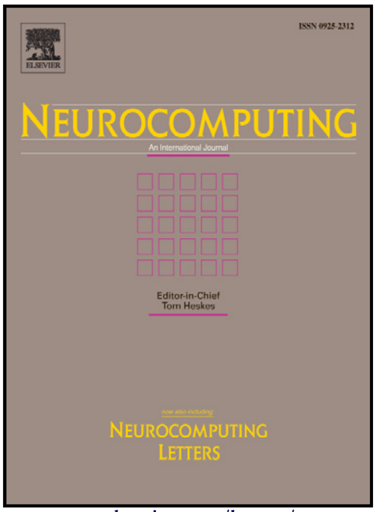
Author's Accepted Manuscript



www.elsevier.com/locate/neucom

Content Based Medical Image Retrieval Using Dictionary Learning

M. Srinivas, R. Ramu Naidu, C.S. Sastry, C. Krishna Mohan

PII: S0925-2312(15)00696-7
DOI: http://dx.doi.org/10.1016/j.neucom.2015.05.036
Reference: NEUCOM15561

To appear in: *Neurocomputing*

Cite this article as: M. Srinivas, R. Ramu Naidu, C.S. Sastry, C. Krishna Mohan, Content Based Medical Image Retrieval Using Dictionary Learning, *Neurocomputing*, http://dx.doi.org/10.1016/j.neucom.2015.05.036

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
00 (2015) 1–19
Neurocomputing

Content Based Medical Image Retrieval Using Dictionary Learning

M. Srinivas [a], R. Ramu Naidu[b], C. S. Sastry[b] and C. Krishna Mohan[a]

VIsual learninG and InteLligence (VIGIL) group [a]
Department of Computer Science and Engineering [a]
Department of Mathematics [b]
Indian Institute of Technology Hyderabad-502205, India [a,b]
Email: {cs10p002, ma11p003, csastry, ckm}@iith.ac.in

Abstract

In this paper, a clustering method using dictionary learning is proposed to group large medical databases. An approach grouping similar images into clusters that are sparsely represented by the dictionaries and learning dictionaries simultaneously via K-SVD is proposed. A query image is matched with the existing dictionaries to identify the dictionary with the sparsest representation using Orthogonal Matching Pursuit (OMP) algorithm. Then images in the cluster associated with this dictionary are compared using a similarity measure to retrieve images similar to the query image. The main features of the method are that it requires no training data and works well on the medical databases which are not restricted to specific context. The performance of the proposed method is examined on IRMA test image database. The experimental results demonstrate the efficacy of the proposed method in the retrieval of medical images.

Keywords: Clustering, Content based image retrieval, Dictionary Learning, Medical X-ray image, Rotation Invariance, Sparse representation, K-SVD, OMP.

1. Introduction

The problem of searching for similar images in a large image repository based on content is called Content Based Image Retrieval (CBIR) [1]. The traditional Text Based Image Retrieval (TBIR) approach has many practical limitations [2] like the images in the collection being annotated manually, which becomes more difficult as the size of the image collection increases. Another important limitation is the inadequacy in representing the image content. CBIR approaches are proposed to overcome the limitations of text based image retrieval.

As more and more hospitals purchase *picture archiving and communication systems* (PACS), the medical imagery world wide is increasingly acquired, transferred and stored digitally [3]. The increasing dependence on modern medical diagnostic techniques like radiology, histopathology and computerized tomography has led to an explosion in the number of medical images stored in hospitals. Digital image retrieval technique is crucial in the emerging field of medical image databases for clinical decision making process. It can retrieve images of similar nature (like same modality and disease) and characteristics. The images of various modalities are becoming an important source of anatomical and functional information for the diagnosis of diseases, medical research and education [4]. In a typical CBIR system in medical domain, subtle differences between images can not be considered irrelevant. Consequently, a Content

Based Medical Image Retrieval (CBMIR) system having a kind of invariance (with respect to any transformation) is of value [5],[6].

The major limitations associated with existing medical CBIR are:1) In most cases, physicians have to browse through a large number of images for identifying similar images, which takes lot of time. 2) Most of the existing tools for searching medical images use text based retrieval techniques. The text based image retrieval suffers from several limitations [7] such as the need for manual annotation. Thus, the existing medical image search and retrieval techniques are not very efficient in terms of time and accuracy. Another important issue in medical CBIR is to find images with similar anatomical regions and diseases. For example, in case of brain tumor images, the tumor can be at any of the different stages and an image of the tumor in a state could be in any orientation [6], [32]. So, there is a need for invariant medical image retrieval technique to find images of a similar (same stage) tumor.

Of late, sparse representation received a lot of attention from the signal and image processing communities. Sparse coding involves the representation of an image as a linear combination of some atoms in a dictionary [12]. It is a powerful tool for efficiently processing data in nontraditional ways. This is mainly due to the fact that signals and images of interest admit sparse representation in some dictionary, which may be identified based on the properties of signals at hand. Recently, dictionaries learnt from the data were found to have potential for several applications. Several interesting dictionary learning methods like K-SVD [8] and Method of Optimal Directions (MOD), [13] were developed to provide each member of database with sparse representation. The emerging filed of compressed sensing has a potential for exploiting sparsity present in medical images. This work is an attempt towards proposing a new CBMIR technique that relies on sparsity based concepts.

In particular, we propose a dictionary based clustering algorithm for grouping the images in medical databases. This clustering technique increases the retrieval speed and improves the accuracy of the results. The dictionary based methods rely on the premise that two signals belonging to the same cluster have decomposition in terms of similar atoms (columns) of a dictionary. Making use of this property, we match the input query with the appropriate cluster. The selection of features for adequately representing the class specific information is an important step in CBIR. For this, we divide the image into four sub-images of equal size. In addition, we consider another sub-image which is of same size as other four sub-images to capture the rich information available at the center of medical images. We then partition each sub-image into concentric circular regions around the center, and consider the mean and variance of pixel intensities in each region as components in the feature vector. Some image retrieval methods were proposed in the literature which made use of SVM [9], [10], [11]. It is to be emphasized here that K-SVD and SVM based methods are different in the sense that K-SVD is a dictionary learning approach banking on the concept of sparsity, which is not the case with SVM. While SVM requires some training data, the way we use K-SVD in the present work does not require any labeled data. The present CBMIR technique centers around images produced in radiology. As color and shape features are of less importance in medical domain [3], we use texture features in the present work.

The work done in this paper has the objective of categorizing (and retrieving) radiological images consisting of different organs, modality, views. We demonstrate the usefulness of our approach through extensive experimental results. For a given N, the number of clusters, we design N dictionaries to represent the clusters. We associate an image of database to a dictionary based on the sparsity criterion. Given a query image, we invoke the concept of sparsity to identify appropriate cluster, wherein we search for relevant images. The rest of the paper is organized as follows: Sections 2 and 3 give brief accounts of a survey of related works and dictionary learning. Section 4 presents the proposed content based medical image retrieval using dictionary learning method. Experiments of CBMIR application are discussed in detail in section 5. Finally, section 6 concludes this paper.

2. Related Work

Chu et al. [16] described a knowledge based image retrieval of computed tomography (CT) and magnetic resonance imaging (MRI) images. In this approach, the brain lesions were automatically segmented and represented through a knowledge based semantic model. Cai et al. [17] proposed a CBIR system for functional dynamic positron emission tomography (PET) images of the human brain, where clusters of tissue time activity from the temporal domain were used in the computation of similarity measure for retrieval. In [18], the delineations of the regions of interest were manually performed on the key frame from the stack of high resolution CT images. These were used as features to represent the entire image.

In the Bag-Of-Words (BOW) [5] framework, the image patches were sampled densely or sparsely by "interest points" detectors and were depicted by local patch descriptors like SIFT. These descriptors were used to classify liver lesions in CT images. In [6], a texture based analysis of lung CT images was proposed through Riesz wavelets. This method used SVM to learn the respective relevance of multi-scale components. Guimond et al. [19] introduced user-selected volume of interest (VOI) for the retrieval of pathological brain MRI images. In [21], group sparse representation with dictionary learning for medical image denoising and fusion was used. Wavelet optimization techniques for content based image retrieval in medical database were described in G. Quellec et al [22]. Linear discriminate analysis (LDA) based selection and feature extraction algorithm for classification and segmentation of one dimensional radar signals and two-dimensional texture and document images using wavelet packet was proposed by Etemand and Chellappa [23]. Recently, similar algorithms for simultaneous sparse signal representation and discrimination were proposed [24]-[29]. In [30], Yi. Chen et al. proposed in-plane rotation and scale invariant clustering using dictionaries. This approach provides Radon-based rotation and scale invariant clustering as applied to content based image retrieval on Smithsonian isolated leaf, Kimia shape and Brodatz texture datasets. Fei et al. [31] described a CT image denoising based on sparse representation using global dictionary. This approach improved low dose CT abdomen image quality through a dictionary learning based denoising method and accelerated the training time at the same time. Different classes of images (produced by different departments such as dermatology, pathology etc) were dealt with differently for applications such as CBIR. An excellent review of the state of art of CBMIR and future directions was presented in [32]. Several multi-resolution analysis techniques via wavelet, ridgelet, and curvelet-based texture descriptors were discussed for CBMIR [33]. The algorithm proposed therein identified various tissues based on the discriminative texture features with the aid of decision tree classification. This method too incorporated some training data for realizing its objectives.

The present paper, nevertheless, has the objective of categorizing medical images that are not restricted to a specific context. In applications of digital radiology such as computer aided diagnosis or case based reasoning, the image category is of importance [3]. It may be emphasized here that our method

- requires no training data for the classification (and retrieval) of medical data, which is in contrast to existing methods
- categorizes medical images that are not restricted to a specific context, unlike many methods available in the literature.

3. On Dictionary Learning

Most of the naturally occurring signals typically carry overwhelming amounts of data in which relevant information is often more difficult to obtain. A sparse representation, wherein a few coefficients capture most of the signal content, makes the processing faster and simpler. Such representations can be generated by decomposing signals via standard bases such as wavelets. But identifying an ideal sparse transform that can be adapted to all signals is a hopeless quest. The Dictionary Learning (DL) methods aim at designing a basis type set (called Dictionary) that provides all signals of a class with sparse representation. Of late, DL was found to have merit for several real life applications [26] [30].

