Interactive Radiographic Image Retrieval System

Malay Kumar Kundu*, Manish Chowdhury*[1] and Sudeb Das*[2]

Machine Intelligence Unit, Indian Statistical Institute, Kolkata-700108, India.
E-mail: malay@isical.ac.in, st.manishc@gmail.com*[1], and to.sudeb@gmial.com*[2]*

Abstract

Background and Objective: Content based medical image retrieval (CBMIR) systems enable fast diagnosis through quantitative assessment of the visual information and is an active research topic over the past few decades. Most of the state-of-the-art CBMIR systems suffer from various problems: computationally expensive due to the usage of high dimensional feature vectors and complex classifier/clustering schemes. Inability to properly handle the '*semantic gap*' and the high intra-class versus inter-class variability problem of the medical image database (like radiographic image database). This yields an exigent demand for developing highly effective and computationally efficient retrieval system.

Methods: We propose a novel interactive two-stage CBMIR system for diverse collection of medical radiographic images. Initially, Pulse Coupled Neural Network based shape features are used to find out the most probable (similar) image classes using a novel '*similarity positional score*' mechanism. This is followed by retrieval using Non-subsampled Contourlet Transform based texture features considering only the images of the pre-identified

[1]Corresponding author. Tel. +91-33-2575 3110/3100, Fax +91-33-2578-3357.

classes. Maximal information compression index is used for unsupervised feature selection to achieve better results. To reduce the semantic gap problem, the proposed system uses a novel fuzzy index based relevance feedback mechanism by incorporating subjectivity of human perception in an analytic manner.

Results: Extensive experiments were carried out to evaluate the effectiveness of the proposed CBMIR system on a subset of Image Retrieval in Medical Applications (IRMA)-2009 database consisting of $10,902$ labeled radiographic images of 57 different modalities. We obtained overall average precision of around 98% after only 2-3 iterations of relevance feedback mechanism. We assessed the results by comparisons with some of the state-of-the-art CBMIR systems for radiographic images.

Conclusions: Unlike most of the existing CBMIR systems, in the proposed two-stage hierarchical framework, main importance is given on constructing efficient and compact feature vector representation, search-space reduction and handling the '*semantic gap*' problem effectively, without compromising the retrieval performance. Experimental results and comparisons show that the proposed system performs efficiently in the radiographic medical image retrieval field.

Keywords: Content Based Medical Image retrieval, Pulse Couple Neural Network, Multiscale Geometric Analysis, Relevance Feedback, Fuzzy logic, Radiographic images.

---

1. Introduction

Over the last few decades, extensive attempts have been made by various researchers to develop Content Based Medical Image Retrieval (CBMIR) systems for efficient management and access of medical image data. In large hospitals and diagnostic centers, several terabytes of digital medical images are generated and stored every year [1, 2, 3]. In today's health-care framework, these images are archived in Picture Archiving and Communication Systems (PACS) [4, 5, 6]. For precise diagnosis and treatment planning, often the medical personals have to browse through similar images in these archives. This necessitates the requirement of developing novel and intelligent techniques for searching images accurately and efficiently from a large medical image database [7, 8, 9]. Such systems have immense applications in todays modern health-care domain: not only these systems can be used to reduce the workload of the medical personals, but also can be used as a cost-effective alternative in places where there is a lack of medical experts [2, 7, 10, 11, 12].

The commonly used textual metadata based image retrieval systems have several shortcomings: costly, time-consuming, dependency on the experience of the human annotator, error-prone (error rate up to 16% have been reported) and inefficiency to represents the true content of the medical image semantically [6, 13]. Another open challenge for automatic categorization of medical images is the inter-class versus intra-class variability problem [6, 7, 11]. For instance, considering the X-ray images of Image Retrieval in Medical Applications (IRMA-2009) database [4, 14], Fig. 1(a1)-(a2) illustrates the high degree of visual content variability within a particular class. Here, all the radiographs of Fig. 1(a1)-(a2) are plain X-ray images in coronal

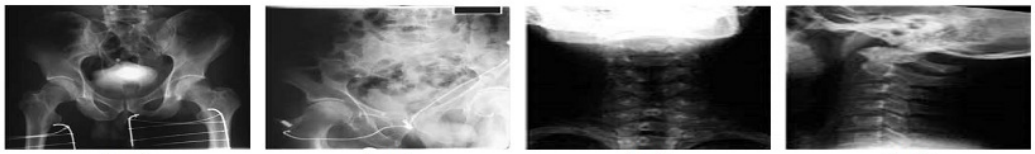


Figure 1: Intra-class variability: (a1)-(a2) vs. inter-class variability: (b1)-(b2) (considering IRMA categorization [4]).

posterioanterior direction of abdomen body region and of the musculoskeletal biosystem-having the same class level. On the other hand, there may be strong visual similarities across different image classes, especially with classes that have the same organ projected at different angles. For example, the image of the Fig. 1(b1) belongs to the "*neck*" class while the image of Fig. 1(b2) belongs to the "*neck side view*" class. This is mainly due to different doses of X-ray, varying orientation, alignment and pathology [6]. Moreover, often the images have different contrast variations and non-uniform intensity background, weak signal-to-noise ratio (SNR), projection and other noises [6, 15, 16, 17].

X-ray radiography is a powerful imaging tool (both in medical and industrial fields), which allows visualization of an object's internal and external structures [12, 14, 18, 19, 20]. Several image retrieval prototypes have been proposed by various researchers for medical radiographs [6, 12, 14, 17, 21, 22, 23, 24]. For example, Keysers et al., [25] have introduced a Bayesian probabilistic framework for radiograph classification based on an appearance-based approach combining tangent distant measure and image distortion model. Presenting an extensive evaluation of different methods for automatic classification, Lehmann et al. in [26] have proposed a radiographic image cate-

gorization scheme using a parallel combination of single classifiers (k-nearest neighbor) based on scaled representations and global texture features. Pinhas et al. [6] have achieved 97.5% accuracy in classifying $1,500$ radiography images of 17 categories using a Gaussian Mixture Model-Kullback Leibler (GMM-KL) framework based on a high-dimensional feature vector. In [27], Rahman et al. have presented a novel machine learning based image prefiltering approach based on a combination of a statistical similarity matching and relevance feedback (RF) scheme. On a database of $9,100$ medical X-ray images of 40 classes Pourghassem et al. [21] have achieved approximately 94% accuracy (merged 17-classes) using a merging-based hierarchical classifier. Avni et al. [28] have proposed a "*bag of visual words*" CBMIR approach based on a non-linear kerna-based support vector machine (SVM). Rahman et al., [24] have designed an iterative CBMIR system using bag of features using a query-specific adaptive linear combination of similarity matching approach by relying on the image classification and feedback information from users. Achieving 93% retrieval accuracy, a hierarchical classification structure based on a novel merging and splitting scheme using shape and texture features is proposed by Fesharaki et al. in [17]. A novel CBMIR system based on a score fusion algorithm for query image classification and a new query expansion method in the relevance feedback level is proposed by Behnam et al. [29]. A novel dictionary learning based clustering method for CBMIR is proposed by Srinivas et al. [30] using mean and variance of pixel intensity values as features based on the K-SVD and orthogonal matching pursuit mechanisms. Recently, few deep neural network based CBMIR systems have been proposed showing promising results [31, 32].

