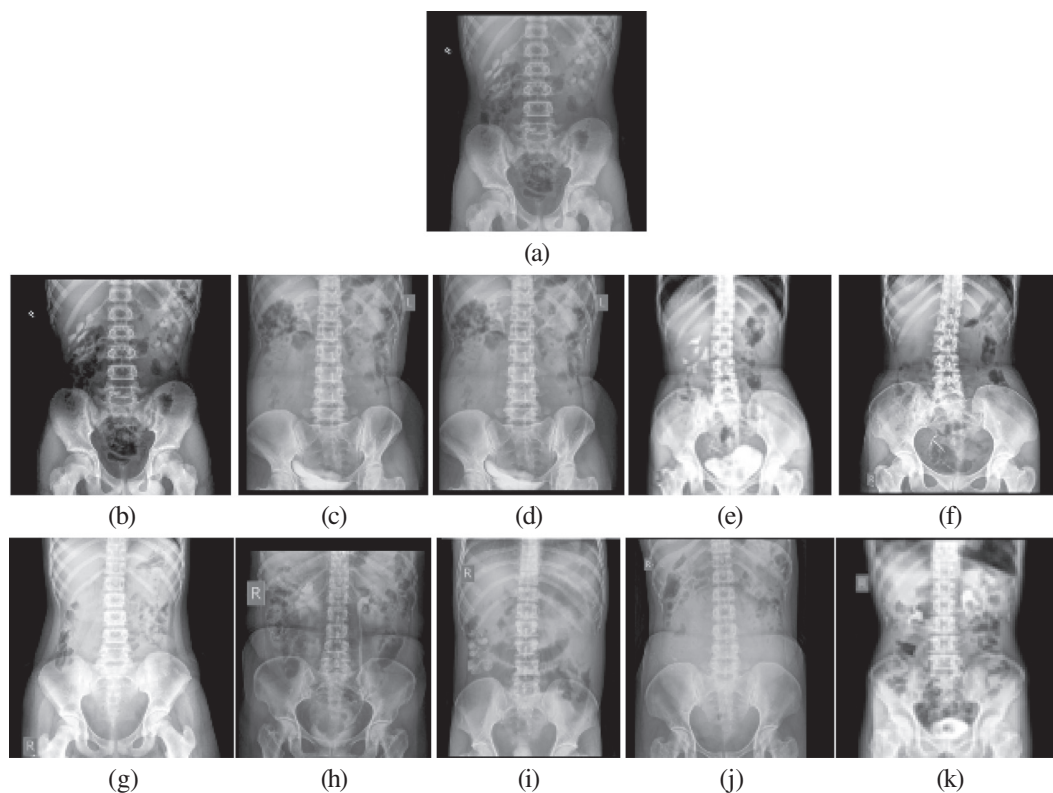

Figure 6   (a) Example query image and (b) retrieval results obtained with the proposed system corresponding to the query image of our experimental database.

dimension is one of the most important factors that determine the amount of storage space and computational complexity. Moreover, it is often necessary and beneficial to limit the dimension of input feature vector fed to a classifier in order to have a good predictive classifier. Thus, the proposed system uses a feature vector of appropriate dimension, which results in good retrieval accuracy while not requiring a great amount of storage space and computational complexity. Thus, the proposed feature vector tradeoff between the retrieval efficiency, computational and storage complexity. Furthermore, the system presented by Rahman et al. (2013) extracts various visual and textual features of an image using various techniques, which is a cumbersome process due to the complementariness of techniques.

4.6. Experiment 6

The experiments are also designed to investigate the robustness of the proposed and existing systems (Rahman et al., 2013; Sudhakar and Bagan, 2014) against scaling, noise, illumination and viewing positions. We observed that the proposed system performs well against scaling, illumination and viewing positions, and also works well against noise to some extent due to the basic properties of the FRAR model.

