A Classification-Driven Similarity Matching Framework for Retrieval of Biomedical Images

Md Mahmudur Rahman
U.S. National Library of Medicine,
National Institutes of Health,
Bethesda, MD, USA
rahmanmm@mail.nih.gov

Sameer K. Antani
U.S. National Library of Medicine,
National Institutes of Health,
Bethesda, MD, USA
santani@mail.nih.gov

George R. Thoma
U.S. National Library of Medicine,
National Institutes of Health,
Bethesda, MD, USA
gthoma@mail.nih.gov

ABSTRACT

This paper presents a classification-driven biomedical image retrieval system to bride the semantic gap by transforming image features to their global categories at different granularity, such as image modality, body part, and orientation. To generate the feature vectors at different levels of abstraction, both the visual concept feature based on the "bag of concepts" model that comprise of local color and texture patches and various low-level global color, edge, and texture-related features are extracted. Since, it is difficult to find a unique feature to compare images effectively for all types of queries, we utilize a similarity fusion approach based on the linear combination of individual features. However, instead of using the commonly used fixed or hard weighting approach, we rely on the image classification to determine the importance of a feature at real time. For this, a supervised multi-class classifier based on the support vector machine (SVM) is trained on a set of sample images and classifier combination techniques based on the rules derived from the Bayes's theorem are explored. After the combined prediction of the classifiers for a query image category, the individual pre-computed weights of different features are adjusted in the similarity matching function for effective query-specific retrieval. Experiment is performed in a diverse medical image collection of 67,000 images of different modalities. It demonstrates the effectiveness of the category-specific similarity fusion approach with a mean average precision (MAP) score of 0.0265 when compared to using only a single feature or equal weighting of each feature in similarity matching.

Categories and Subject Descriptors

H.3 [**Information Storage and Retrieval**]: H.3.3 Information Search and Retrieval; I.4.8 [**Image Processing and Computer Vision**]: Applications

General Terms

Algorithms,Design,Experimentation,Performance

MIR'10, March 29–31, 2010, Philadelphia, Pennsylvania, USA.

Keywords

Medical Imaging, Content-based Image Retrieval, Classification, Support Vector Machine, Classifier Combination, Similarity Matching

1. INTRODUCTION

The digital imaging revolution in the medical domain over the past three decades has changed the way the present-day physicians diagnose and treat diseases. There is an increasing number of digital images of diverse modalities being produced using sophisticated image acquisition devices in the hospitals and medical centers. These images are playing an important role in diagnosis, medical research and education. In medical imaging, the focus now includes more effective post-processing, organization, and retrieval. In this context, it is a goal to support the clinical decision making by retrieving and displaying the relevant cases using all available information, such as electronic patient records (text), as well as images (visual content). Search results in medical collections might be improved by combining text attribute-based search capability with low-level visual features computed directly on the image content, commonly known as the content-based image retrieval (CBIR) [1].

During the last decade, several medical CBIR prototypes have been proposed [2–6]. Majority of these are developed around a specific imaging modality, e.g., the ASSERT system [3] is designed for high resolution computed tomography (HRCT) images of the lung and the SPIRS [1] system [4] for digitized X-rays of the spine. However, few research projects have a goal create CBIR systems for heterogeneous image collections [5–7], e.g., the IRMA (Image Retrieval in Medical Applications) system [5] is an ongoing project [2] that retrieves from a large set of radiological images of different anatomical regions, acquisition views, and biological systems based on various low-level texture features. In general, CBIR systems for the heterogeneous medical image collection can be classified on the basis of their image feature extraction and processing capabilities. Some systems utilize low-level visual features computed over the entire image while others compute the features on localized regions.

Some approaches have been also explored recently to classify medical image collections into multiple semantic categories for effective retrieval [8–11]. For example, the automatic categorization of radiological images is examined