Given a set of vectors $\{\mathbf{v}_i\}_{i=1}^n$, the K-SVD based DL method finds the dictionary D by solving the following optimization problem:

$$(D, \Phi) = \arg\min_{\hat{D},\hat{\Phi}} \|V - \hat{D}\hat{\Phi}\|_F^2 \text{ subject to } \|\hat{\gamma}_i\|_0 \leq T_0 \ \forall i, \quad (1)$$

where $\hat{\gamma}_i$ represents i^{th} column of $\hat{\Phi}$, V is the matrix whose columns are v_i, and T_0 is the sparsity parameter. The columns of $\hat{\Phi}$ represent the sparse solutions of $\{\mathbf{v}_i\}_{i=1}^n$ in terms of dictionary D. Here, $\|A\|_F$ denotes the Frobenius norm which is defined as $\|A\|_F = \sqrt{\sum_{ij} A^2{}_{ij}}$ and $\|v\|_0$ stands for the number of nonzero components in v, that is, $\|v\|_0 = |\{i|v_i \neq 0\}|$. The K-SVD algorithm alternates between sparse coding and dictionary update steps. Various efficient pursuit algorithms were proposed in the literature for sparse coding [14],[15]. The simplest one among all is the Orthogonal Matching Pursuit (OMP) algorithm [15]. In sparse coding step, dictionary D is fixed and representation vectors γ_i are identified for each example $\mathbf{v_i}$. Then, the dictionary is updated atom by atom in an efficient way.

4. CBMIR using Dictionary Learning

In this section, we propose a method for clustering data using dictionary learning. The present work is inspired by the ideas embedded in [30] and differs from it as follows :

- The sparsity seeking Dictionary Learning approaches typically exploit the framework of under-determined setting and hence work under some implicit assumptions on the database. In applications, nevertheless, one often encounters databases which are not big enough. As a result, the sparsity-promoting under-determined framework could not be deployed efficiently. We come to this point in the section dealing with simulation work. The present work does not require any implicit assumptions on the sizes of database and its members.
- As Radon transform is $O(N^2 logN)$ procedure, the present approach avoids using it. This, of course, results in some computational savings.

The problems stated above could be addressed by down sampling the images or by projecting them to lower dimensional spaces. Instead, the present work extracts a small set of features that describe the images well for CBMIR.

4.1. Feature Extraction

Two types of feature extraction methods are considered to represent the content of medical images. In the first feature extraction method, an image is partitioned into concentric circular regions of equal area, which provides invariant representation as shown in Fig. 1 (a).

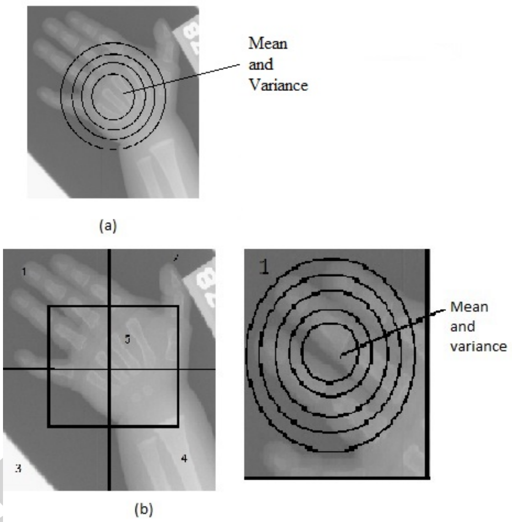


Figure 1: Proposed feature extraction techniques: (a) Image is partitioned into concentric circular regions of equal area. (b) Image is divided into sub-images and each sub-image is partitioned into concentric circular regions of equal area.

The mean and variance of pixel intensity in a circular region become components of the feature vector $\mathbf{y_j}$ and are defined as follows :

$$\mathbf{y_{j,2k-1}} = \frac{1}{P_k} \sum_{l \in J_k} I_l \tag{2}$$

$$\mathbf{y_{j,2k}} = \sum_{l \in J_k} I_l^2 - \mathbf{y^2_{j,2k-1}}, \tag{3}$$

for $j = 1, 2, \ldots, M$ and $k = 1, 2, \ldots, L$ with L standing for the number of concentring circles and the index set J_k $(k = 1, 2, \ldots, L)$ for the pixels that fall in k^{th} concentric circular region. The number I_l stands for l^{th} pixel value.

Finally, $\mathbf{y_j} = (\mathbf{y_{j,l}}, \ldots, \mathbf{y_{j,2L}})$. This approach accomplishes the rotation invariant representation of the contents of an image.

In the second feature extraction method, an image is divided into four blocks resulting in four sub-images shown in Fig. 1 (b). Also another sub-image which is of same block size as other four sub-images is considered in order to capture the rich information available at the center of medical images. Each sub-image is partitioned into concentric circular regions of equal area from which the mean and variance of pixel intensity values are computed. This feature extraction method is more suitable for medical image databases because of the rich information of medical images available at the center of images. Unlike natural images, most of the medical images are taken under standardized conditions [5], which makes them possess somewhat rich information around the center. This is, however, not a prerequisite for our method.

4.2. Proposed Method

In this section, we discuss in detail the proposed content based medical image retrieval technique using dictionary learning. To begin with, we extract the feature vectors consisting of mean and variance of pixel intensity values (as in (2) and (3)) from the images in the database. We then form initial clusters by applying K-means clustering algorithm on the extracted features. Subsequently, we generate a dictionary for each cluster using K-SVD method. By assigning the images that are sparsely represented by the dictionary, we create a new cluster for each dictionary. These clusters in turn are used to update the Dictionaries through K-SVD algorithm. The updated dictionaries are then used to update the clusters using OMP algorithm. This process goes on iteratively back and forth between dictionary and sparsity updates, until clusters converge. We match the query image with the existing dictionaries to identify the dictionary that gives the sparest representation. The images in the cluster associated with this dictionary are compared using a similarity measure to retrieve images similar to the query image. The entire process of proposed content based medical image retrieval is (shown Fig. 2) elaborated in more detail through the following steps:

Let $\{\mathbf{y}_j\}_{j=1}^{M}$ represent the set of feature vectors of the database of M images. Suppose N is the number of clusters. We obtain the initial clusters by solving the following K-means algorithm:

$$\{C_1, C_2, \ldots, C_N\} = \arg \min_{\{\hat{C}_1, \hat{C}_2, \ldots, \hat{C}_N\}} \sum_{i=1}^{N} \sum_{\mathbf{y_j} \in \hat{C}_i} ||\mathbf{y_j} - \hat{\mu}_i||, \tag{4}$$

where $\hat{\mu}_i$ is the mean of points in $\hat{C}_i$. Let C_i be the matrix containing cluster members as columns corresponding to the i^{th} cluster. Then the proposed method may be summarized as follows:

- *Dictionary update*: From the initial clusters $C_1, C_2, \ldots, C_N$, we obtain the dictionaries $D_1, D_2, \ldots, D_N$ by using the K-SVD approach (1) as follows:

  $$(D_i, \Phi_i) = \arg \min_{\hat{D}_i, \hat{\Phi}_i} ||C_\mathbf{i} - \hat{D}_i \hat{\Phi}_i||_F^2 \text{ subject to } ||\hat{\gamma}_i||_0 \leq T_0 \quad \forall i, \tag{5}$$

  satisfying $C_i \approx D_i \Phi_i, \quad i = 1, 2, \ldots, N$.

- *Cluster assignment*: From the Dictionaries so learnt, we update the clusters in this step, which is done based on the premise that two images belonging to the same cluster have decomposition in terms of similar dictionary atoms. Let D be the concatenation of all dictionaries as $[D_1, D_2, \ldots, D_N]$. Our proposed method considers obtaining the sparsest representation of $\mathbf{y_j}$ ($j = 1, 2, .., M$) in an appropriate dictionary $D_{\hat{i}}$ from:

  $$\begin{aligned} \alpha^j &= \arg \min_\omega ||\mathbf{y_j} - D\omega||_2^2 \quad \text{subject to } ||\omega||_0 \leq T_0, \\ \hat{i} &= \arg \min_i ||\mathbf{y_j} - D\delta_i(\alpha^j)||_2^2 \quad j = 1, 2, \cdots, M, \end{aligned} \tag{6}$$

  where δ_i is a characteristic function that selects the coefficients. Then $\mathbf{y_j}$ is assigned to $C_{\hat{i}}$ which is associated with the $\hat{i}^{th}$ dictionary. In (6), the first step finds T_0 number of atoms from D that sparsely describe $\mathbf{y_j}$, while the second step identifies the concentration of atoms from a particular dictionary.

We repeat the cluster assignment and dictionary update steps till there is no significant change in the clusters C_i ($i = 1, 2, .., N$). The above mentioned clustering methodology may be summarized as the following optimization problem :

$$\min_{\{D_i\},\{C_i\}} \sum_{j=1}^{M} \sum_{\mathbf{y_j} \in C_i} \min_{\alpha} \|\mathbf{y}_j - D\delta_i(\alpha)\|_2^2 + \gamma\|\alpha\|_1, \tag{7}$$

where $\gamma > 0$. Given a query image $\mathbf{x_q}$, in line with (6), we find the cluster that is closest to the query image by identifying the corresponding dictionary admitting representation to $\mathbf{x_q}$ from the following optimization problem:

$$\beta \quad = \quad \arg\min_{\omega} \|\mathbf{x_q} - D\omega\|_2^2 \quad \text{subject to } \|\omega\|_0 \leq T_0. \tag{8}$$

We consider the cluster $C_{\hat{i}}$ to be most relevant to the query image if $\hat{i}$ satisfies $\|\mathbf{x_q} - D\delta_{\hat{i}}(\beta)\|_2 < \|\mathbf{x_q} - D\delta_j(\beta)\|_2 \ \ \forall j \neq \hat{i}$. In the event of tie, that is,

$$\|\mathbf{x_q} - D\delta_{\hat{i}}(\beta)\| = \|\mathbf{x_q} - D\delta_{\hat{l}}(\beta)\| < \|\mathbf{x_q} - D\delta_j(\beta)\| \ \forall j \neq \hat{l}, \hat{i}, \tag{9}$$

we search for relevance of $\mathbf{x_q}$ in $C_{\hat{i}}$ and $C_{\hat{l}}$.

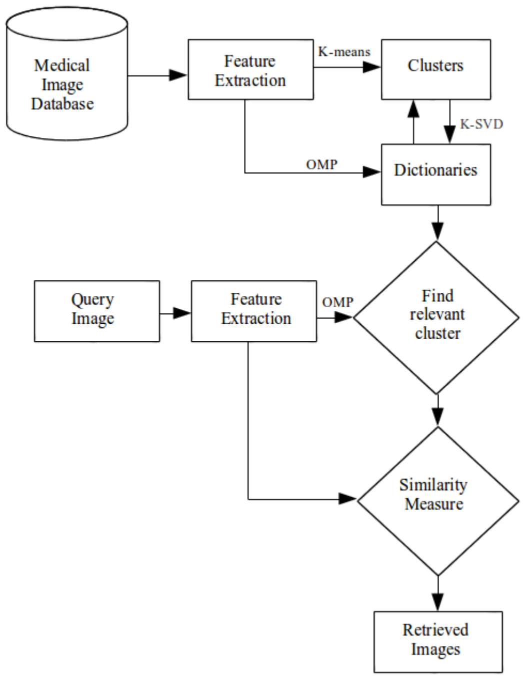


Figure 2: Block diagram of the proposed method

After determining the most relevant cluster, we identify the relevant images within the cluster using a similarity metric. To evaluate similarity between images based on the selected features, an appropriate similarity/dissimilarity metric needs to be chosen. A large class of similarity measures are used in the literature [34]. In this paper, we use three type of similarity metrics, namely, Euclidean distance (ED), Mahalanobis distance (MD) and Cross correlation (CC). The proposed algorithm is summarized as follows:

To summarize, the method proposed herein is an unsupervised technique being effective on the X-ray images of different contexts for the purpose of CBIR.