The computational efficiency of the proposed and existing systems is measured with respect to CPU utilization time for extracting the feature vector and search time of image database for a given query image.
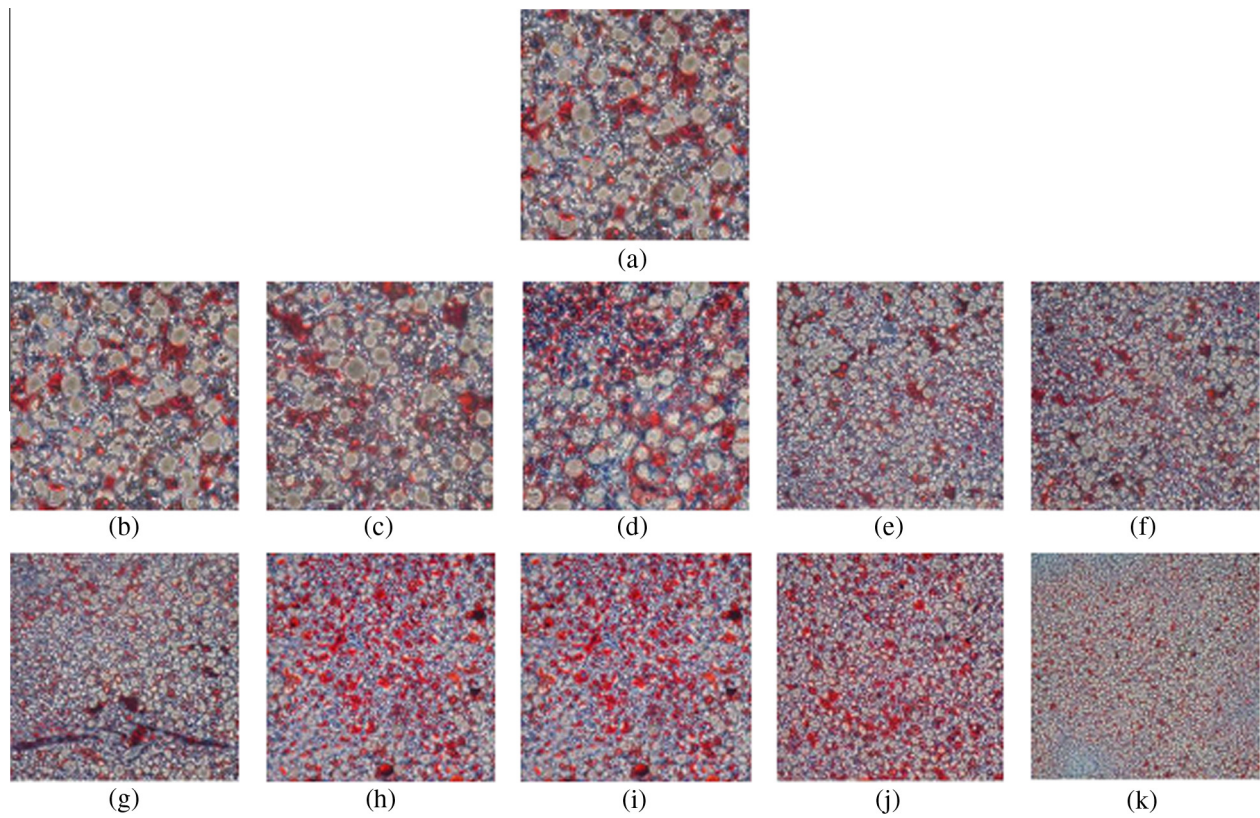
Figure 7 (a) Example query image and (b) retrieval results obtained with the proposed system corresponding to the query image of our experimental database [Image (b), (c) and (d): Magnification 20×, Image (e)–(k): Magnification 10×].
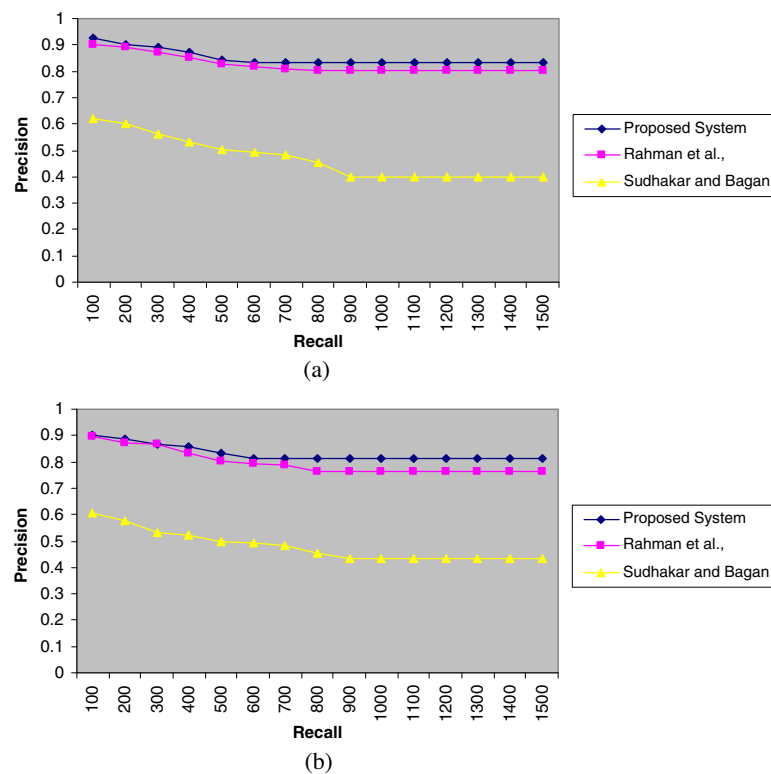

Figure 8 Average precision versus recall of the proposed and existing systems for the databases used in (a) Rahman et al. (2013); (b) Sudhakar and Bagan (2014).