Most of the above mentioned CBMIR systems are non-hierarchical and based on single classification/clustering technique using different combinations of feature representation schemes. But, it is difficult for a non-hierarchical classification/clustering method to effectively handle the overlapping (inter-vs-intra class variability) of different classes of medical images [21, 17]. Not only that due to the use of relatively expensive classifier/clustering scheme, overall cost of computation remains fairly high in these systems. Moreover, in case of hierarchical CBMIR systems, if pre-classification of the query image is wrong, then the retrieval results will be wrong, which results in lowering the accuracy of the system. Furthermore, the dimensions of the image representative feature vectors used in these systems are quite high. This also leads to further computational burden. Even using such computationally expensive classifier/clustering schemes along with high dimensional feature representation technique, it is often not possible for the existing CBMIR schemes to effectively capture the true semantic contents of the medical images and efficiently maps these to the user's response. Therefore, it becomes necessary to develop new interactive CBMIR system with low computational cost and hight retrieval performance.

In this paper, we have proposed a novel CBMIR system to overcome these aforementioned problems of the existing systems. The proposed system is a two-stage image retrieval system and two different kinds of image features are used for image representation in two different stages. The main contributions of this paper are as follows: (1) we model the image retrieval problem as a hierarchical two-stage classification problem. In the first stage mammalian vision system mechanism imitating Pulse Coupled Neural Network

(PCNN) [33] based global shape features are used to coarsely identify the top 'p' probable classes similar to the query image. For finer retrieval, Multiscale Geometric Analysis of Non-subsampled Contourlet Transform (NSCT) [34] based local textural features are used in the second stage. The compaction of feature set is done class wise in IRMA image database using Maximal Information Compression Index (MICI) [35]. This reduces the overall search space which results in a faster and more accurate retrieval. (2) To find out the top 'p' probable classes in the first stage of the proposed system we have developed a novel visual proximity measure named "*similarity positional score*" (SPS) which is inversely proportional to distance metric. Based on this score, top probable classes similar to the class of the query image are identified which restricts the final search (second stage) only within the apparent classes comprising a very small portion of the whole IRMA-2009 image database. This reduces the search space and the mis-classification error of any classifier which classifies the data into a particular class instead of probable classes. (3) In addition, we have developed a novel *fuzzy index* based relevance feedback mechanism (*fuzzy-index*-RFM) to handle the uncertainty and ambiguity arising out of user feedback which is purely individualistic in nature. This reduces the "*semantic gap*" problem. Extensive experiments were carried out to evaluate the performance which show that the proposed system is efficient for radiographic image retrieval.

The rest of the paper is organized as follows: In Section 2, we describe the image representative feature extraction procedure based on PCNN and NSCT. The Section 3 describes the feature selection algorithm. The novel SPS mechanism is described in Section 4. The *fuzzy-index*-RFM and the

proposed CBMIR system is presented in Section 5 and Section 6, respectively. The Section 7, contains the experimental findings and we draw conclusion in Section 8.

2. Feature Extraction for Image Representation

The performance of a CBMIR system is largely dependent on how effectively the database's images are represented in terms of feature vector. Some images can be easily distinguishable by the global feature characteristics such as shape, geometric property, orientation etc., whereas there may be some other images which can be distinguishable by local feature characteristics such as color, texture, positional distribution etc. A judicious combination of both kind of features can help in achieving high accuracy at lower computation cost [36]. Feature dimension is also an important issue for achieving greater computational efficiency. Lower the dimension of the feature vector used, the higher will be the computational efficiency [2, 7]. Furthermore, total computational cost depends upon the number of comparisons to be made for matching which is proportional to the total number of images in the database. Keeping these above mentioned problems in mind, a hierarchical two stage retrieval scheme is used in the proposed framework.

In this work, two different novel feature representation schemes namely PCNN-shape and NSCT-texture are used in two hierarchical stages, respectively. In both the stages, the input images are gray level X-ray images taken from IRMA-2009 image database. The first stage is a coarse classification where PCNN based segmentation scheme is used for extracting the shape features. Due to its adaptive thresholding nature, PCNN serves as

an efficient tool for edge sharpening as well as for ambient noise removal in the mega voltage X-ray images [15]. PCNN mimicking the global coupling and pulse synchronization characteristics of mammalian vision system has been successfully applied in various image processing and analysis applications like segmentation, enhancement, fusion etc. This shows that PCNN has very good capability in capturing various features of image like texture, edge, shape, salient regions etc. But before doing segmentation, some pre-processing is required due to the intensity variations, low contrast and high rate of noise in X-ray images [17]. Hence, at first, median filter is used for noise reduction [16]. After noise reduction, adaptive histogram equalization is used for contrast enhancement, considering exponential distribution of histogram. Even though, adaptive histogram equalization provides improvement of the image contrast with no destructive effects on the areas with higher contrast, but it increases the noise on the image. As a result, after adaptive histogram equalization, a median filter is used again on the X-ray images to reduce the noise [17].

After pre-processing, segmentation is carried out using PCNN [37] and 7 global shape features are computed in each iteration. These features are perimeter, area, euler-number, eccentricity, major-axis-length, minor-axis-length and compactness. Therefore, considering 10 iterations (chosen experimentally to achieve satisfactory performance with low computational complexity), each image I is represented with 70 PCNN-shape features:

$$FM^{I}_{PCNN\_shape} = [f^{1}_{\text{PCNN\_shape}}, ..., f^{70}_{\text{PCNN\_shape}}] \tag{1}$$

In the second stage, NSCT-texture features are used for finer retrieval.

In the literature, it is reported that several different transforms like Wavelet Transform (WT), Curvelet Transform (CVT), Contourlet Transform (CNT) etc., based features are used for retrieval of medical images [2, 7, 38]. But these techniques lack shift-invariance and anisotropy properties. To overcome these problems, we adopt NSCT, which is a fully shift-invariant, multiscale, and multidirection expansion with fast implementability and anisotropy property. It also has better directional decomposition and frequency localization capability [34]. NSCT is constructed by combining the Non-Subsampled Pyramid Filter Bank and Non-Subsampled Directional Filter Bank, and the resulting filtering structure approximates the ideal partition of the frequency plane. NSCT coefficients are capable of capturing the fine details present in the image, which is very much essential in image retrieval, as radiographic images are full of subtle image information. Moreover, it has been reported that Human Visual System (HVS) has characteristics like multiscale, rotational invariance and anisotropy [39]. This behavior is well matched by NSCT decomposition. Based on these above facts, in the proposed CBMIR system, NSCT is used as the feature representative scheme for finer retrieval. The details about PCNN and NSCT are available in [33] and [34], respectively.

Each image I of size $N \times N$ is decomposed into l $(= 3)$ number of multiscales and multidirection decomposition of "*levels*"$= [0, 1, 2]$. This results in $(2^0 + 2^1 + 2^2) = 7$ number of high-frequency subbands (HFSs). The image, prior to NSCT decomposition, is pre-processed using the before mentioned enhancement mechanism. Each subband having the same size $N \times N$ that of the original image is sub-divided into W number of blocks of size $w \times w$, where $W = \frac{N^2}{w^2}$. Standard deviation of each block is used to represent the co-

efficients of that block. The low-frequency subband (LFS) that contains the DC information is ignored due to its relatively small contribution in terms of standard deviation compare to that of AC subbands. Thus, the dimension of the NSCT-texture feature vectors becomes $7 \times W = 448$, where $w = 16$, $N = 128$:

$$FM^{I}_{NSCT\_texture} = [f^{1}_{\text{NSCT\_texture}}, ..., f^{448}_{\text{NSCT\_texture}}] \tag{2}$$