5. Experimental Results

In this section, we present our experimental results in detail and carry out an analysis of results by assigning different values to the parameters associated with our method.

Algorithm 1 :Summary of the proposed CBMIR algorithm
Step 1. Extract features from the members of medical database
Step 2. Apply K-means clustering algorithm on the extracted features to generate initial clusters
Step 3. Generate dictionary for each cluster using K-SVD method
Step 4. Create new cluster for each dictionary by assigning the images that are sparsely represented by it
Step 5. Repeat steps 3 and 4 till clusters converge
Step 6. For the query image $\mathbf{x_q}$, search for relevance in $C_{\hat{i}}$, where $D_{\hat{i}}$ provides sparest representation to $\mathbf{x_q}$.

5.1. Database Description

Majority of medical images are generally gray scale ones such as X-ray, CT etc. The ImageCLEF medical image database is made available by IRMA[1] group from the University Hospital of Aachen, Germany. This collection compiles anonymous radiographs, which were arbitrarily selected from a routine at the Department of Diagnostic Radiology, Aachen University of Technology (RWTH), Aachen, Germany. The imagery represents different ages, genders, view positions and pathologies, resulting, therefore, in varying image quality.

In the IRMA database considered for CBMIR application, each image is of size 120×120 pixels. For evaluating proposed CBMIR method, 2600 sample images of skull, breast, chest, hand etc are selected. The database members when considered in matrix form as columns result in a matrix of size $(120)^2 \times 2600$. This matrix being tall and slim may not in general provide sparse representation to $\mathbf{x_q}$, the query image. Consequently, to bring CBMIR problem into the rich theory of compressed sensing, which is based on the under-determined setting, one needs to generate feature vectors of database members.

5.2. Feature Extraction

In the first feature extraction method (FE-I), each image is partitioned into 17 concentric circular regions such that each circular region has the same number of pixels as the other region. The mean and variance of these circular regions are used to design the feature vector. Hence, the size of each feature vector is 34×1 (due to 17 means and 17 variances) for each image. In the second feature extraction method (FE-II), image is partitioned into five sub-images and each sub-image is partitioned into 4 concentric circular regions, such that each circular region possesses same number of pixels as others. The mean and variance of pixel intensity in a circular region become components of the feature vector and size of each feature vector is 40×1 (due to 4 means and 4 variances from each of 5 sub images). This procedure is applied to all the database members and 14 more images are used for testing.

5.3. Performance Evaluation and Results

The performance of the medical image retrieval task is measured in terms of recall $R = N_c/N_m$ and precision $P = N_c \ / \ (N_c + N_f)$ where N_m is the total number of actual (or similar) images, N_c is the number of images detected correctly and N_f is the number of false alarms. A good performance requires both recall and precision to be high, that is, close to unity. Recall is the portion of total relevant images retrieved whereas precision indicates the capability to retrieve relevant images. A compromise between recall and precision is obtained by using a measure combining both as $F_1 = \frac{2\times(R\times P)}{R+P}$. Ideally, F_1 should be close to unity. Given some of retrieved images, the average retrieval performance is defined as the average number of relevant images retrieved over all query images of a particular class. We compare the performance of proposed method with that of CBMIR obtained by K-means and fuzzy C-means clustering algorithms on the same image database. We evaluate the experimental results on proposed, K-means and fuzzy C-means clustering algorithms using two different types of feature extraction methods on the same image database.

We evaluate the performance of the proposed method with three different cluster sizes, viz N=3, 4, 5. The precision and recall of dictionaries of different column sizes viz, 60, 65, 70, 75 80 and 80 with optimum cluster size are shown in Tables [I-VI]. Tables I, III and V receptively represent the average precision and recall for N being 3, 4 and 5 using the proposed method, fuzzy C-means and K-means clustering methods and using first feature extraction

[1]www.irma-project.org

method. Tables II, IV and VI receptively represent the average precision and recall for N being 3, 4 and 5 using the proposed method, fuzzy C-means and K-means clustering methods and using second feature extraction method. The performance in Table I was computed against the top 10 most accurately retrieved images for each test image using first feature extraction method and Euclidean distance as similarity measure. Through proposed method, for 3 clusters, the performance of 92.1% precision and 79.6% recall was obtained. Similarly, the results for 4 clusters provided the performance of 90.7% precision and 78.8% recall, and for 5 clusters the performance of 87.8% precision and 74.7% recall was obtained. The fuzzy C-means clustering algorithm with 3 clusters, resulted in the performance of 67.8% precision and 68.2% recall. In other cases, the performance of fuzzy C-means clustering algorithm was less accurate. The K-means clustering algorithm provided the performance of 62.1% precision and 32.6% recall. In other cases, the performance of K-means clustering algorithm was further less.

Table II shows the performance of evaluation obtained with the second feature extraction method and Euclidean distance as similarity measure. Through the the second method of feature extraction, the performance of 97.14% precision and 80.1% recall were obtained. Fuzzy C-means clustering algorithm provided the performance of 74.8% precision and 60% recall and K-means clustering algorithm resulted in the performance of 62.6% precision and 48% recall. From our simulation results, it may be concluded that the 2^{nd} feature extraction method gives better performance than the 1^{st} method.

Tables III and IV represent the average precision and recall of proposed, fuzzy C-means and K-means clustering methods using first and second feature extraction methods respectively, with cross correlation as similarity measure. It can be inferred from Tables III and IV that the proposed method using second feature extraction method (93.7% precision and 83.2% recall) provided better performance than the fuzzy C-means and K-means clustering algorithms.

Tables V and VI represent the average precision and recall of proposed, fuzzy C-means and K-means clustering methods using first and second feature extraction methods respectively by using Mahalanobis distance as similarity measure. From the results in Tables V and VI, it can be concluded that the proposed method performs better (62.8% precision and 47.2% recall) than fuzzy C-means and K-means clustering methods.

Tables VII and VIII represent the average precision and recall of proposed methods with increasing and decreasing number of concentric circular regions for the first feature extraction method by using Euclidean distance as similarity measure. The results obtained in the Table VII portray that decreasing the number of concentric regions (<17) for feature extraction yielded less performance. This is because of the reduction in feature vector size and the creation of non-optimal dictionaries for clustering. On the other hand, increasing the number of concentric circles beyond 17 yielded no significant improvement in the performance. Table IX depicts the results of the proposed method with various dictionary and cluster sizes. It can be noted that the dictionary size of 65 yielded the good performance.

In Fig. 3, on every row, the 1^{st} element represents the query image while the other represent those retrieved by the proposed method with Euclidean distance as similarity metric. Fig. 4 shows the average precision and recall of the proposed, fuzzy C-means and K-means clustering methods using first and second feature extraction methods with three different similarity measures. Among these, the proposed method has better performance (97.1% precision and 80.1% recall) with the Euclidean distance based similarity measure as shown in Fig. 4 (a) and Fig. 4 (d).

Figs. 5, 6, 7, 8, 9 and 10 show the average performance of F_1 measure over all query images of the proposed, fuzzy C-means and K-means clustering methods using first and second feature extraction methods with three different similarity measures. Among these, proposed method with second feature extraction method has better F_1 performance measure (87.5% when N=3) with the Euclidean distance based similarity measure as shown in Fig. 6(a).

Figs. 5 and 6 show the average performance of F_1 measure over all query images of the proposed method using the first and second feature extraction methods respectively with three different similarity measures. Among these configurations, the proposed method using first feature extraction method has F_1 measure of 85.1% when N=3 with Euclidean distance as similarity measure, which is shown in Fig. 5 (a). The second feature extraction method has F_1 measure of 87.5% when N=3 clusters with euclidean distance similarity measure as shown in Fig. 6(a).

Figs. 7 and 8 show the average performance of F_1 measure over all query images of the fuzzy C-means clustering method using first and second feature extraction methods respectively with three different similarity measures. With cross correlation as similarity metric, fuzzy C-means as clustering technique, FE1 and FE2 provided the performances 68.2% and 55.5% as F_1 measure, which is shown in Fig. 7(a) and Fig. 8(a).

Figs. 9 and 10 show the average performance of F_1 measure over all query images of the K-means clustering method. With cross correlation as similarity metric, fuzzy C-means as clustering technique, FE1 and FE2 provided the performances 72.5% and 68.1% as F_1 measure, which are shown in Fig. 9(a) and Fig. 10(a). Fig. 11 shows comparison between retrieval time and feature vector size for different cluster sizes. This plot indicates that increasing the feature vector size contributes to an increase in retrieval time as expected. Fig. 12 shows the cluster and its associated dictionary obtained after five iterations. Fig. 13 shows the features obtained from x-ray images of the hand and hip at various orientations 10, 20 and 30. It can be noted from the values of the features that the feature vectors of rotated classes of images are discriminative and are almost same for the rotated copies of an image. This implies automatically the rotation invariance of the proposed first feature extraction method.

5.4. Comparison with different features

Haralick [35] proposed the use of Gray-Level Co-occurrence Matrix (GLCM) and texture features for retrieval of medical images. Initially, GLCMs were computed and texture features were calculated based on the extracted GLCMs. However, the major drawback of this method is that it is computationally expensive and time-consuming [36]. The Table X shows the top 10 similar images retrieved based on similarity measure (Euclidean distance) using both Haralick features and the features proposed in the manuscript. It can be noted that the proposed features give best retrieval results of 90% as compared to texture features. Additionally, we implemented our Dictionary based CBMIR technique inputting Haralick as well as proposed features and the retrieval performances so obtained are shown in the XI Table. The performance of proposed method with circular features gives better retrieval results of 97.1% as compared to Haralick features. In Fig. 14, on every column, the first image represents query image while the remaining are those retrieved by the proposed and Haralick features with Euclidean distance as similarity metric. The Haralick feature and the proposed method perform equally well on the query image of a hand. However, for other query images, the proposed method gives better results as compared to Haralick features.

In SIFT method [5],[41], the key points so extracted mainly depend on local properties of representation of an object and hence cannot be expected to capture properties of solid objects. SIFT descriptors mainly depend on the properties of the projection of an object but not on the properties of the object itself. These details may be found in [41]. It may be observed from Fig. 15 below that SIFT extracts less number of key points on X-Ray medical images than on natural images. This is because X-ray images contain very less number of edge corners. Consequently, SIFT based feature extraction methods are unlikely to be suitable for X-ray medical images.