[1]http://archive.nlm.nih.gov/proj/spirs/
[2]http://irma-project.org/

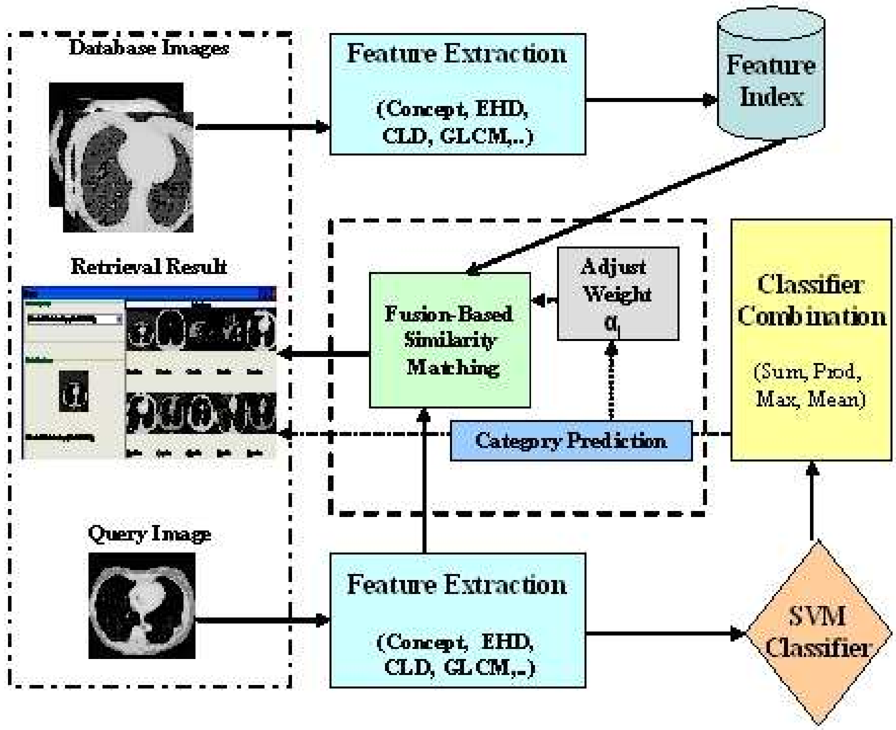

Figure 1: Block diagram of the classification-driven similarity fusion framework.

in [8] by utilizing a combination of low-level global texture features with low-resolution scaled images and a K-nearest-neighbors (KNN) classifier. In [9], the performances of two medical image categorization architectures with and without a learning scheme are evaluated on images based on modality, body part, and orientations. These approaches demonstrated promising results for medical image classification at a global level. However, the approaches of these kinds do not relate classification to retrieval in a direct manner, instead only stressed the usefulness of it for the image annotation and pre-filtering purposes.

We present a classification-driven image retrieval framework by transforming various low-level and concept-based image features at different levels to their global categories. The category-specific information are utilized directly to adjust the feature weights in a fusion-based similarity matching function for the ranked-based retrieval. The global categories, such as image modalities, body parts, orientations, and distinct visual features are inferred from the entire images instead of the individual object based semantics. For this, several supervised multi-class classifiers based on the support vector machine (SVM) [12] are trained for individual features on a set of sample images to associate the image features to their global categories. The classifier combination techniques based on the rules derived from the Bayes's theory [13] are then explored to obtain a final prediction about the query image category. Hence, for the SVM training, the initial inputs are the feature vectors of the sample images in which each vector is associated with a category label out of the M predefined categories. To generate the feature vectors at different levels of abstraction, both the visual concept feature based on the *"bag of concepts"* model [14] and various low-level global color, edge, and texture-related features are extracted. These feature descriptors at different levels of image representation are in diversified forms and are often complementary in nature.

The block diagram of the proposed image retrieval framework is shown in Fig. 1. As can be seen from the top portion of Fig. 1, various image features are extracted in a feature extraction sub-system and stored in a feature index for the database images for later similarity matching with query images as shown in the middle of the figure. For a query image, similar feature extraction is performed as database images as shown in the bottom portion of Fig. 1. However, instead of performing the similarity matching, the global category of a query image is determined at first in this approach. The query features are sent to the corresponding SVM classifiers to get probabilistic outputs or class confidence scores of each category for individual features. Next, the outputs are sent to the classifier combination sub-system, which produces a ranking of the M category labels and finally classify a query image to the category with the highest obtained probability value. Based on the on-line category prediction, the pre-defined category-specific feature weights are utilized in a linear combination of similarity matching function as shown in the middle portion of Fig. 1. In this scheme, for example, a color feature would have more weight for microscopic pathology and dermatology images, whereas edge and texture-related features would have more weights for the radiographs.

The rest of the paper is organized as follows. Section 2 describe the image categorization approach at a global level by utilizing the SVM and the classifier combination approach. Section 3 presents the feature representations approaches both at the concept and content level. The similarity fusion approach based on image categorization is presented in Section 1. The experiments and the analysis of the results are presented in Sections 5 and 6 respectively, and finally Section 7 provides the conclusion.