3. Feature Selection

The success of any image recognition task depends on the adopted feature representation scheme. A good practice is to reduce the feature space to ideally preserve only the most informative features. In this work, we retain the relevant PCNN-shape and NSCT-texture based features by adopting a feature selection algorithm based on the Maximal Information Compression Index (MICI) [35]. This unsupervised method does not rely on searching techniques, thus avoiding over-consumption of calculation time, but it removes redundancies among the features by evaluating their mutual dissimilarities. In details, given two random variables x and y, let Σ be their covariance matrix with entries Σ_{xx}, Σ_{yy}, respectively and $\Sigma_{xy} = \Sigma_{yx}$. The value of MICI for x and y is defined as the smallest eigenvalue $\lambda_2(x, y)$ of the covariance matrix Σ, which is given by

$$\lambda_2(x, y) = \frac{1}{2}\left[\Sigma_{xx} + \Sigma_{yy} - \sqrt{(\Sigma_{xx} - \Sigma_{yy})^2 + 4\Sigma^2_{xy}}\right] \tag{3}$$

The value of $\lambda_2(x, y)$ is zero when the features represented by x and y are linearly dependent (hence, similar) and it increases as the degree of de-

pendency decreases. The algorithm clusters the features into homogeneous groups by exploiting this MICI-based feature dissimilarity and retains a representative feature for each cluster. The set of representatives forms the reduced set of features and its size is controlled by a user-defined parameter (see [35] for further details). Hereafter, we denote $FMR^{I}_{PCNN\_shape}$ and $FMR^{I}_{NSCT\_texture}$ as the reduced PCNN-shape and NSCT-texture feature vectors associated to image I, respectively.

4. Computation of Approximate Class of Image

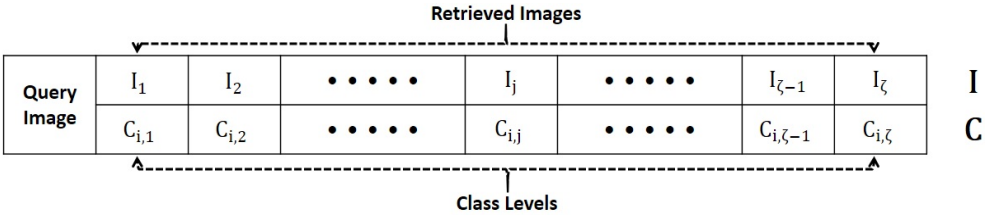


Figure 2: Representation of retrieved images considering $\zeta = 20$ with class level information.

The basic aim of this operation is to reduce the search space and to improve the retrieval accuracy simultaneously. The class information of each X-ray image of the stored database is represented by pre-defined class label. In the first phase, a query image is presented to the system. Based on the PCNN-shape features, most similar images are retrieved according to Euclidean distance measure with respect to the query image. The first ζ ("*scope size*") number of most similar retrieved images are used to identify the top "p" number of most probable classes' information based on the proposed SPS (s_i) where i represents the individual class level, $1 \leqslant i \leqslant L$

considering total L number of different classes of X-ray images stored in the database.

Considering the scope size ($\zeta =$) 20, Fig. 2 describes the retrieved image representation where I_1 is the most similar and I_ζ is the least similar image to the query image. The corresponding class levels ($C_{i,j}$) of the retrieved images are also shown in Fig. 2, where $C_{i,j}$ represents that the retrieved image I_j corresponding to the position j in the scope size ζ belongs to the class i; where, $1 \leqslant i \leqslant L$ and $1 \leqslant j \leqslant \zeta$. Using the class levels and positional information of the retrieved images in ζ, SPSs are computed.

Initially, all the s_i are set to zero considering equal importance for all the different classes of images present in the database. Assuming, among the ζ number of retrieved images $I_{j_1}, I_{j_2}, ..., I_{j_m}$ of $\mathbf{I}$ belong to the same class i i.e. the corresponding cells $C_{i,j_1}, C_{i,j_2}, ..., C_{i,j_m}$ of $\mathbf{C}$ contain the same class level i, where j_m represents the position of the retrieved images in ζ and m is the number of retrieved images of the ith particular class. The s_i is computed as:

$$s_i = \frac{j_1 + j_2 + ... + j_m}{\zeta}; 1 \leqslant j_1, ..., j_m \leqslant \zeta \tag{4}$$

Let the number of distinct classes present in $\mathbf{C}$ be $i_1, i_2, ..., i_n$; where $1 \leqslant i_1, ..., i_n \leqslant L$. Assuming, $\mathbf{A} = \{s_{i_1}, s_{i_2}, ..., s_{i_n}\}$ representing the set of calculated s_i, the sorted (in descending order) set $\mathbf{B} = \{s_{i'_1}, s_{i'_2}, ..., s_{i'_n}\}$ is computed, where $s_{i'_1} \geq s_{i'_2} \geq ... \geq s_{i'_n}$ and $\{i'_1, i'_2, ..., i'_n\}$ is a permutation of $\{i_1, i_2, ..., i_n\}$.

For fixed ζ (here, $\zeta = 20$) a threshold is calculated in the following way

$$\delta = 1 + \frac{\zeta - 1}{\zeta} + \frac{\zeta - 2}{\zeta} + ... + \frac{\zeta - \zeta'}{\zeta}; \text{where } \zeta' = \frac{\zeta}{2} + 1 \tag{5}$$

The first p classes $i'_1, i'_2, ..., i'_p$ are chosen from the set $\{i'_1, i'_2, ..., i'_n\}$ where $p \leq n$ and $s_{i'_1} \geq s_{i'_2} \geq ... \geq s_{i'_p}$ such that

$$s_{i'_1} + s_{i'_2} + ... + s_{i'_p} \geq \delta \tag{6}$$

5. Fuzzy Index Based Relevance Feedback Mechanism

Human perception of image similarity is subjective, semantic and task dependent. Hence, CBMIR system cannot deliver satisfactory result in general by merely relying on low-level features extracted in an unsupervised way. The user's response has to be incorporated in the retrieval loop as a source of feedback. By doing so, the system can iteratively adopt to the query results to meet user-specific requirements. In formal CBMIR system, an image I is usually represented by an N dimensional feature vector: $I = \{f_q |\ q = 1, 2, 3, ..., N,\ f_q \in \mathbb{R}^{\mathbb{N}}\}$, and f_q represents the q^{th} feature component. The commonly used decision function for measuring similarity between query image (I_{qr}) and the database image I, is represented as

$$D(I, I_{qr}) = \sum_{q=1}^{N} w_q ||f_q(I) - f_q(I_{qr})|| \tag{7}$$

where $\|f_q(I) - f_q(I_{qr})\|$ is the distance between the q^{th} feature component of an image I and the q^{th} feature component of the query image I_{qr} and w_q is the weight assigned to it. The weight should be adjusted such that the feature has small variation over the relevant images and large variation over

the irrelevant images. Recently, RFM has been used in CBMIR systems to overcome the "*semantic gap*" problem [27, 40, 24]. As, human perception and responses are subjective as well as ambiguous in nature. Therefore, this subjectivity and ambiguity has to be properly handled for designing an efficient RFM. Fuzzy set theoretic [41] models try to mimic human reasoning and has the capability of handling uncertainty arising from ill-defined, incomplete, defective and imprecise input. Fuzzy logic based RFM is proved to be very effective in CBMIR based applications [42]. In light of this, we have developed a novel relevance feedback mechanism based on index of fuzziness (*fuzzy-index*-RFM).

Assuming, k similar images $I_s = \{I_1, I_2, \ldots, I_k\}$ are returned to the user by the proposed system. From these returned images, the user marks the relevant and irrelevant images. Let, $I_c = \{I_r, I_{ir}, I_s\}$, where I_r = relevant images, I_{ir} = irrelevant images, I_s = system's returned images: $I_r \cup I_{ir} = I_s$ and $I_r \cap I_{ir} = \varnothing$. The information from I_r, I_{ir} and I_s are combined to compute the relative importance of the individual features by fuzzy Feature Evaluation Index (FEI).