We implemented our Dictionary algorithm by considering Fisher [37], [38] and Vlad [39], [40] vectors as feature vectors. The performances obtained with Fisher vectors (21.4 % precision and 18.2 % recall) and Vlad (26.4 % precision and 23.1 % recall) are below par. The Fisher and Vlad features being derivatives of SIFT features result in not so discriminative features giving there by the poor retrieval performance.

As the result proposed in [3], [5], [6] and [33], to name a few, used either labeled data or concentrated on the medical data of specific context, comparison is not justified in true sense. Nevertheless, it may be emphasized that, despite having different objectives, our method gives competitive performance.

6. Conclusion and future work

In this paper, a novel dictionary learning based clustering method for content based medical image retrieval is proposed. Mean and variance of pixel intensity values are used as features and K-SVD method is used to generate dictionaries for each cluster. The performance of the proposed method is evaluated using IRMA database. The first feature extraction (FE1) method aims at providing rotation invariant CBIR, while the second (FE2) method aims at taking into consideration the rich information available at the center. Our experimental results show that FE2 method gives superior performance than FE1 on IRMA database. The extensive experimental work is carried out with different cluster sizes, different number of concentric circular regions, different column sizes for dictionaries, different similarity metrics and with different initial clustering algorithms. Our observation is that when N is 3 (that is, number of concentric circles is 17), one achieves better F_1 performance of 87.5% with Euclidean distance as similarity metric.

The supervised dictionary learning (DL) exploits image/region labels, match/non-match features for discriminative sparse coding. The accurate extraction of attributes (such as feather, car, tree etc) from database members is key to the effectiveness of DL based classification. As accurate attribute extraction is not guaranteed always, Yue Gao et. al. [20] proposed a novel weakly unsupervised DL method. They incorporated cheaply available visual attributes very effectively into dictionary learning and demonstrated the efficacy, scalability of the method in large scale image retrieval and classification tasks. In fact, adding supervised information to facilitate image retrieval is an interesting, challenging and ongoing research. It is even more challenging in dealing with medical databases. This is because it is very difficult to define suitable attributes from medical databases, which is in contrast to the databases of general images where attributes such as tree, car and feather etc can work well. Inspired by the ideas in [20], our future efforts shall attempt to identify and incorporate visual attributes into DL to improve upon the performance of CBMIR. In addition, as medical images come with different transformations (such as scaling), our future work shall further aim at addressing the invariant CBMIR with respect to other transformations as well.

Acknowledgement

Authors would like to thank Dr. T.M. Deserno, Department of Medical Informatics, RWTH Aachen, Germany for making the original IRMA Database available for research purposes. One of the authors (CSS) is thankful to the CSIR for its support (25(0219)/13/EMR-II). Authors are grateful to the anonymous reviewers for their thorough review and constructive comments, which improved the manuscript substantially. Finally, authors thank Prof. Andrea Vedaldi for making the software of Vlad and Fisher vectors available.

References

[1] A.W.M. Smeulder, M. Worring, S. Santini, A. Gupta, R. Jain, “Content based image retrieval at the end of the early years,” IEEE Trans. Pattern Analysis Machine Intell., vo1. 22, no. 12, pp. 1349-1380, 2000.

[2] H. Pourghassem, H. Ghassemian, “Content based medical image classification using a new hierarchical merging scheme,” Computerized Medical Imaging and Graphics., vol. 32, no. 8, pp. 651-661, 2008.

[3] T. M. Lehmann, Guld. Mark. O, ‘Deselaers. Thomas, Keysers. Daniel, Schubert. Henning, Spitzer. Klaus, Ney. Hermann, B. B. Wein, “Automatic categorization of medical images for content-based retrieval and data mining,” Computerized Medical Imaging and Graphics., vol.9, no. 2, pp. 143-155, 2005.

[4] H. D. Tagare, C. Jafe, J. Duncan, “Medical Image Databases: A Content Based Retrieval Approach,” JAmer. Medical. Informat.Assoc, vol. 4, no.3, pp. 184-198, 1997.

[5] W. Yang, Z. Lu, Yu. M, M. Huang, Q. Feng, W. Chen, “Content-based retrieval of focal liver lesions using Bag-of-Visual-Words representations of single- and multiphase contrast enhanced CT images,” Journal of Digital Imaging, vol. 25, pp. 708-719, 2011.

[6] Depeursinge. Adrien, Foncubierta. Rodriguez. Antonio, Van. de. Ville. Dimitri, Muller. Henning, “Multiscale lung texture signature learning using the Riesz transform,” Medical Image Computing and Computer-Assisted Intervention-MICCAI 2012, pp.517-524, 2004.

[7] G. Csurka, S. Clinchant, G. Jacquet,“Medical Image Modality Classification and Retrieval,” in Proc. IEEE Int. Conf. CBMI pp.193-198, 2011.

[8] M. Aharon, M. Elad, A. M. Bruckstein,“The K-SVD: An Algorithm for Designing of Overcomplete Dictionaries for Sparse Representation,” IEEE Trans. Signal Process., vol. 54 no. 11, pp. 4311-4322, 2006.

[9] Tian. Tian Chang, Jun. Feng, Hong. Wei. Liu and Horace. H. S. I,“Clustered Microcalcification Detection Based on a Multiple Kernel Support Vector Machine with Grouped Features (GF-SVM),” Proc. IEEE Conf. Pattern Recognition (ICPR), pp.1-4, 2008.

[10] Guanqun. Bao, Kaveh. Pahlavan, “Motion Estimation of the Endoscopy Capsule using Region-based Kernel SVM Classifier,” in Proc. IEEE Int. Conf. Electro/Information Technology (EIT), pp. 1-5, 2013.

[11] Amjad. Zaim, Taeil. Yi, and Rick. Keck, “Feature-Based Classification of Prostate Ultrasound Images using Multiwavelet and Kernel Support Vector Machines ,” in Proc. IEEE Int. Conf. Neural Networks (IJCNN 2007), pp. 278 - 281, 2007.

[12] I. Ramirez, P. Sprechmann, Guillermo. Spiro,“Classification and Clustering via Dictionary Learning with Structured Incoherence and Shared Features,” in Proc. IEEE Conf. Computer Vision and Pattern Recognition, June 2010,

[13] K. Kreutz-Delgado, J. F. Murray, B. D. Rao, K. Engan, T. W. Lee, and T. Sejnowski, “Dictionary Learning Algorithms for Sparse Representation, Neural Computation, no 15, pp. 349396, 2003.

[14] K. Engan, S. Aase, J. Hakon-Husoy, “Method of optimal directions for frame design,” in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process., vol. 5, pp. 2443-2446, 1999.

[15] S. Mallat, Z. Zhang,“Matching Pursuits with Time Frequency Dictionaries,” IEEE Trans. Signal Processing., Vol. 41, no.12, pp. 3397-3415, 1993.

[16] W. W. Chu, C. C. Hsu, A. F. Cardenas, R. K. Taira, “Knowledge Based Image Retrieval with Spatial and Temporal Constructs,” IEEE Trans. Knowledge. Data Eng., vol. 10, no. 6, pp. 872-888, Dec. 1998.

[17] W. Cai, D. Feng, R. Fulton,“Content-Based Retrieval of Dynamic PET Functional Images,” IEEE Trans. Information Technology in Biomedicine., vol. 4, no. 2, pp. 152-158, Jun. 2000.

[18] C. Shyu, C. Brodley, A. Kak, A. Kosaka, A. Aisen, L. Broderick,“ASSERT: A physician-in-loop Content-Based Image Retrieval System for HRCT Image Databases,” Computer Vision and Image Understanding., vol. 75, pp. 111-132, 1999.

[19] A. Guimond, G. Subsol,“Automatic MRI Database Exploration and Applications,” Int. Journal. Pattern Recognit. Artif. Intell., vol. 11, pp. 1345-1365, 1997.

[20] Yue Gao, Rongrong Ji , Wei Liu, Qionghai Dai and Gang Hua, “ Weakly Supervised Visual Dictionary Learning by Harnessing Image Attributes,” IEEE Trans. on Image Processing, Vol. 23, No. 12, 5400-5411, 2014.

[21] S. Li, H. Yin, L. Fang,“Group Sparse Representation with Dictionary Learning for Medical Image Denoising and Fusion,” IEEE Tans. Biomedical Engineering, vol. 59, no. 12, Dec 2012.

[22] G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. Roux, “Wavelet optimization for content-based image retrieval in medical databases,” Journal. Medical Image Analysis., vol 14, pp. 227-241, 2010.

[23] K. Etemand, R. Chellapa,"Separability Based Multiscale Basis Selection and Feature Extraction for Signal and Image Classification," IEEE Trans. Image Proc. vol. 7, no. 10, pp. 1453-1465, Oct. 1998.
[24] F. Rodriguez, G. Sapiro, "Sparse Representation for Image Classification: Learning Discriminative and Reconstructive non Parametric Dictionaries," Tech. Report, University of Minnesota, Dec. 2007.
[25] K. Huang, S. Aviyente, "Sparse representation for signal classification," NIPS, vol. 19, pp. 609-616, 2007.
[26] Q. Zhang, B. Li, "Discriminative k-svd for dictionary learning in face recognition," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 2691-2698, 2010.
[27] J. Mairal, F. Bach, J. Pnce, G. Sapiro, A. Zisserman, "Discriminative learned dictionaries for local image analysis," Proc. IEEE Conf. Computer Vision and Pattern Recognition, Anchorage, AL, June 2008
[28] http://lear.inrialpes.fr/people/jegou/data.php
[29] M. Ranzato, F. Hung, Y. Boureau, Y. LeCun,"Unsupervised Learning of Invariant Feature Hierarchies with Applications to Object Recognition," Proc. IEEE Conf. Computer Vision and Pattern Recognition, pp. 1-8, 2007.
[30] Y. C. Chen, C. S. Sastry, V. M. Patel, P. J. Phillips, R. Chellappa. ,"In-Plane Rotation and Scale Invariant Clustering Using Dictionaries," IEEE Tans. Image Processing, vol. 22, No. 6, pp.2166-2180, June 2013.
[31] F. Yu, Y.Chen, L. Luo,"CT Image Denoising Based on Sparse Representation Using Global Dictionary," Proc. IEEE Conf. Complex Medical Engineering, pp. 408-411, 2013.
[32] Akgul, Ceyhun. Burak, et al., "Content-based image retrieval in radiology: current status and future directions," Journal of Digital Imaging, vol.24 no. 2, pp. 208-222, 2011.
[33] Dettori. Lucia, Lindsay. Semler, "A comparison of wavelet, ridgelet, and curvelet-based texture classification algorithms in computed tomography," Computers in Biology and Medicine, vol. 37, no. 4, pp. 486-498, 2007 .
[34] S. Hyukcha,"Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions," Proc. of Int. Journal of Mathematical Models and Methods in Applied Sciences, vol.1, pp. 300-307, 2007.
[35] K. Shanmugam, R. M. Haralick, I. H. Dinstein, "Textural features for image classification," IEEE Transactions on Systems, Man and Cybernetics, vol.3 , pp. 610-621, 1973.
[36] Tahir. MA, Bouridane. A, Kurugollu. F, "An FPGA based coprocessor for GLCM and Haralick texture features and their application in prostate cancer classification," Proc. Analog Integrated Circuits Signal Process Vol.43, No. 2, pp.205215, 2005.
[37] F. Perronnin, C. Dance. "Fisher kernels on visual vocabularies for image categorization," Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR '07), pp. 1-8, 2007.
[38] Florent. Perronnin, Jorge, Snchez, Thomas. Mensink, "Improving the fisher kernel for large-scale image classification," In Proc. ECCV, pp. 143-156, 2010.
[39] H. Jegou, M. Douze, C. Schmid, P. Perez, "Aggregating local descriptors into a compact image representation," In Proc. CVPR, pp. 1-8, 2010.
[40] R. Arandjelovic, A. Zisserman, "All about VLAD." In Proc. CVPR, pp. 1-8, 2013.
[41] T. Bakken, "An Evaluation of the SIFT algorithm for CBIR," Telnor R&I Research Note N30/2007.