In the proposed system, the relevance judgments are performed using the ground truth and the subjectivity of the individual user. The result of the proposed system indicates its usefulness to clinical applications, education and researches. The proposed work is implemented with the system configuration: Pentium® Dual core personal computer with 2.20 GHz processor; 2 GB RAM; Java and Oracle.

5. Conclusion

In this paper, a framework based on the FRAR model with Bayesian approach is proposed for retrieving heterogeneous medical images. The proposed unified framework is used to extract all (color and its spatial information, shape and its spatial information and micro-textures) the proposed features of gray-scale and color medical images. The proposed system is evaluated on various databases consisting of images of various modalities. The average retrieval rate obtained by the proposed system is significantly better than that of the existing systems at low computational and storage cost. We have described a new version of EOAC, which outperforms the conventional EOAC for both gray-scale and color medical images. The result of the proposed system indicates its usefulness to various clinical aspects, education and researches.

Acknowledgement

References

Avni, Uri, Greenspan, Hayit, Konen, Eli, Sharon, Michal, Goldberger, Jacob, 2010. X-ray categorization and retrieval on the organ and pathology level using patch-based visual words. IEEE Trans. Med. Imag. 30 (3), 733–746.

Cai, W., Feng, D.D., Fulton, R., 2000. Content-based retrieval of dynamic PET functional images. IEEE Trans. Inform. Technol. Biomed. 4 (2), 152–158.

Chen, Wenjin, Meer, Peter, Georgescu, Bogdan, He, Wei, Goodell, Lauri A., Foran, David J., 2005. Image mining for investigative pathology using optimized feature extraction and data fusion. Comput. Methods Programs Biomed. 79, 59–72.

Chen, Jin, Wang, Cheng, Wang, Runsheng, 2009. Adaptive binary tree for fast SVM multi class classification. Neurocomputing 72, 3370–3375.

Chu, W.W., Hsu, C.C., Cardenas, A.F., Taira, R.K., 1998. Knowledge-based image retrieval with spatial and temporal constructs. IEEE Trans. Knowledge Data Eng. 10, 872–888.

Chun, Y.D., Kim, N.C., Jang, I.H., 2008. Content-based image retrieval using multiresolution color and texture features. IEEE Trans. Multimedia 10 (6), 1073–1084.

El-Kwae, Y.E.A., Xu, H., Kabuka, M.R., 2000. Content-based retrieval in picture archiving and communication systems. IEEE Trans. Knowledge Data Eng. 13 (2), 70–81.

Güld, M.O., Christian, Thies, Benedikt, Fischer, Lehmann, T.M., 2007. A generic concept for the implementation of medical image retrieval systems. Int. J. Med. Informatics 76, 252–259.

Hsu, William, Antani, Sameer, Long, L.Rodney, Neve, Leif, Thoma, George R., 2009. SPIRS: A Web-based image retrieval system for large biomedical databases. Int. J. Med. Inform. 78S, S13–S24.

Huang, J., Kumar, S.R., Mitra, M., Zhu, W., Zabih, R., 1997. Image indexing using color correlograms. Int. Conf. Comput. Vision Pattern Recogn., 762–768

Karkanis, S.A., Iakovidis, D.K., Maroulis, D.E., Karras, D.A., Tzivras, M., 2003. Computer-aided tumor detection in endoscopic video using color wavelet features. IEEE Trans. Inf. Technol. Biomed. 7 (3), 141–152.

Lehmann, T.M., Güld, M.O., Deselaers, T., Keysers, D., Schubert, H., Spitzer, K., Ney, H., Wein, B.B., 2005. Automatic categorization of medical images for content-based retrieval and data mining. Comput. Med. Imaging Graph. 29, 143–155.

Liu, G.H., Li, Z.Y., Zhang, L., Xu, Y., 2011. Image retrieval based on micro-structure descriptor. Pattern Recogn. 44 (9), 2123–2133.

Long, L.R., Thoma, G.R., 2001. Landmarking and feature localization in spine X-rays. J. Electron. Imag. 10 (4), 939–956.

Mahmoudi, F., Shanbehzadeh, J., Eftekhari, A.M., Soltanian-Zadeh, H., 2003. Image retrieval based on shape similarity by edge orientation autocorrelogram. Pattern Recogn. 36, 1725–1736.