A fuzziness measure of a fuzzy set express the average amount of ambiguity in making a decision whether an element belongs to the set or not [41]. One such measure, the index of fuzziness denotes the degree of ambiguity or fuzziness present in a set by measuring the distance between the membership values of the fuzzy elements of the fuzzy set A and its nearest ordinary set $\tilde{A}$. The index of fuzziness is defined as

$$\begin{aligned}\nu_p(A) &= (\frac{2}{n^{\frac{1}{p}}})d^p(A,\tilde{A}) \\ &= (\frac{2}{n^{\frac{1}{p}}})[\sum_{i=1}^{n}\{min(\mu_A(x_i), 1-\mu_A(x_i))\}^p]^{1/p}\end{aligned} \tag{8}$$

where $d(A,\tilde{A})$ denotes the distance between the fuzzy set A and its nearest ordinary set $\tilde{A}$ and n is the number of elements in the fuzzy set. An ordinary set $\tilde{A}$ nearest to the fuzzy set A is defined as

$$\begin{aligned}\mu_{\tilde{A}}(x_i) &= 0 \quad \text{if} \quad \mu_A(x_i) \leq 0.5 \\ &= 1 \quad \text{if} \quad \mu_A(x_i) > 0.5\end{aligned} \tag{9}$$

The value of p depends on the type of distance function used. For example, $p = 1$ for generalized Hamming distances whereas $p = 2$ for Euclidean distance. The corresponding indices of fuzziness are called the "*linear index of fuzziness*" $\nu_{L1}(A)$ and "*quadratic index of fuzziness*" $\nu_Q(A)$, respectively. In this proposed CBMIR system we have used the "*quadratic index of fuzziness*" for finding out the relative importance w_q of the individual feature f_q. The $\mu_A(x_i)$ is computed using a standard S-type membership function [41]:

In our case, A represents the q^{th} column of $I_c = \{I_r, I_{ir}, I_s\}$ with x_i denoting the corresponding feature value and $\mu(x_i)$ as the corresponding fuzzy membership value. The FEI for the q^{th} feature is computed as follows:

$$FEI_q = \frac{\nu_Q(f_q^{I_s})}{\nu_Q(f_q^{I_r}) + \nu_Q(f_q^{I_{ir}})} \tag{10}$$

Lower value of FEI_q indicates better quality of importance of the q^{th} feature in recognizing and discriminating different classes. The weight w_q is computed as a function of the evaluated (FEI_q) as shown below:

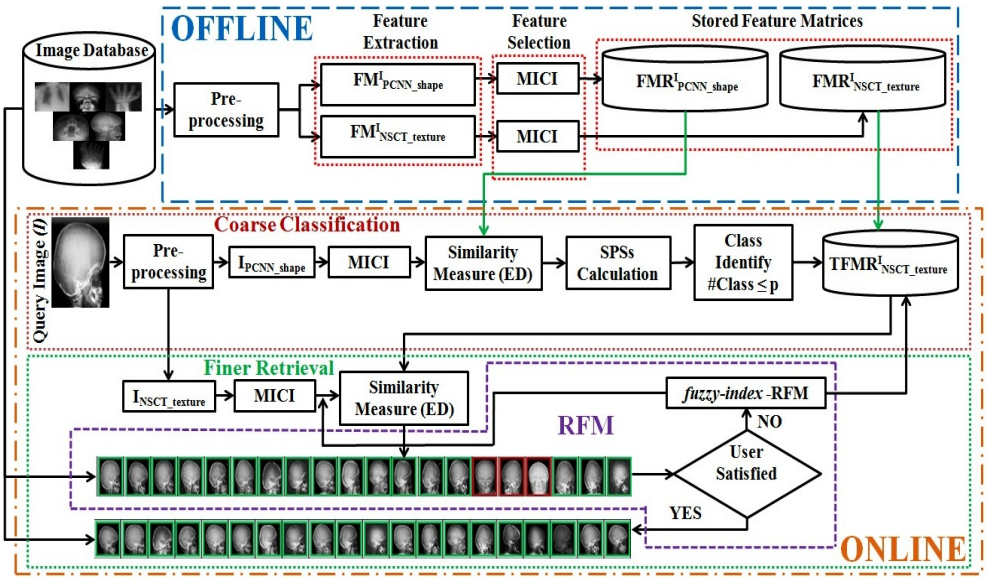


Figure 3: Overview of the proposed CBMIR system.

$$w_q = FEI_q \tag{11}$$

In the first pass, all features are considered to be equally important. Hence, $w_1 = w_2 ... = w_q = ... = w_N = 1$. The feature spaces of the relevant images are therefore altered in a similar fashion after updating the components with w_q. As a result, the ranks of the relevant images are not affected much.

6. Summary of the Proposed CBMIR System

The proposed CBMIR system consists of two main phases as shown in the block diagram of Fig. 3: an "OFFLINE" phase and an "ONLINE" phase. In the "OFFLINE" phase two different features ($FM^I_{PCNN\_shape}$ and $FM^I_{NSCT\_texture}$) are extracted from each image in the stored image database after pre-processing. The dimension of the extracted feature vectors are

reduced by MICI feature selection scheme. The reduced feature vectors are then stored in two different feature database $FMR^{I}_{PCNN\_shape}$ and $FMR^{I}_{NSCT\_texture}$ along with the class levels of the images. The "ONLINE" phase of the proposed system consists of two steps described below.

Coarse Classification:

1. Input the query image I.
2. Compute the Euclidean distances between the compact PCNN-shape feature vector of the query image (I_{PCNN_shape}) and each reduced feature vector of $FMR^{I}_{PCNN\_shape}$.
3. Sort the distances in ascending order.
4. Retrieve the class level information of the top ζ images corresponding to the first ζ sorted distances.
5. Compute the SPSs for the top ζ images and sort them in ascending order.
6. Choose the first p classes which satisfy Eq. 6.

Finer Retrieval:

1. Construct a temporary feature database $TFMR^{I}_{NSCT\_texture}$ from the $FMR^{I}_{NSCT\_texture}$ database consisting only the feature vectors corresponding to the images belonging to the top p identified image classes.
2. Compute the Euclidean distances between the compact NSCT based query feature vector $I_{NSCT\_texture}$ and each reduced feature vector of $TFMR^{I}_{NSCT\_texture}$.
3. Sort the distances in ascending order and display the top ζ images.
4. If the user is satisfied with the displayed result, then stop.

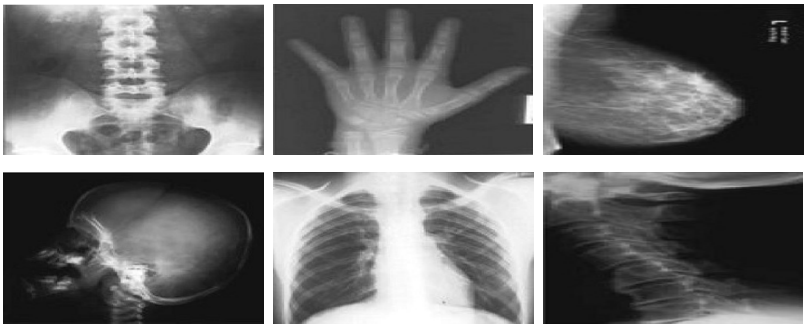


Figure 4: Example radiographs taken from the IRMA database: (left to right) top row: abdomen, limbs, breast; bottom row: skull, chest, and spine.

5. Otherwise, take the user's feedback response by marking the relevant images from the irrelevant images in the displayed images.
6. Compute the fuzzy feature evaluation index based upon the user's response (described in section 5).
7. Go to step 2, taking into consideration the ranked fuzzy features to recompute the Euclidean distances.