2. GLOBAL IMAGE CATEGORIZATION BY MULTI-CLASS SVM

The variation of the medical image categories at a global level can be effectively modeled by using any supervised learning-based classification techniques [15]. The task of a supervised learner or classifier is to predict the label of a newly encountered unknown data after having seen only a small number of training examples. To achieve this, the learner has to generalize from the presented data to unseen situations in a reasonable way. Due to the empirical success or good generalization ability of SVM [12], it is utilized to classify images to multiple categories. In its basic formulation, the SVM is a binary classification method that constructs a decision surface and maximizing the inter-class boundary between the samples. Stated mathematically,

$$f(\mathbf{x}) = \operatorname{sign}\left(\sum_{i=1}^{N} \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b\right) \tag{1}$$

where $\mathbf{x} \in \Re^d$ is an input vector, $\mathbf{x}_i$ is a training sample vector along with its label $y_i \in (+1, -1)^N$, b is a bias and K is a kernel function which maps the vectors into a higher dimensional space by the non-linear mapping function $\phi : \Re^d \rightarrow \Re^l$, where $l > d$ or l could even be infinity.

A number of methods have been proposed for the extension of the SVM to multi-class classification problems [16]. We utilize one such a method by combining all pairwise comparisons of binary SVM classifiers, known as *one-against-one* or pairwise coupling (PWC) [17]. During the testing of a feature vector $\mathbf{x}$, each classifier votes for one class. The winning class is the one with the largest number of accumulated votes. For the SVMs training, a set of M labels are assigned as $C = \{c_1, \cdots, c_i, \cdots, c_M\}$, where each $c_i \in C$ characterizes a global concept category. The initial input to the system is the sets of feature vectors along with their manually assigned corresponding concept labels. During the testing phase, each category will assign a probability or confidence score to each image I_j as

$$p_m(\mathbf{x}_j) = P(y = m \mid \mathbf{x}_j), \text{for } 1 \leq \text{m} \leq \text{M} \tag{2}$$

The pairwise class probabilities r_{uv} are estimated as an approximation of the original pairwise class probabilities μ_{uv} following the setting of the PWC ensemble in [17]:

$$r_{uv} \approx P(y = u \mid y = u \textit{ or } v, \mathbf{x}_j) \approx \frac{1}{1 + e^{A\hat{d}+B}} \tag{3}$$

where A and B are the parameters estimated by minimizing the negative log-likelihood function, and $\hat{d}$ are the decision values of the training data.

2.1 Classifier Combination

Feature descriptors at different levels of image representation are in diversified forms and are often complementary in nature. Different features represent image data from different viewpoints; hence the simultaneous use of different feature sets can lead to a better or robust classification result. For simultaneous use of different features, a traditional method is to concatenate different feature vectors together into a single composite feature vector. However, it is rather unwise to concatenate them together since the dimension of a composite feature vector becomes much higher than any of individual feature vectors. Hence, multiple classifiers are needed to deal with different features resulting in a general problem of combining those classifiers to yield improved performance. In general, a classifier combination is defined as the instances of the classifiers with different structures trained on the distinct feature spaces. It has been realized that such systems can be more robust and more accurate than using a single classifier alone [13].

Four popular classifier combination techniques derived from the Bayes's theory, such as the product, sum, max, and mean rules [13, 15] are considered for the expert combination strategies. Since the outputs of the classifiers are to be used in combination, the *a posterior* probabilities in the range of $[0, 1]$ for each category will serve this purpose. In these combination rules, *a priori* probabilities are assumed to be equal and the decision is made by the following formula in terms of the *a posteriori* probabilities yielded by the respective classifiers as

$$\omega_m \Longleftrightarrow \max_{1 \leq m \leq M} p_m^O, \; O \in \{\text{prod}, \text{sum}, \text{max}, \text{mean}\} \tag{4}$$

where the product, sum, max, and mean rules are defined as

$$p_m^{\text{prod}} = \frac{\prod_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j)}{\sum_{m=1}^{M} \prod_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j)} \tag{5}$$

$$p_m^{\text{sum}} = \frac{\sum_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j)}{\sum_{m=1}^{M} \sum_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j)} \tag{6}$$

$$p_m^{\max} = \max_{1 \leq r \leq R} p_r(\omega_m|\mathbf{x}_j) \tag{7}$$

and

$$p_m^{\text{mean}} = \frac{1}{R} \sum_{r=1}^{R} p_r(\omega_m|\mathbf{x}_j) \tag{8}$$

Here, $p_r(\omega_m|\mathbf{x}_j)$ is the *a posteriori* probabilities yielded by an expert r for $1 \leq r \leq R$. In the product rule, it is assumed that the representations used are conditionally statistically independent. In addition to the conditional independence assumption of the product rule, the sum rule assumes that the probability distribution will not deviate significantly from the a priori probabilities. Classifier combination based on these two rules often performs better than the other rules, such as max and mean [13].