Table I Performance measures (%) of the Proposed, Fuzzy *C*-means and *K*-means clustering methods obtained with the first feature extraction method and the Euclidean distance as similarity measure.

Query	Proposed Method-I			Fuzzy C-Means-I			K-Means Clustering-I		
Images/Clusters	N=3	N=4	N=5	N=3	N=4	N=5	N=3	N=4	N=5
1.png	100	90	100	0	0	50	50	0	45
2.png	80	90	100	10	0	40	70	25	10
3.png	70	70	100	100	40	60	60	20	30
4.png	100	100	100	100	50	60	80	100	55
5.png	100	100	80	90	50	50	60	85	45
6.png	100	100	100	80	60	40	80	100	100
7.png	90	100	80	50	100	20	50	0	10
8.png	100	80	100	100	50	0	50	0	60
9.png	70	80	80	80	40	60	70	20	30
10.png	100	100	80	90	50	50	70	80	50
11.png	100	90	100	20	10	50	50	0	50
12.png	90	100	60	90	50	40	70	90	80
13.png	100	80	80	90	40	0	50	0	10
14.png	90	90	90	50	60	20	60	10	50
precision (%)	**92.1**	90.7	87.8	**67.8**	42.8	38.5	**62.1**	37.8	41.4
recall(%)	**79.6**	78.8	74.7	**68.2**	62	60.7	**32.6**	35.5	60.2

Table II Performance measures (%) of the Proposed, Fuzzy *C*-means and *K*-means clustering methods using second feature extraction method and Euclidean distance as similarity measure.

Query	Proposed Method-II			Fuzzy C-Means-II			K-Means Clustering-II		
Images/Clusters	N=3	N=4	N=5	N=3	N=4	N=5	N=3	N=4	N=5
1.png	100	100	100	70	60	60	30	30	0
2.png	100	100	90	80	70	60	40	10	20
3.png	90	70	100	80	70	50	57.1	50	50
4.png	100	80	100	90	50	40	50	40	50
5.png	100	100	90	90	70	20	70	40	100
6.png	90	90	80	60	40	80	80	20	70
7.png	100	100	100	50	70	70	100	100	50
8.png	100	100	80	25	0	0	80	40	0
9.png	90	88.8	90	85.7	83.3	87.5	100	90	43
10.png	100	90	90	90	90	80	90	60	32
11.png	100	77.7	60	83.3	50	16.6	100	60	0
12.png	100	100	90	77.7	80	80	80	40	50
13.png	90	90	40	83.3	60	100	0	40	30
14.png	100	100	100	83.3	20	0	0	40	20
precision(%)	**97.14**	91.8	86.4	**74.8**	54	48	**62.6**	45	43
recall(%)	**80.1**	83.2	76.9	**60**	58.2	68	**48**	38	32

Figure 3: On each row, the first one is query image while the remaining are those retrieved by the proposed dictionary base CBMIR.

Table III Performance measures (%) of the Proposed, Fuzzy *C*-means and *K*-means clustering methods using first feature extraction method and cross correlation as similarity measure.

Query	Proposed Method-I			Fuzzy C-Means-I			K-Means Clustering-I		
Images/Clusters	N=3	N=4	N=5	N=3	N=4	N=5	N=3	N=4	N=5
1.png	90	60	50	20	30	60	68.4	0	60
2.png	70	70	50	20	30	50	47.3	0	44.4
3.png	100	90	80	100	80	80	65	35	0
4.png	90	60	80	100	50	90	33.3	55	50
5.png	70	80	90	100	50	50	33.3	50	50
6.png	100	100	100	80	60	60	91.6	25	55
7.png	90	100	100	100	100	30	100	25	0
8.png	40	30	20	50	30	0	37.5	0	0
9.png	90	80	80	90	70	70	60	30	0
10.png	80	90	80	100	50	50	40	50	50
11.png	80	60	50	20	30	50	63.2	0	70
12.png	100	80	60	70	50	50	80	30	50
13.png	60	40	20	40	30	10	32.4	10	0
14.png	50	50	25	50	20	0	30	0	20
precision(%)	**79.2**	70.7	68.5	**67.1**	48.5	46.4	**55.8**	22.1	32.1
recall(%)	**65.7**	65	68	**69.8**	75	71.4	**57.4**	47.4	47

Table IV Performance measures (%) of the Proposed, Fuzzy *C*-means and *K*-means clustering methods using second feature extraction method and cross correlation as similarity measure.

Query	Proposed Method-II			Fuzzy C-Means-II			K-Means Clustering-II		
Images/Clusters	N=3	N=4	N=5	N=3	N=4	N=5	N=3	N=4	N=5
1.png	90	100	90	70	90	70	70	70	40
2.png	100	100	90	80	90	80	80	60	70
3.png	97	90	94.1	60	40	50	57.1	50	67
4.png	90	90	100	50	50	30	90	67	63
5.png	100	100	50	50	90	30	90	0	70
6.png	90	90	100	90	50	100	90	78	90
7.png	100	80	80	100	50	50	0	50	0
8.png	95	80	100	100	0	20	83.3	50	0
9.png	100	80	100	87.5	40	50	50	10	90
10.png	100	80	100	90	50	40	70	90	80
11.png	90	60	80	83.3	90	70	70	70	70
12.png	90	100	100	88.9	50	90	100	70	50
13.png	80	70	20	83.3	10	30	50	40	50
14.png	90	100	100	83.3	20	20	60	40	50
precision(%)	**93.7**	87.1	86	**79.6**	51.4	52.1	**68.6**	53.2	56.4
recall(%)	**83.2**	80.1	76.8	**69.6**	50.1	64.8	**71.9**	50.4	49.8

Table V Performance measures (%) of the Proposed, Fuzzy C-means and K-Means clustering methods using first feature extraction method and Mahalanobis distance as similarity measure.

Query	Proposed Method-I			Fuzzy C-Means-I			K-Means Clustering -I		
Images/Clusters	N=3	N=4	N=5	N=3	N=4	N=5	N=3	N=4	N=5
1.png	60	70	100	0	0	50	31.5	0	35
2.png	40	30	60	0	0	33.3	26.5	0	05
3.png	40	50	50	60	20	28.5	30	10	05
4.png	30	40	60	70	50	50	18	50	45
5.png	30	70	30	70	50	50	18	55	45
6.png	50	70	50	20	20	0	16.6	05	30
7.png	90	70	50	50	80	0	52.6	03	0
8.png	100	90	90	70	30	0	75	0	30
9.png	50	50	50	60	20	30	30	10	0
10.png	30	40	50	70	60	50	20	50	30
11.png	60	60	90	0	10	60	30	0	30
12.png	50	70	40	40	30	0	20	30	40
13.png	90	80	80	30	20	10	70	20	40
14.png	80	90	70	40	30	10	40	0	45
precision(%)	57.1	60.7	**62.1**	**41.4**	30	26.5	**34.1**	19.2	27
recall(%)	56.4	54.6	**60**	**42.9**	49.3	32	**41.5**	30.7	39.5

Table VI Performance measures (%) of the Proposed, Fuzzy C-means and K-Means clustering method using second feature extraction method and Mahalanobis distance as similarity measure.

Query	Proposed Method-II			Fuzzy C-Means-II			K-Means Clustering-II		
Images/Clusters	N=3	N=4	N=5	N=3	N=4	N=5	N=3	N=4	N=5
1.png	40	70	70	0	30	30	30	30	0
2.png	30	60	60	50	20	20	20	10	30
3.png	100	40	80	29	43	38	20	19	35.2
4.png	40	30	50	38	20	33.3	33.3	17	62.2
5.png	40	60	30	50	20	0	33.3	20	62.2
6.png	50	60	70	50	40	40	60	20	59
7.png	80	83.3	70	71.4	40	75	100	0	0
8.png	90	90	50	88	20	0	67	0	0
9.png	10	20	80	14.2	16.6	50	10	20	0
10.png	30	20	50	10	10	0	30	20	50
11.png	60	30	30	50	30	30	30	30	30
12.png	30	30	60	33.3	20	40	70	20	20
13.png	50	50	90	66.6	0	0	30	10	20
14.png	50	80	90	66.6	16.6	0	30	10	60
precision(%)	50	53.8	**62.8**	**44**	23.3	25.4	**40.2**	16.1	30.6
recall(%)	49.3	49.4	**47.2**	**38.7**	29.1	27	**55.6**	24.7	31.6

Table VII Performance measures (%) of the proposed method with decreasing feature vector size (No.of concentric circles is 7) using Euclidean distance, Cross correlation and Mahalanobis distance as similarity measure.

Query	Euclidean distance			Cross correlation			Mahalanobis distance		
Images/Clusters	N=3	N=4	N=5	N=3	N=4	N=5	N=3	N=4	N=5
1.png	80	80	30	50	30	30	30	60	10
2.png	100	90	30	90	40	30	30	30	20
3.png	90	90	90	100	90	90	70	50	50
4.png	100	100	70	90	90	80	30	30	50
5.png	100	90	90	100	100	80	50	50	20
6.png	40	90	70	70	100	90	20	70	30
7.png	100	80	80	100	90	60	40	20	20
8.png	90	100	70	40	20	10	70	90	70
9.png	100	100	100	100	100	90	10	30	30
10.png	60	100	80	80	90	100	10	70	60
11.png	90	90	90	70	40	80	10	60	40
12.png	30	50	60	40	50	70	30	30	20
13.png	60	100	80	20	70	40	30	50	50
14.png	90	100	70	60	60	10	90	90	80
precision(%)	80.7	**90**	72.1	**72.1**	69.2	61.4	37.1	**52.1**	39.2
recall(%)	67.9	**80.4**	54.9	**66.3**	61.8	50.2	35	**53.4**	36.2

Table VIII Performance measure (%) of the proposed method with increasing feature vector size (No.of concentric circles=23) using Euclidean distance, cross correlation and Mahalanobis distance as similarity measure.