7. Experimental Results and Discussion

Extensive experiments were carried out to evaluate the performance (both quantitatively and qualitatively) of the proposed CBMIR system.

7.1. Experimental Setup

The experiments were carried out on a Dell Precision T7400 PC with 4GB RAM using MATLAB R2014b. All the experiments were performed on the diverse radiographic medical image collection of the IRMA-2009 dataset [13, 4]. The X-ray images are noisy with irregular brightness and contrast, and sometimes contain dominant visual artifacts such as artificial limbs and X-ray

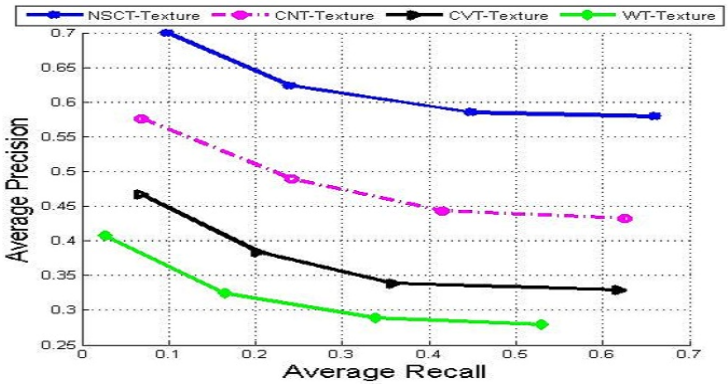


Figure 5: Performance comparison of different transforms in terms of AP and AR curves.

frame borders. This database contains $14,410$ images divided into 193 distinct categories, where the number of images in each category varies largely (from 1 to $2,314$ images). In the experiments, we only took the categories which contain at least 50 images. This resulted in a subset of database containing $10,902$ images of 57 categories. The images in this database are organized into various semantic categories (by labelling them with the four-axis IRMA code) by expert human observer according to the imaging modality, the examined region, the image orientation with respect to the body, and the biological system under evaluation. Some of the example radiographic images of different classes are shown in Fig. 4. We used "*pyrexc*" and "*sinc*" as the pyramid filter and the orientation filter in NSCT decomposition, respectively. The performance on the database was evaluated by considering each image as the query image (leaving-one-out) and by measuring the classification accuracy as well as Average Precision (AP) and Recall (AR). During the experiments, top 20 retrieved images were used to compute the precision and recall values.

7.2. Effectiveness: Feature Representation and Selection

The first set of experiments we conducted were to evaluate the effectiveness of the feature representation technique used in the proposed CBMIR system. For these experiments, we only considered the results obtained through the proposed CBMIR system without *fuzzy-index*-RFM.

At first, we evaluated the performance of NSCT against other different transforms (WT, CVT, CNT) and the outcome is described in Fig. 5. We considered all the $10,902$ (of 57 different classes) images and Euclidean distance measure considering different scope sizes (5, 10, 15, 20). In each run we took 1 image from the $10,902$ images as the query image and the rest $10,901$ images as the stored labeled image database. It is clear from the Fig. 5 that NSCT performs much better than WT, CVT and CNT in CBMIR domain.

Moreover, we tried to select the most effective PCNN-shape and NSCT-texture features after applying MICI for efficient retrieval. The idea was to select the optimal number of features that allows for the best category discrimination considering both shape and texture. To this end, we trained a multi-class Least-Squares Support Vector Machine (LS-SVM) classifier using a one-versus-one strategy with a 10-fold cross-validation on the parametrization. A share of 60% labeled samples from the dataset was selected as the training set, while the remaining 40% was reserved for the test set. We evaluated the performance of the classifier using different shares of selected features (20%, 40%, 60% and 100%) in terms of misclassification error on IRMA dataset. According to the results reported in Fig. 6, we can clearly see that the feature selection step has an impact on the retrieval performance. The best results are achieved with a share of 40% PCNN-shape and

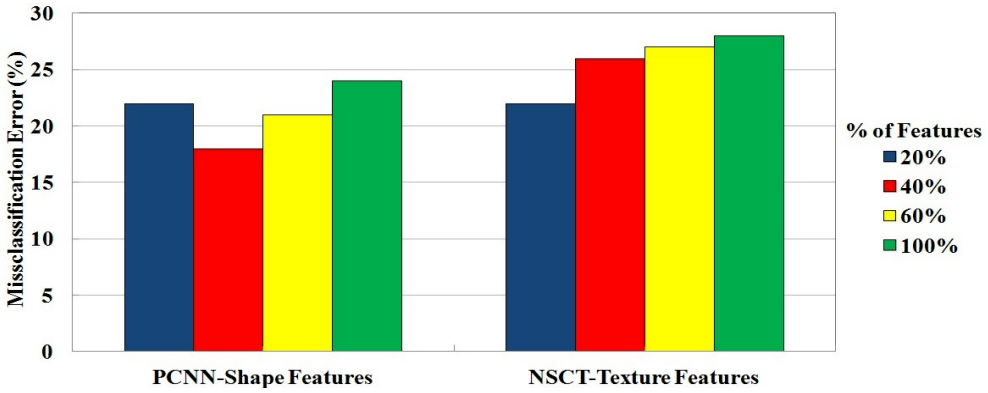


Figure 6: Performance effectiveness of different shares of PCNN-shape and NSCT-texture features.

20% NSCT-texture feature on IRMA dataset. Thus, empirically, the dimensions of PCNN-shape and NSCT-texture feature vectors were reduced from 70 and 448 to 28 and 90, respectively.

The next experiment we performed was to find out the best sequence of feature combination for radiographic images considering two different feature representations (selected PCNN-shape and NSCT-texture features from the previous experiments). The graph of Fig. 7 shows the performance of compact PCNN-shape and NSCT-texture feature vectors with various combinations, considering various scope size ζ (5, 10, 15 and 20 images/frame). The graph shows the performance of the CBMIR system using compact PCNN-shape and NSCT-texture features, combination of compact PCNN-shape + NSCT-texture features, compact NSCT-texture followed by compact PCNN-shape and compact PCNN-shape followed by compact NSCT-texture, respectively. Here, "$\Rightarrow$" means "followed by". For e.g., PCNN-shape $\Rightarrow$ NSCT-texture means PCNN-shape followed by NSCT-texture. The numbers in the

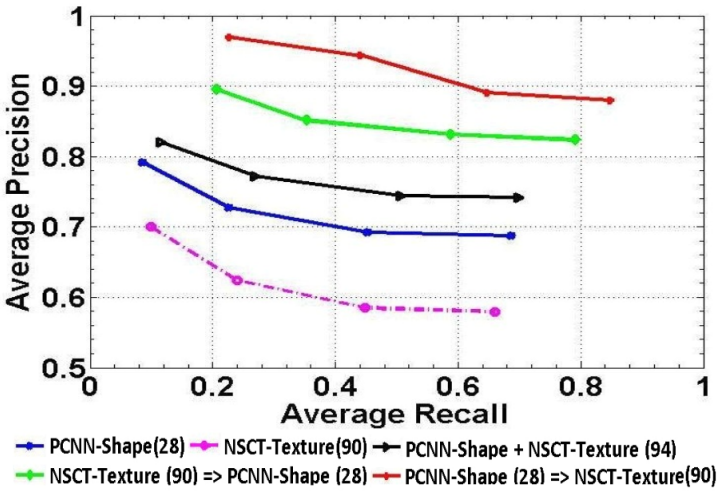


Figure 7: Performance comparison of different feature sequence combinations.

bracket denote the dimension of the selected feature representation. From the graph of Fig. 7, we can see that the compact NSCT performs worse than compact PCNN-shape features whereas the combination of compact PCNN-shape + NSCT-texture performs better than only PCNN-shape and NSCT-texture based feature representation schemes. Therefore, the proposed feature representation scheme (PCNN-shape followed by NSCT-texture) provides the best retrieval results. We also conducted another experiment to find out the time requirement of different feature representation schemes. Table 1 shows the average retrieval time using all and selected features with Euclidean distance measure and without considering RFM. From Table 1, it can be seen that the "PCNN-shape⇒ NSCT-texture" performs well in terms of classification accuracy while having a satisfactory time requirement as compared to the other feature combinations.