Müller, Henning, Michoux, Nicolas, Bandon, David, Geissbuhler, Antoine, 2004. A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. Int. J. Med. Informatics 73, 1–23.

Müller, Henning, Rosset, Antoine, Garcia, Arnaud, Vallée, Jean.-Paul, Geissbuhler, Antoine, 2005. Benefits of content-based visual data access in radiology. Radio Graph. 25 (3), 849–858.

Nezamabadi-pour, Hossein, Kabir, Ehsanollah, 2009. Concept learning by fuzzy k-NN classification and relevance feedback for efficient image retrieval. Expert Syst. Appl. 36, 5948–5954.

Orphanoudakis, S.C., Chornaki, C., Kostomanolakis, S., 1994. I2C – a system for the indexing, storage and retrieval of medical images by content. Med. Inf. 19 (2), 109–122.

Penatti, O.A.B., Valle, E., Torres, R.D.S., 2012. Comparative study of global color and texture descriptors for web image retrieval. J. Vis. Commun. Image R. 23, 359–380.

Rahman, Md. Mahmudur, Bhattacharya, Prabir, Desai, Bipin C., 2007. A framework for medical image retrieval using machine learning and statistical similarity matching techniques with relevance feedback. IEEE Trans. Inf. Technol. Biomed. 11 (1), 58–69.

Rahman, Mahmudur Md., Desai, Bipin C., Bhattacharya, Prabir, 2008. Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion. Comput. Med. Imaging Graph. 32, 95–108.

Rahman, Mahmudur Md., Bhattacharya, Parbir, Desai, Bipin C., 2009. A unified image retrieval framework on local visual and semantic concept based feature space. J. Vis. Commun. Image R. 20 (7), 450–462.

Rahman, Md. Mahmudur, Antani, Sameer K., Thoma, George R., 2011. A learning-based similarity fusion and filtering approach for biomedical image retrieval using svm classification and relevance feedback. IEEE Trans. Inf. Technol. Biomed. 15 (4), 640–646.

Rahman, Mahmudur Md., You, Daekeun, Simpson, Matthew S., Demner-Fushman, Dina, Antani, Sameer K., Thoma, George R., 2013. Multimodal biomedical image retrieval using hierarchical classification and modality fusion. Int. J. Multimed. Inf. Retr. 2, 159–173.

Rui, Y., Huang, T.S., Chang, S.-F., 1999. Image retrieval: current techniques, promising directions, and open issues. J. Vis. Commun. Image R. 10, 39–62.

Seetharaman, K., Krishnamoorthi, R., 2007. A statistical framework based on a family of full range autoregressive models for edge extraction. Pattern Recogn. Lett. 28 (7), 759–770.

Seetharaman, K., Palanivel, N., 2013. Texture characterization, representation, description, and classification based on full range Gaussian markov random field model with Bayesian approach. Int. J. Image Data Fusion, 1–24.

Shyu, C.R., Brodley, C.E., Kak, A.C., Kosaka, A., Aisen, A.M., Broderick, L.S., 1999. ASSERT: A physician-in-the-loop content-based image retrieval system for HRCT image databases. Comput. Vis. Image Understand 75, 111–132.

Squire, D., Muller, W., Muller, H., 1998. Relevance feedback and term weighting techniques for content-based image retrieval. Technical Report, Computer Vision Group, University of Geneva, 98.05.

Stricker, M., Orengo, M., 1995. Similarity of color images. in: Proc. SPIE Storage and Retrieval for Image and Video Databases.

Sudhakar, M.S., Bagan, Bhoopathy K., 2014. An effective biomedical image retrieval framework in a fuzzy feature space employing Phase Congruency and GeoSOM. Applied Soft Computing, Article in press.

Tang, H. Lilian, Hanka, Rudolf, Ip, H. Horace, 2003. Histological image retrieval based on semantic content analysis. IEEE Trans. Inf. Technol. Biomed. 7 (1), 26–36.

Wei, C.-H., Li, Y., Huang, P.J., 2011. Mammogram retrieval through machine learning within BI-RADS standards. J. Biomed. Inform. 44 (4), 607–614.

White, D.A., Jain, R, 1996. Algorithms and strategies for similarity retrieval. Tech. Rep. VCL-96-101, Visual Computing Laboratory, University of California, San Diego.

Yu, Sung-Nien, Chiang, Chih-Tsung, 2004. Similarity Searching for Chest CT Images Based on Object Features and Spatial Relation Maps. Proceedings of the 26th Annual International Conference of the IEEE EMBS San Francisco, CA, USA, pp. 1–5.