Query	Euclidean distance			Cross correlation			Mahalanobis distance		
Images/Clusters	N=3	N=4	N=5	N=3	N=4	N=5	N=3	N=4	N=5
1.png	100	90	80	90	70	60	50	40	20
2.png	70	80	90	80	100	80	40	20	30
3.png	90	90	70	100	100	100	60	40	50
4.png	80	100	100	90	100	90	40	30	20
5.png	100	100	90	100	100	100	30	30	40
6.png	90	90	70	100	90	90	20	50	30
7.png	100	90	90	100	100	100	60	50	40
8.png	100	90	100	20	20	10	60	70	80
9.png	100	100	90	100	100	100	40	20	50
10.png	70	70	70	80	70	100	50	40	30
11.png	80	90	80	80	100	80	20	20	40
12.png	90	90	80	90	90	80	40	70	60
13.png	100	80	100	80	70	90	60	60	80
14.png	100	90	90	60	40	30	90	80	90
precision(%)	**91.4**	89.2	85.7	**83.5**	82	79.2	**47.8**	44.2	47.1
recall(%)	**71.6**	76.3	73.8	**68.9**	71.9	68.7	**47.4**	45.5	43.3

Table IX Performance measure (%) of the proposed method with different dictionary sizes

	Proposed method-I			Proposed method-II		
Column size of D_i/Clusters	N=3	N=4	N=5	N=3	N=4	N=5
60	89	82.3	82	93	91	86.4
65	**92.1**	90	**87.8**	**97.1**	**91.8**	**93**
70	86.2	**90.7**	82	91.2	89.3	89
75	88.1	82	80.2	92.3	90	90
80	86.4	80.4	86	93.2	88.6	91.3
85	88.3	78.4	84	95	89.1	88.6

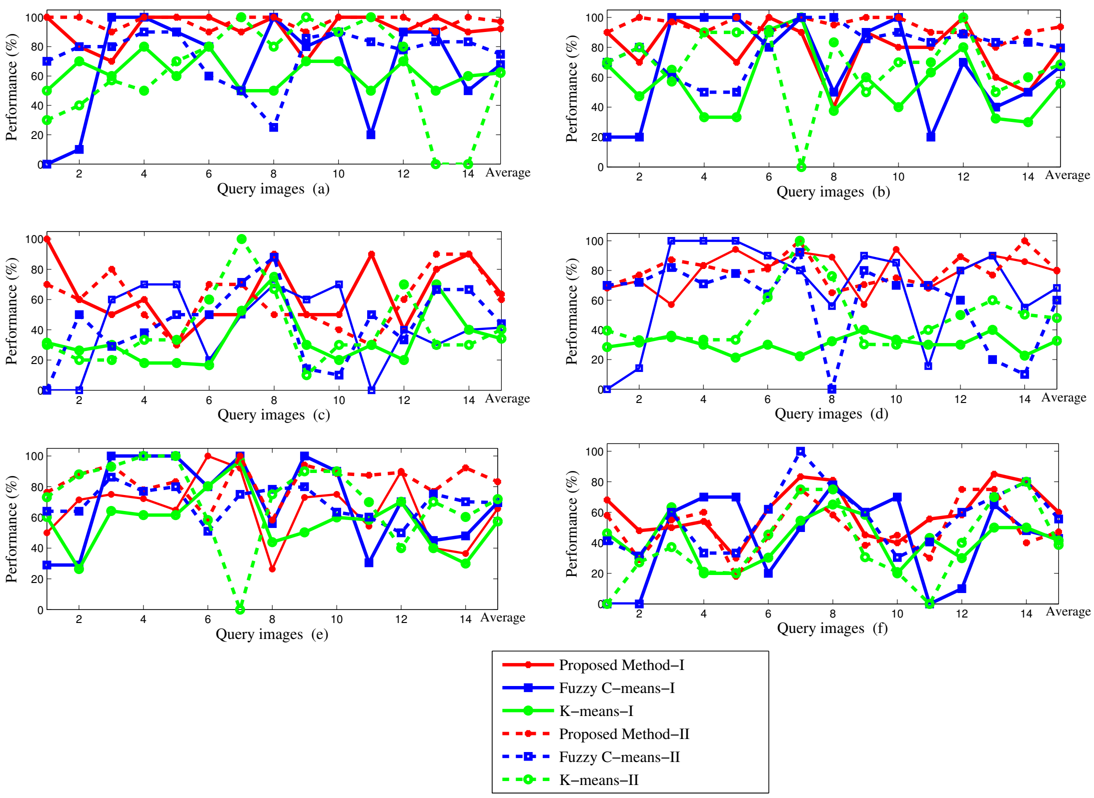


Figure 4: Comparing average precision and recall of proposed, fuzzy C-means and K-means clustering methods using first (I) and second (II) feature extraction methods with three different distance similarity measures. (a) Best performance of precision (%) using euclidean distance as similarity measure: (b) Best performance of precision (%) using cross correlation as similarity measure. (c) Best performance of precision (%) using Mahalanobis distance as similarity measure. (d) Best performance of recall (%) using euclidean distance as similarity measure. (e) Best performance of recall (%) using cross correlation as similarity measure, (f) Best performance of recall (%) using Mahalanobis distance as similarity measure. Here, x-axis refers to different query images and the y-axis refers to F_1 performance.

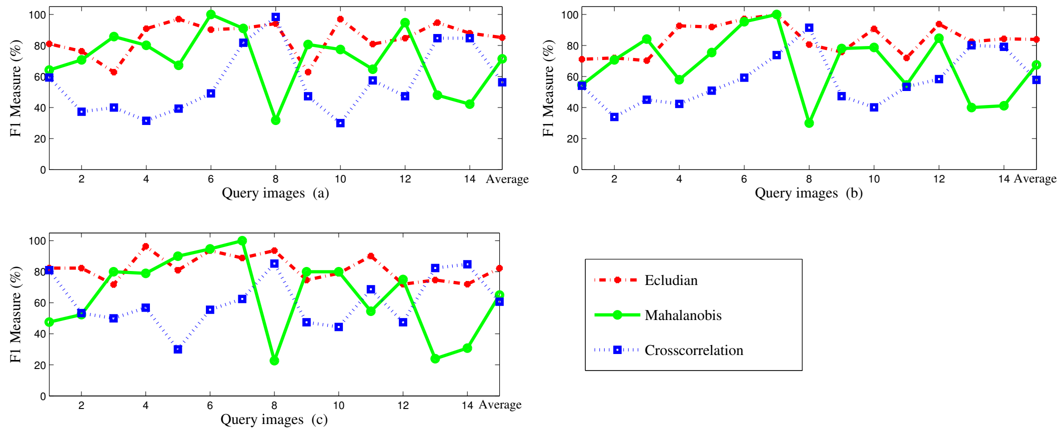


Figure 5: Comparing F_1 measure of proposed method (PM-I) using first feature extraction method with three different distance similarity measures: (a) Best F_1 measure of (%) proposed method (PM-I) when N=3 Clusters, (b) Best F_1 measure of (%) proposed method (PM-I) when N=4 Clusters, (c) Best F_1 measure of (%) proposed method (PM-I) when N=5 Clusters. The x and y axes in these plots refer respectively to different query images and the associated F_1 performances.

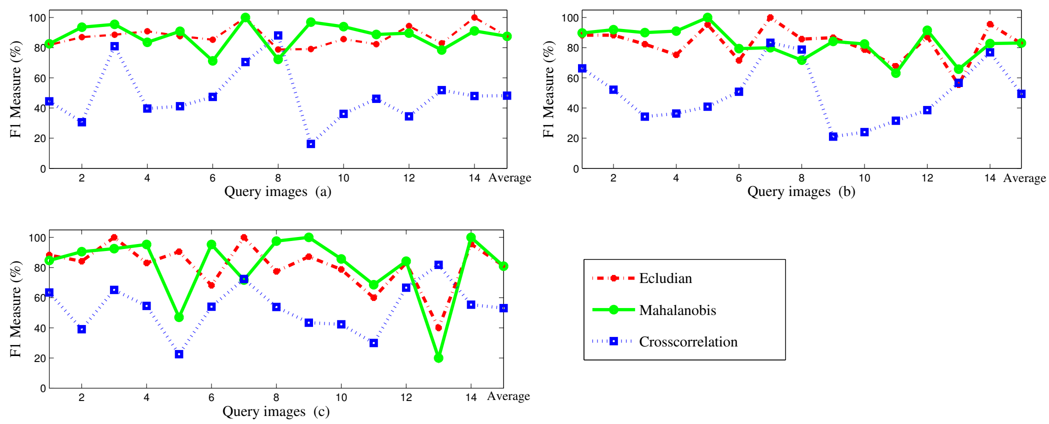


Figure 6: Comparing F_1 measure of proposed method (PM-II) using second feature extraction method with three different distance similarity measures: (a) Best F_1 measure of (%) proposed method (PM-II) when N=3 Clusters, (b) Best F_1 measure of (%) proposed method (PM-II) when N=4 Clusters, (c) Best F_1 measure of (%) proposed method (PM-II) when N=5 Clusters.

Table X: Retrieval performance (%) of the texture and proposed features with euclidean distance as similarity measure (that is feature extraction step followed by similarity measure without dictionary concept)

Query Image	**Haralick features**	**Proposed features**	**Query Image**	**Haralick features**	**Proposed features**
1.png	100	90	**8.png**	40	100
2.png	100	90	**9.png**	50	90
3.png	70	70	**10.png**	70	80
4.png	40	100	**11.png**	100	90
5.png	60	100	**12.png**	60	70
6.png	40	90	**13.png**	50	90
7.png	80	100	**14.png**	20	100
Average				**63**	**90**

Table XI: Retrieval performance (%) of proposed approach using texture features with euclidean distance as similarity measure (that is feature extraction step followed by dictionary based retrieval).