To support our choice of setting the scope size ζ as 20 (which is a general

Different Methods	With All Features		With Compact Features	
	Accuracy(%)	Time(sec)	Accuracy(%)	Time(sec)
PCNN-shape	87.19	1.104	85.89	0.525
NSCT-texture	75.12	2.915	74.45	1.146
PCNN-shape ⇒ NSCT-texture	93.69	3.102	93.03	1.169

Table 1: Comparisons of different features vectors in terms of time requirement and classification accuracy.

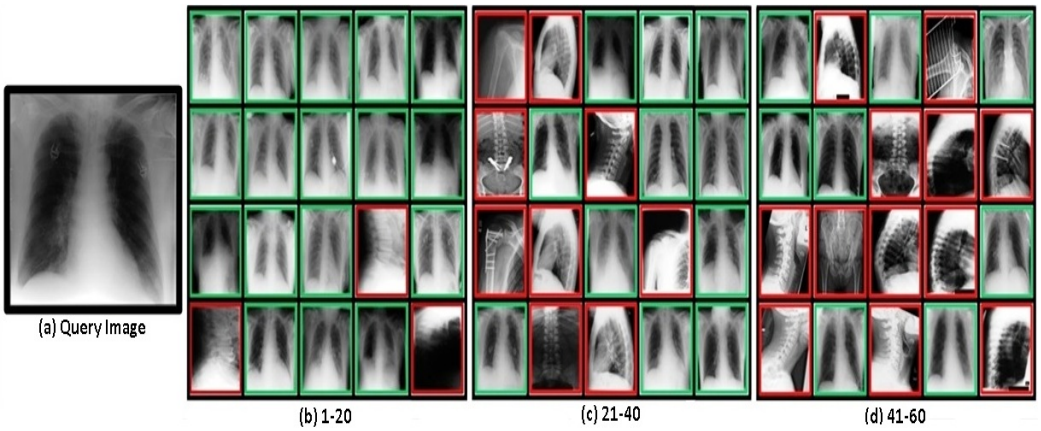


Figure 8: Performance comparison considering different scope size ζ. Green boxes $\rightarrow$ relevant images, red boxes $\rightarrow$ irrelevant images.

and popular convention), we evaluated the performance of the proposed CBMIR system with different scope sizes - starting from 5 to 60 in step-size 5. An example is given in Fig. 8 which shows the fact that increasing the scope size would result in showing more and more number of irrelevant images. In this case, it becomes difficult for the medical personal to provide correct response to the system as the person has to choose the relevant/irrelevant images from a large number of displayed images. Moreover, the effectiveness of the RFM also depends on the correctness of the user's response. With a large scope size, there is a chance of introducing erroneous responses to the

RFM.

7.3. Effectiveness: Fuzzy Index Based RFM

To test the effectiveness of the novel *fuzzy-index*-RFM, top 20 retrieved images were used to compute the average precision values. The graph of Fig. 9, shows the effectiveness of the proposed system in terms of average precision for different phases of the system and the number of iterations. We can clearly see the incremental improvement in the retrieval accuracy achieved by the proposed system over the different stages of the system. The proposed CBMIR system achieves near about 78% average precision after the coarse retrieval (using compact PCNN-shape feature). Based on the retrieval results provided by the coarse retrieval steps, the system works on a reduced search space (consisting only the images of the top p classes returned by coarse retrieval steps). The AP increases to approximately 93% in the finer retrieval stage (using compact NSCT-texture features). If the user is not satisfied with the retrieval result, based on the feedback response of the user, the proposed *fuzzy-index*-RFM tries to improve the result further. From the graph of Fig. 9, we can see that the proposed *fuzzy-index*-RFM achieves the highest retrieval accuracy after only $3 - 4$ iterations.

The Fig. 10 shows the visual results (considering $\zeta = 20$) achieved by the different steps of the proposed CBMIR system for some query images (left column's image marked by black box). The row 1 of the right column shows the coarse classification results, when retrieval is done on whole database. The row 2 of the right column shows the results of limited classes, which has been selected using similarity position score. Based on the user feedback, final results are shown in the last row (row 3 of the right column) of the

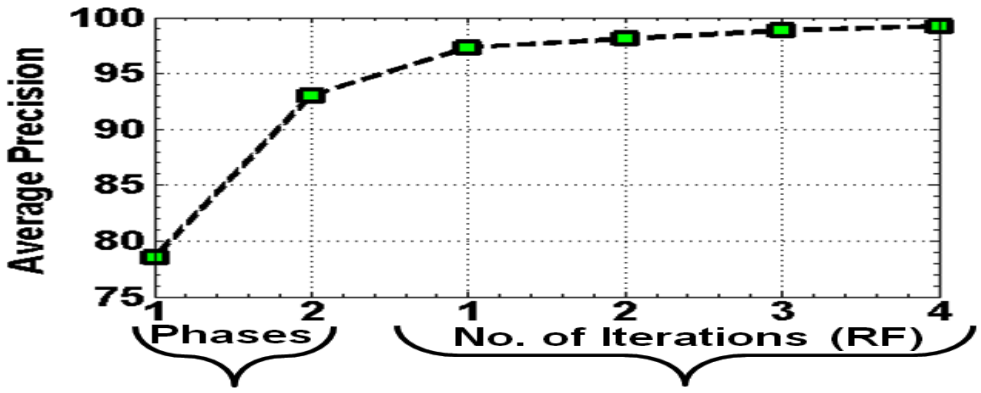
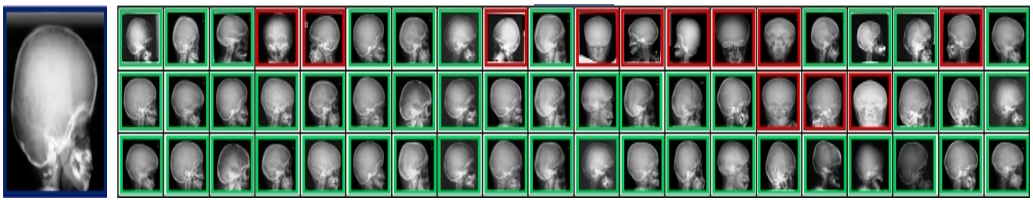


Figure 9: Effectiveness of the proposed *fuzzy-index*-RFM.

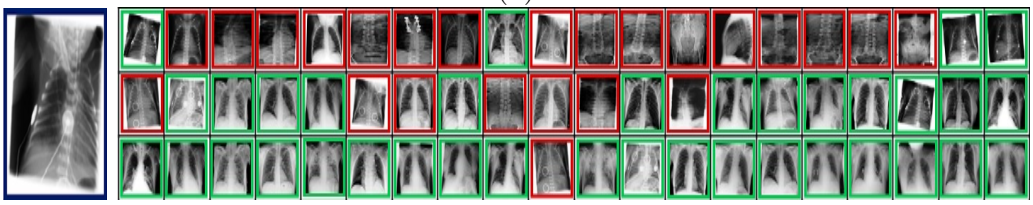
respective query image. We can see that all the retrieved images after the 3^{rd} iteration of the *fuzzy-index*-RFM are almost from the same class as that of the query image with improved ranking. In Fig. 10(b), after using the proposed *fuzzy-index*-RFM, only one irrelevant image present in the top 20 retrieved images as shown in row 3. This might because that the pathology information has not being considered in our proposed system. Hence, from the qualitative and quantitative results, we can say that our system achieves AP of approximately 93% without using *fuzzy-index*-RFM and it increases to approximately 98% by using proposed *fuzzy-index*-RFM on the subset of IRMA-2009 database's images.