Query Images	**Number of Clusters**					
	N=3		**N=4**		**N=5**	
	precision	**recall**	**precision**	**recall**	**precision**	**recall**
1.png	90	62.5	30	30	55	55
2.png	28.5	28.5	60	48.9	70	58.3
3.png	60	55.2	70	63.6	44.1	48
4.png	50	44.4	68.4	60	60	50
5.png	50	33.3	41.3	48.6	52.5	59.3
6.png	30	34.1	70	40.7	45	36.8
7.png	75	62.8	80	65.3	73.3	73.3
8.png	38.8	62.5	38	48.2	60	50
9.png	80	80	44.4	44.4	66.6	66.6
10.png	60	59.2	40	40.9	60	66.6
11.png	20	28.5	61.1	45.9	50	50
12.png	40	22.4	20	28.5	30	21.4
13.png	40	17.6	50	50	20	15
14.png	30	19.4	25	25	23.5	20
Average	**51**	**31**	**48.5**	**45.8**	**48.7**	**45**

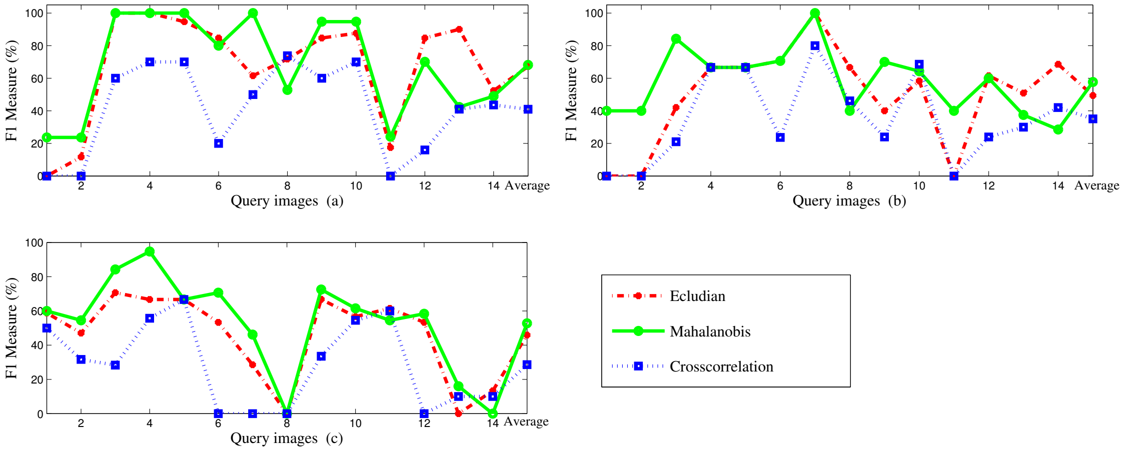


Figure 7: Comparing F_1 measure of fuzzy C-means (FC-I) using first feature extraction method with three different distance similarity measures: (a) Best F_1 measure of (%) fuzzy C-means (FC-I) when N=3 Clusters, (b) Best F_1 measure of (%) fuzzy C-means (FC-I) when N=4 Clusters, (c) Best F_1 measure of (%) fuzzy C-means (FC-I) when N=5 Clusters.

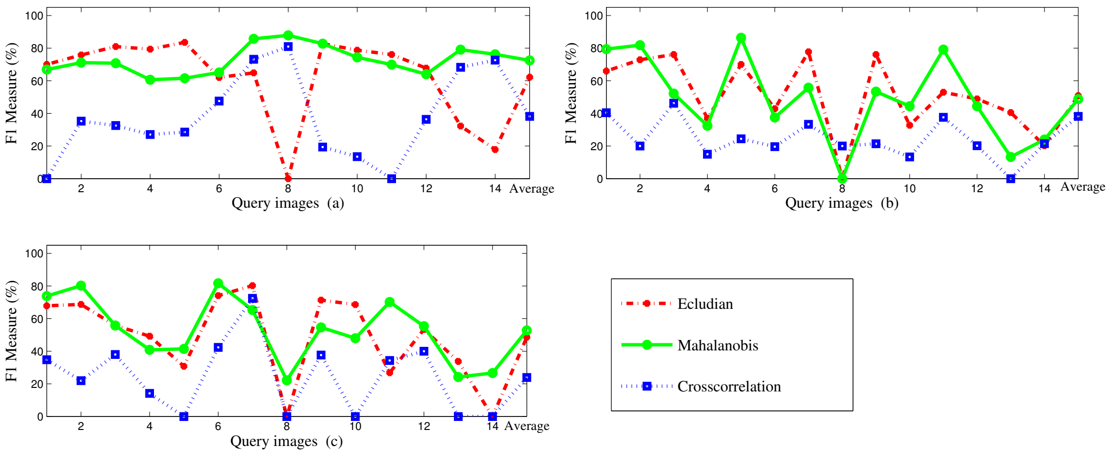


Figure 8: Comparing F_1 measure of fuzzy C-means (FC-II) using second feature extraction method with three different distance similarity measures: (a) Best F_1 measure of (%) fuzzy C-means (FC-II) when N=3 Clusters, (b) Best F_1 measure of (%) fuzzy C-means (FC-II) when N=4 Clusters, (c) Best F_1 measure of (%) fuzzy C-means (FC-II) when N=5 Clusters.

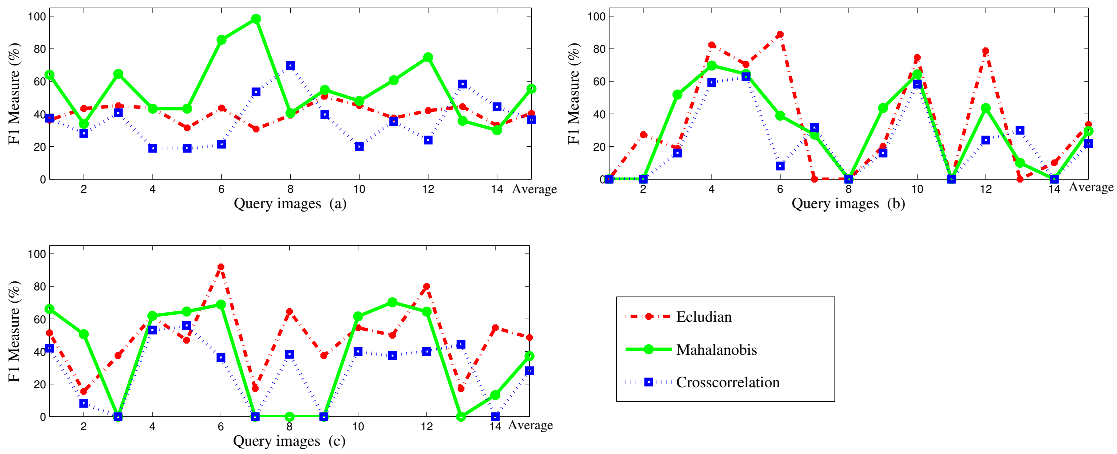


Figure 9: Comparing F_1 measure of K-means (KM-I) using first feature extraction method with three different distance similarity measures: (a) Best F_1 measure of (%) K-means when N=3 Clusters, (b) Best F_1 measure of (%) K-means when N=4 Clusters, (c) Best F_1 measure of (%) K-means when N=5 Clusters.

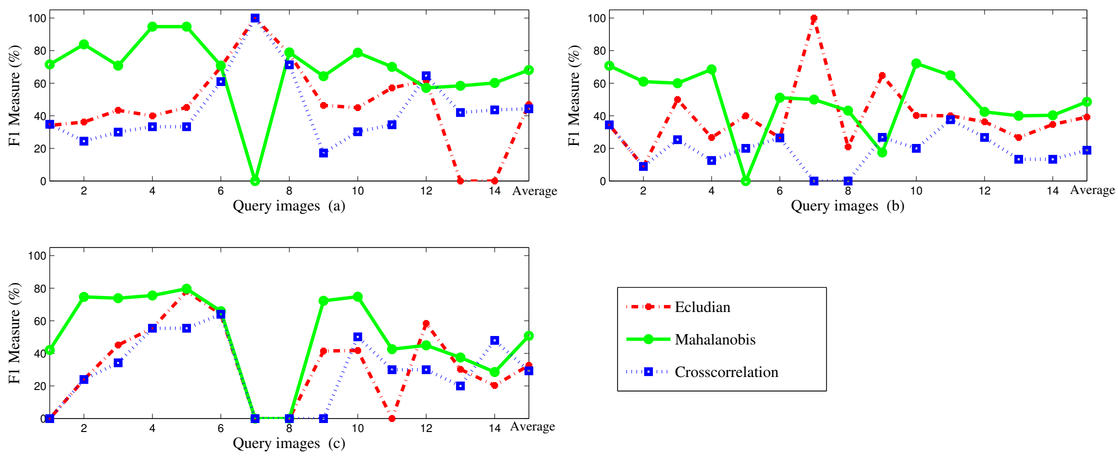


Figure 10: Comparing F_1 measure of K-means (KM-II) using second feature extraction method with three different distance similarity measures: (a) Best F_1 measure of (%) K-means when N=3 Clusters, (b) Best F_1 measure of (%) K-means when N=4 Clusters, (c) Best F_1 measure of (%) K-means when N=5 Clusters.

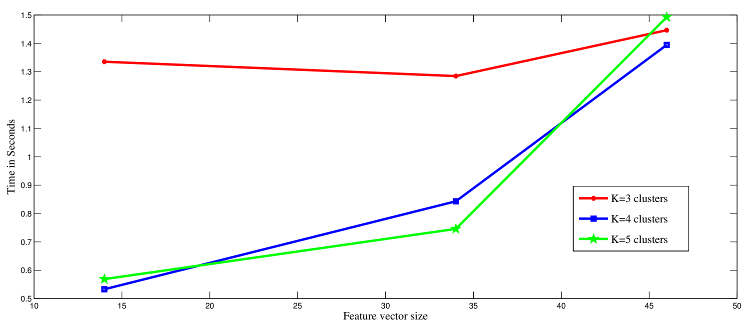
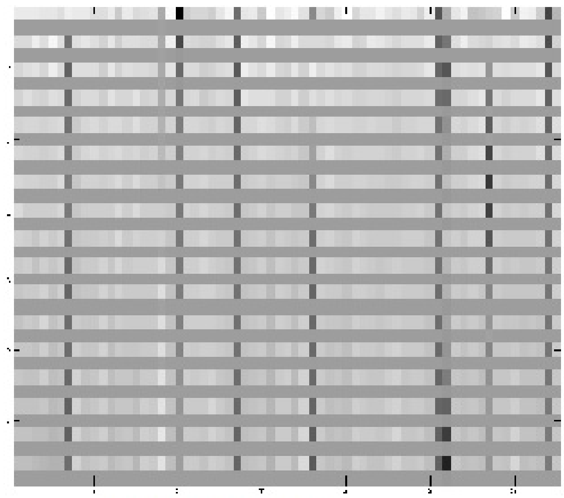


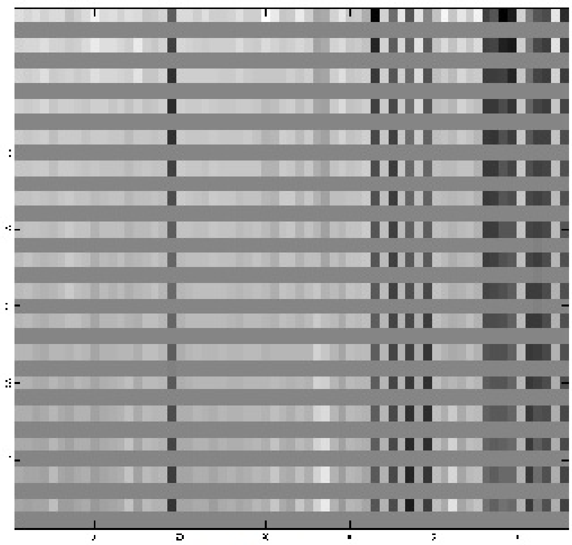
Figure 11: Comparison between retrieval time and feature vector size for different cluster sizes.

(a) Intial Cluster images



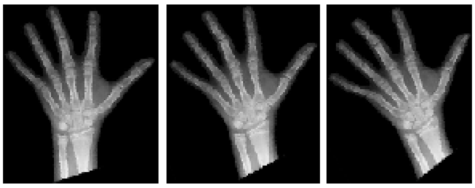
(a1) Intial Dictionary

(b) Final Cluster images

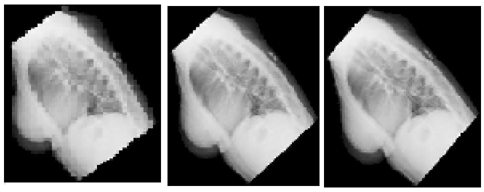


(b1) Final Dictionary

Figure 12: Clustered images and corresponding dictionary. (a) is initial cluster and (a1) is corresponding initial dictionary. (b) is final cluster and (b1) is corresponding final dictionary.



(a) Hand Images



(b) Hip Images

10	20	30
149.231121	145.574371	145.535469
0.03203661	0.03661327	0.03457666
134.747768	134.022321	133.272321
0.02092857	0.02678571	0.02339286
126.951835	124.603211	128.522936
0.01376147	0.01587156	0.01376147
136.006757	136.171171	136.963964
0.02702703	0.02009009	0.02252252
122.451327	122.575221	120.471239
0.02212389	0.02964602	0.02884956
125.135135	125.207207	124.092342
0.1036036	0.11351351	0.09459459
118.027778	116.111111	118.25463
0.00462963	0.00462963	0.04018519
100.15367	100.295872	100.912844
0.08256881	0.08834862	0.88018349
96.3640351	96.2412281	94.7149123
0.14719298	0.14333333	0.14473684
86.0550459	85.2591743	82.3027523
0.17431193	0.17889908	0.18100917
78.25	76.9391892	78.045045
0.12162162	0.11261261	0.12477477
62.5069444	68.0763889	66.7060185
0.25462963	0.29907407	0.28703704
62.8662281	62.9276316	61.247807
0.25964912	0.24508772	0.24561404
53.7991453	54.0128205	53.2393162
0.4017094	0.44017094	0.43333333
51.8725	51.2475	52.9025
0.515	0.43	0.59
47.0811966	48.3119658	49.982906
0.44871795	0.42735043	0.42478632
44.9041667	46.5805556	44.3152778
0.48888889	0.49444444	0.49444444

(c) Hand Feature vectors

10	20	30
165.029018	166.85177	166.827434
0.19642857	0.1919469	0.19159292
171.898148	171.824201	174.022831
0.06481481	0.16415525	0.18721461
170.945652	170.490909	176.863636
0.06956522	0.03636364	0.06363636
166.643182	169.196833	166.497738
0.02272727	0.02345233	0.02239819
164.752315	167.80543	169.572398
0.07777778	0.09502262	0.07692308
162.136161	165.463636	167.093182
0.05392857	0.00454545	0.05454545
162.03271	164.162844	164.993119
0.20560748	0.03669725	0.04587156
162.464912	164.638249	163.133641
0.02631579	0.03225806	0.03354839
163.392202	165.173423	163.673423
0.02293578	0.08558559	0.06756757
164.584091	164.325221	165.809735
0.03636364	0.01327434	0.04867257
165.441441	166.073059	167.977169
0.05702703	0.05022831	0.06392694
165.975225	166.140625	167.939732
0.02702703	0.03375	0.02160714
165.90708	166.050926	167.215278
0.10079646	0.10648148	0.10648148
167.138095	165.878661	166.987448
0.0847619	0.07154812	0.08786611
169.057377	165.866667	166.942857
0.05327869	0.05904762	0.05238095
170.186893	166.316143	169.459641
0.15339806	0.18834081	0.11210762
171.102703	168.226415	172.411051
0.21081081	0.17773585	0.17789757

(d) Hip Feature vector

Figure 13: (a) Rotated copies of hand images. (b) Rotated copies of Hip images. (c) Feature vectors of rotated hand images (d) Feature vectors of rotated hip images. These feature vectors indicate that the features generated are discriminative and rotation independent approximately in the discrete setting.

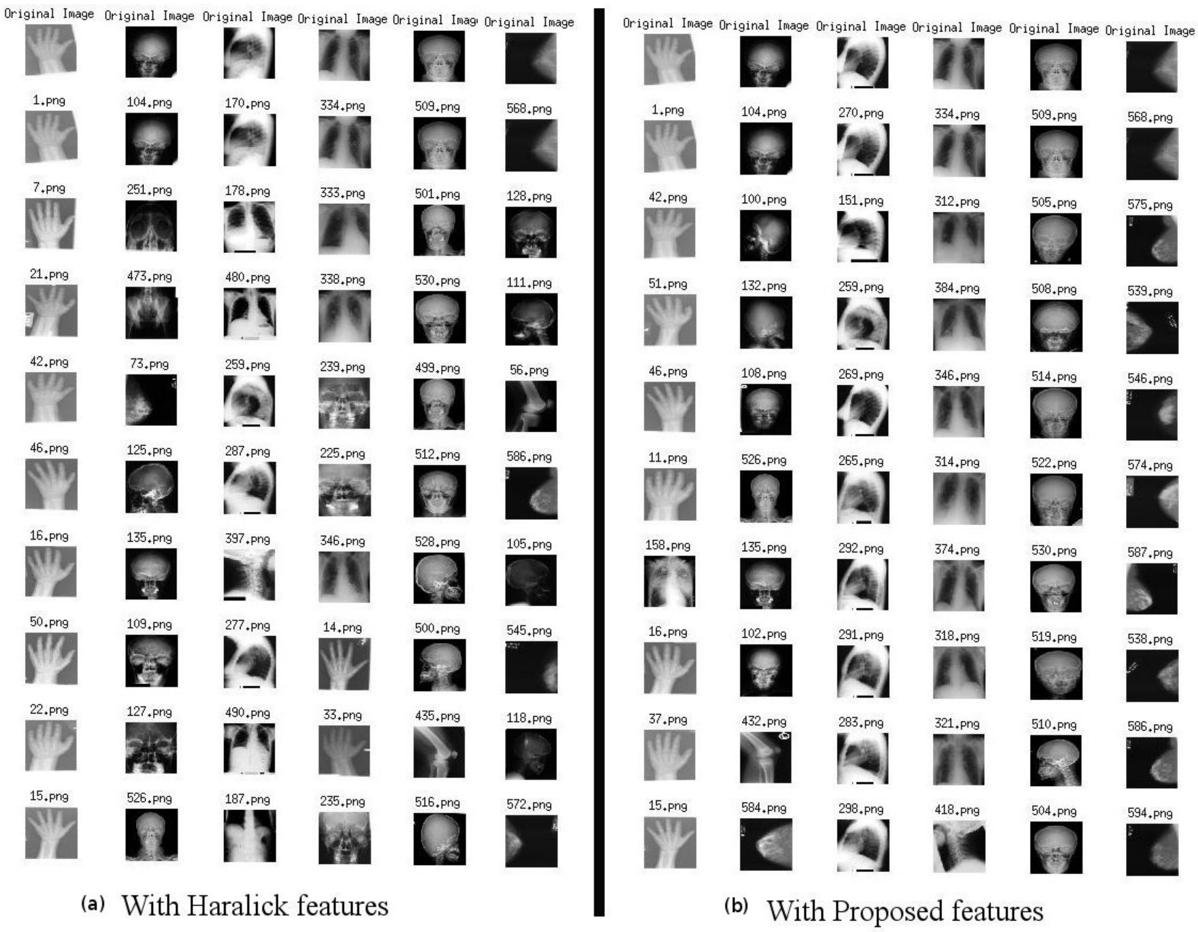


Figure 14: On every column, the first image represents query image while the remaining are those retrieved by the proposed and Haralick features with Euclidean distance as similarity metric.

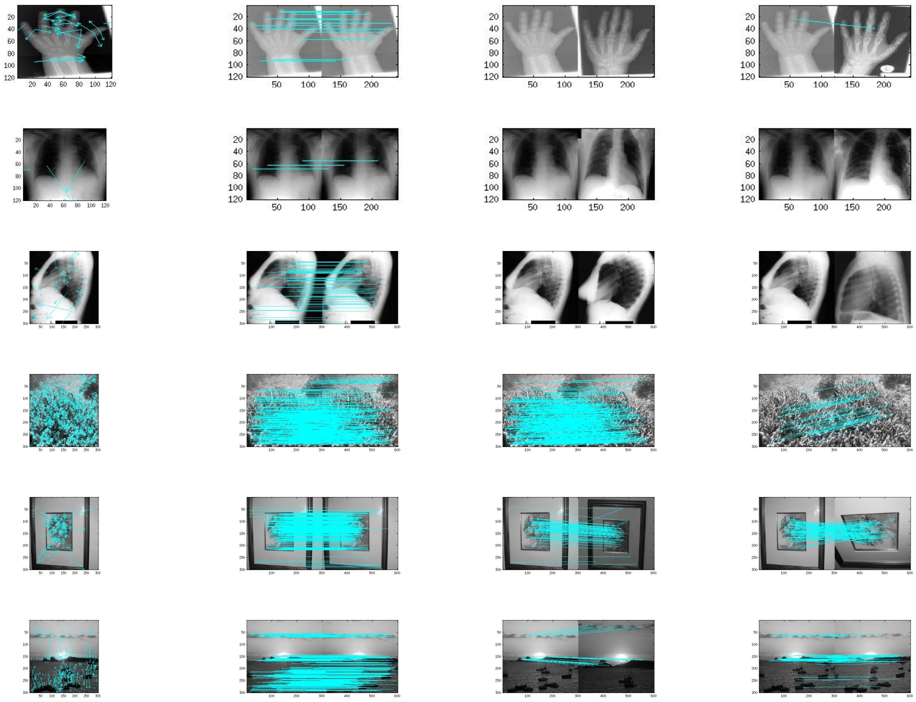


Figure 15: This figure shows the SIFT descriptors extracted from medical images (IRMA database) and natural images (Holiday dataset [28]). The first three rows correspond to medical images and the last three rows are of natural images. The first image on each row depicts the class and the SIFT descriptors extracted from it. For the 1st class specific image on each row, other three images of same class are matched in terms of SIFT descriptors. This figures shows that between any pair of images of same class, medical images admit less number of descriptors than the nonmedical images.