7.4. Effectiveness: Comparisons with State-of-the-Art CBMIR Systems

An exact comparison across different existing CBMIR schemes reported in the literature is a complex task. This is due to the different usages of medical image database in these studies. However, to further justify the proposed system, we compared it against several state-of-the-art CBMIR schemes for medical images [25, 26, 27, 21, 43, 24, 17, 30], without involving



(a)



(b)

Figure 10: Visual performance of the proposed system (black box → query image, green box → relevant image and red box → irrelevant image): Left column's image is the query image. Row 1 of right column shows the coarse classification results ((a): 11/20; (b): 4/20). Row 2 of right column shows the results when limited search from pre-selected number of classes ((a): 17/20; (b): 13/20, ranking improved) using similarity positional score. Row 3 of right column shows the refinement of second result with *fuzzy-index*-RFM after 3^{rd} iteration ((a):20/20; (b): 19/20, ranking improved).

Scheme	#Images	#Class	Accuracy (%)
Keysers et al. [25]	1,617	6	92.00%
Lehmann et al. [26]	6,231	81	85.50%
Rahman et al. [27]	5,000	20	81.96%
Pourghassem et al. [21]	9,100	40	94.00% (17 merged classes)
Iakovidis et al. [43]	10,000	116	90%
Rahman et al. [24]	5,000	30	81.01%
Fesharaki et al. [17]	2,158	18	93.60%
Srinivas et al. [30]	2,600	–	87.50%
Proposed	10,902	57	93.03%

Table 2: Comparisons with state-of-the-art CBMIR systems. **#Images**: Number of images in the database; **#Class**: Number of image classes in the database.

the RFM in the retrieval loop. The comparison is summarized in Table 2. It can be seen from the Table 2 that many [25, 17, 30] of the existing CBMIR systems works efficiently for small image database ($< 3,000$ images) having very limited classes (< 20 categories) of medical images. But, the effectiveness of these systems for larger medical image database consisting of much different modalities of medical images is unknown. The schemes described in [27, 24] both achieved approximately 81% accuracy on medical image database having $5,000$ images, but considering only 20 and 30 image classes, respectively. Only the CBMIR systems proposed in [26, 21, 43] have considered bigger image database ($> 5,000$ images) consisting of relatively higher number of different classes (≥ 40) of images. Specifically, the scheme of [21] reported average classification accuracy of 94% on a database of $9,100$

images from 40 classes in which 17 classes are merged. But, considering all the 40 different image classes-the accuracy reduces to approximately 64% [21]. The scheme described in [43] considered the largest diverse image database ($10,000$ images of 116 classes) among all the existing techniques mentioned in the Table 2. But even with a much higher dimensional feature representation scheme (compared to the proposed feature vectors), this scheme achieved approximately 90% accuracy. Compared with these above mentioned methods, the proposed system performs (comparably) better (93%) because of the use of a novel hierarchial two-stage retrieval procedure along with the novel SPS and fuzzy-index RFM. The global (shape) to local (texture) hierarchial system follows the vision system (VS) characteristics. The use of PCNN mimicking the mammalian VS, provides an efficient descriptor to effectively represent the shape characteristics of the radiographic images using the global coupling and pulse-synchronization properties. On the other hand, the use of NSCT as a multiscale geometric analysis tool to represent the texture of the x-ray images is also proven to be an effective solution. This also leads to a lower dimensional (using MICI) compact feature representation. It should also be noted that inclusion of the novel fuzzy-index RFM to the proposed system also enhances the retrieval performance (approximately 98%). The proposed system does not use any computationally expensive machine learning techniques (like support vector machine, fuzzy C-means, decision tree etc.) which makes it easily implementable in hardware and requires only compact feature vector representations of the stored database's images with their class levels during retrieval.

8. Conclusion

In this paper, we have proposed a novel efficient system for retrieval of radiographic medical images based on a two-stage hierarchical classification system. Main importance is given on constructing effective feature vector representing the image's content and search-space reduction without compromising the performance. Furthermore, to handle the uncertainty and ambiguity of the user response, a novel *fuzzy-index*-RFM is incorporated to the proposed system. Experimental evaluations of the proposed system have shown that our approach is effective and improves the retrieval performance of CBMIR systems significantly compared to other state-of-the-art methods. In future, we intend to explore our proposed scheme for the retrieval of various kinds of medical images, such as endoscopic, colonoscopy images etc. We are also interested in considering the generalizations of our proposed scheme to accommodate different types of features, different similarity measures and improved RFM within our framework.

Acknowledgements

The collection of images used in this paper is courtesy of Dr. T. M. Lehmann, Image Retrieval in Medical Application (IRMA) Group, Department of Medical Informatics, RWTH-Aachen, Germany (http://irma-project .org). Malay K. Kundu acknowledges the Indian National Academy of Engineering (INAE) for their support through INAE Distinguished Professor fellowship. This work was supported by Machine Intelligence Unit, Indian Statistical Institute, Kolkata-108, India, (Internal Academic Project).

References

[1] T. Weyand, T. Deselaers, Combining content-based image retrieval with textual information retrieval, Research Project, RWTH Aachen (2005) 1–23.

[2] K. H. Hwang, H. Lee, D. Choi, Medical image retrieval: Past and Present, Healthc Inform Res. 18 (1) (2012) 3–9.

[3] H. Shen, D. Ma, Y. Zhao, H. Sun, S. Sun, R. Ye, L. Huang, B. Lang, Y. San, MIAPS: A web-based system for remotely accessing and presenting medical images, Computer Methods and Programs in Biomedicine 113 (2014) 266–283.

[4] T. M. Lehmann, M. O. Guld, C. Thies, B. Fischer, D. Keysers, M. Kohnen, Content-based image retrieval in medical applications for picture archiving and communication systems, in: Proc. SPIE Med. Img.: PACS and Integrated Medical Inf. Syst: Design and Evaluation, Vol. 5033, 2003, pp. 109–117.

[5] C. T. Jr., A. J. M. Traina, M. R. B. Araujo, J. M. Bueno, F. J. T. Chino, H. Razente, P. M. A.-Marques, Using an image-extended relational database to support content-based image retrieval in a PACS, Computer Methods and Programs in Biomedicine 80 (1) (2005) 571–583.

[6] H. Greenspan, A. T. Pinhas, Medical image categorization and retrieval for PACS using the GMM-KL framework, IEEE Trans. Inf. Technol. Biomed. 11 (2) (2007) 190–202.

[7] C. B. Akgl, D. L. Rubin, S. Napel, C. F. Beaulieu, H. Greenspan, B. Acar, Content-based image retrieval in radiology: Current status and future directions, J. of Digital Imaging 2 (24) (2011) 208–222.

[8] M. Chowdhury, S. Das, M. K. Kundu, Effective classification of radiographic medical images using LS-SVM and NSCT based retrieval system, in: Proc. 5th Int. Conf. on Computers And Devices for Communication (CODEC 2012), 2012, pp. 1–4.

[9] P. Welter, B. Fischer, R. W. Gunther, T. M. Deserno, Generic integration of content-based image retrieval in computer-aided diagnosis, Computer Methods and Programs in Biomedicine 108 (2012) 589–599.

[10] H. D. Tagare, C. C. Jaffe, J. Duncan, Medical image databases: a content-based retrieval approach, J. Am. Med. Inform. Assoc. 4 (3) (1997) 184–198.

[11] A. Kumar, J. Kim, W. Cai, M. Fulham, D. Feng, Content-based medical image retrieval: a survey of applications to multidimensional and multimodality data, J. Digit Imaging. 26 (6) (2013) 1025–1039.

[12] H. Muller, N. Michoux, D. Bandon, A. Geissbuhler, A review of content-based image retrieval systems in medical applications – clinical benefits and future directions, Int. J. Medical Inf. 73 (1) (2004) 1–23.

[13] M. O. Guld, M. Kohnen, D. Keysers, H. Schubert, B. B. Wein, J. Bredno, T. M. Lehmann, Quality of DICOM header information for image categorization, Proc. SPIE Int. Symp. Med. Img. 4685 (2002) 280–287.

[14] J. K. Cramer, A. G. S. de Herrera, S. B. D. D. Fushman, S. Antani, H. Muller, Evaluating performance of biomedical image retrieval systemsan overview of the medical image retrieval task at imageCLEF 2004–2013, Computerized Medical Imaging and Graphics 39 (2015) 55–61.

[15] K. Iftekharuddin, M. Prajna, S. Samanth, M. Indhukuri, Mega voltage x-ray image segmentation and ambient noise removal, in: Proc of Second Joint EMBS/BMES Conference, Vol. 2, 2002, pp. 1111–1113.

[16] M. Kirchner, J. Fridrich, On detection of median filtering in digital images, in: Proc. of Media Forensics and Security II, part of the IS&T-SPIE Electronic Imaging Symposium, Vol. 7541, 2010, pp. 754110–754121.

[17] N. J. Fesharaki, H. Pourghassem, Medical X-ray image hierarchical classification using a merging and splitting scheme in feature space, J Med Signals Sens 3 (3) (2013) 150–163.

[18] C. Hong, J. Zhu, Hypergraph-based multi-example ranking with sparse representation for transductive learning image retrieval, Neurocomputing 101 (2013) 94–103.

[19] N. Nacereddine, D. Ziou, L. Hamami, Fusion-based shape descriptior for weld defect radiographic image retrieval, Int. J. Adv. Manuf. Technol. 68 (2013) 2815–2832.

[20] L. E. Bray, S. J. Tsai, E. S. Jimenez, Exploring the feasibility of traditional image querying tasks for industrial radiographs, in: Proc. SPIE Medical Applications of Radiation Detectors, Vol. 9594, 2015, pp. 1–12.

[21] H. Pourghassem, H. Ghassemian, Content-based medical image classification using a new hierarchical merging scheme, Comput. Med. Imag. Grap. 32 (2008) 651–661.

[22] M. M. Rahman, B. C. Desai, P. Bhattacharya, Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion, Comput. Med. Imag. Graph. 32 (2) (2008) 95–108.

[23] B. C. Ko, S. H. Kim, J.-Y. Nam, X-ray image classification using random forests with local wavelet-based cs-local binary patterns, J Digit Imaging. 24 (6) (2011) 1141–1151.

[24] M. M. Rahman, S. K. Antani, G. R. Thoma, 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) (2011) 640–646.

[25] D. Keysers, J. Dahmen, H. Ney, B. B. Wein, T. M. Lehmann, Statistical framework for model-based image retrieval in medical applications, J. Electronic Imaging 12 (1) (2003) 59–68.

[26] T. M. Lehmann, M. O. Guld, T. Deselaers, D. Keysers, H. Schubert, K. Spitzer, Automatic categorization of medical images for content-based retrieval and data mining, Comput. Med. Imaging. Graph 29 (2) (2005) 143–155.

[27] M. M. Rahman, P. Bhattacharya, B. C. Desai, A framework for medical image retrieval using machine learning and statistical similarity

matching techniques with relevance feedback, IEEE Trans. Inf. Technol. Biomed. 11 (1) (2007) 58–69.

[28] U. Avni, H. Greenspan, E. Konen, M. Sharon, J. Goldberger, X-ray categorization and retrieval on the organ and pathology level using patch-based visual words, IEEE Trans. Med. Imaging 30 (3) (2011) 733–746.

[29] M. Behnam, H. Pourghassem, Optimal query-based relevance feedback in medical image retrieval using score fusion-based classification, J. Digit. Imaging 28 (2015) 160–178.

[30] M. Srinivas, R. R. Naidu, C. Sastry, C. K. Mohan, Content based medical image retrieval using dictionary learning, Neurocomputing 168 (2015) 880–895.

[31] X. Liu, H. R. Tixhoosh, J. Kofman, Generating binary tags for fast medical image retrieval based on convolutional nets and radon transform, in: Proc of Int. Joint. Conf. on Neural Networks, IEEE, 2016, pp. 1–6.

[32] M. Chowdhury, S. R. Bulo, R. Moreno, M. K. Kundu, O. Smedby, An efficient radiographic image retrieval system using convolution neural network, in: Proc of Int. Conf. on Pattern Recognition, IEEE, 2016, pp. 1–4.

[33] J. Johnson, M. Padgett, PCNN models and applications, IEEE TNN 10 (3) (1999) 480–498.

[34] A. L. da Cunha, J. Zhou, M. N. Do, The nonsubsampled contourlet transform: Theory, design, and applications, IEEE Trans. Image Process. 15 (10) (2006) 3089–3101.

[35] P. Mitra, C. A. Murthy, S. K. Pal, Unsupervised feature selection using feature similarity, IEEE Trans. Pattern Anal Mach Intell 24 (3) (2002) 301–312.

[36] L.-j. Zhi, S.-m. Zhang, D.-z. Zhao, H. Zhao, S.-k. Lin, A new two-step method for medical image retrieval, in: Proc of Int. Conf. on Inf. Tech. and Computer Sci., IEEE, 2009, pp. 305–309.

[37] Y. Chen, S.-K. Park, Y. Ma, R. Ala, A new automatic parameter setting method of a simplified PCNN for image segmentation, IEEE Trans. Neural Networks 22 (6) (2011) 880–892.

[38] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, Wavelet optimization for content-based image retrieval in medical databases, Medical Image Analysis 14 (2) (2010) 227–241.

[39] A. M. Garcia-Perez, The perceived image: Efficient modelling of visual inhomogeneity, Spatial Vision 6 (2) (1992) 89–99.

[40] X. Xu, D.-J. Lee, S. K. Antani, L. R. Long, J. K. Archibald, Using relevance feedback with short-term memory for content-based spine X-ray image retrieval, Neurocomputing 72 (1012) (2009) 2259–2269.

[41] S. K. Pal, D. D. Majumder, Fuzzy Mathematical Approach To Pattern Recognition, Wiley Press, New York, NY, USA, 1985.

[42] K.-H. Yap, K. Wu, Fuzzy relevance feedback in content-based image retrieval systems using radial basis function network, in: Proc. of the Int. Conf. on Multimedia and Expo, Vol. 3, 2005, pp. 177–180.

[43] D. K. Iakovidis, N. Pelekis, E. E. Kotsifakos, I. Kopanakis, H. Karanikas, Y. Theodoridis, A pattern similarity scheme for medical image retrieval, IEEE Trans. Inf. Technol. Biomed. 13 (4) (2009) 442–450.

Highlights:

1. An interactive CBMIR system using two-stage classification.

2. Two new compact feature representations using PCNN and NSCT.

3. New "*similarity positional score*" to automatically identify the top "*p*" probable classes in the first stage.

4. Reducing the "*semantic gap*" problem using a new fuzzy based relevance feedback mechanism.

5. Large and diversified experimental findings on popular and large radiographic image database (IRMA-2009).