The multi-class SVM classifiers on different feature spaces (will be described in the following section) are combined or fused by the above rules and will finally classify an unknown query image to the category with the highest obtained probability value.

3. IMAGE FEATURE REPRESENTATION

The success of a CBIR system depends on the underlying image representation, usually in the form of a feature vector. To generate the feature vectors at different levels of abstraction, we extract both the concept-based feature [14] based on "bag of concepts" model that comprise of color and texture patches from local image regions and various low-level global color, edge, and texture-related image features.

3.1 Concept-Based Image Representation

In a heterogeneous medical image collection, it is possible to identify specific local patches in images that are perceptually and/or semantically distinguishable, such as homogeneous texture patterns in grey level radiological images, differential color and texture structures in microscopic pathology and dermoscopic images. The variation in the local patches can be effectively modeled as visual concepts (keywords) [14] by using supervised learning-based classification techniques, such as the SVM [12].

For concept model generation, we utilize the similar (as described in previous section) voting-based multi-class SVM known as *one-against-one* or pairwise coupling (PWC) [17]. In order to perform the learning, a set of L labels are assigned as $C = \{c_1, \cdots, c_i, \cdots, c_L\}$, where each $c_i \in C$ characterizes a visual concept (keyword). The training set of the local patches that are generated by a fixed-partition based approach and represented by a combination of color and texture moment-based features. For SVM training, the initial input to the system is the feature vector set of the patches along with their manually assigned corresponding concept labels. Images in the data set are annotated with visual concept labels by fixed partitioning each image I_j into l regions as $\{\mathbf{x}_{1_j}, \cdots, \mathbf{x}_{k_j}, \cdots, \mathbf{x}_{l_j}\}$, where each $\mathbf{x}_{k_j} \in \Re^d$ is a combined color and texture feature vector. For each $\mathbf{x}_{k_j}$, the visual concept probabilities are determined by the prediction of the multi-class SVMs as [17]

$$p_{ik_j} = P(y = i \mid \mathbf{x}_{k_j}),\ 1 \leq i \leq L. \quad (9)$$

Finally, the category label of x_{k_j} is determined as c_m, which is the label of the category with the maximum probability score. Hence, instead of the low-level feature-based representation, an entire image is thus represented as a two-dimensional index linked to the visual keyword (concept) labels. Based on this encoding scheme, an image I_j is represented as a vector of weighted visual keywords as

$$\mathbf{f}_j^I = \ [w_{1_j}, \cdots, w_{i_j}, \cdots w_{L_j}]^{\mathrm{T}} \quad (10)$$

where each w_{i_j} denotes the weight a visual keyword $c_i, 1 \leq i \leq L$ in image I_j, depending on its information content. The popular *"tf-idf"* term-weighting scheme [18] is used in this work, where the element w_{i_j} is expressed as the product of local and global weights.

3.2 Low Level Global Feature Representation

In addition to the visual concept feature, we extract the following global features.

Color Feature: To represent the spatial structure of images, we utilize the Color Layout Descriptor (CLD) of MPEG-7 [19]. The CLD represents the spatial layout of the images in a very compact form. It is obtained by applying the discrete cosine transformation (DCT) on the 2-D array of local representative colors in the $YCbCr$ color space where Y is the luma component and Cb and Cr are the blue and red chroma components. Each channel is represented by 8 bits and each of the 3 channels is averaged separately for the 8×8 image blocks. In this work, a CLD with only 10 Y, 3 Cb, and 3 Cr, is extracted to form a 16-dimensional feature vector.

Image are also represented as Color Coherence Vector (CCV) [20] where a color's coherence is defined as the degree to which pixels of that color are members of large similarly-colored regions. Each pixel is classified in a given color bucket as either coherent or incoherent, based on whether or not it is part of a large similarly-colored region. A color coherence vector (CCV) stores the number of coherent versus incoherent pixels with each color. By separating coherent pixels from incoherent pixels, CCV's provide finer distinctions than color histograms.

Edge Feature: To represent the global shape/edge feature, the spatial distribution of edges are utilized by the Edge Histogram Descriptor (EHD) [19]. The EHD represents local edge distribution in an image by dividing the image into 4×4 sub-images and generating a histogram from the edges present in each of these sub-images. Edges in the image are categorized into five types, namely vertical, horizontal, 45° diagonal, 135° diagonal and non-directional edges. Finally, a histogram with $16 \times 5 = 80$ bins is obtained, corresponding to a 80-dimensional feature vector.

In addition, a histogram of edge direction is constructed where the edge information contained in the images is processed and generated by using the Canny edge detection (with $\sigma = 1$, Gaussian masks of size = 9, low threshold = 1, and high threshold = 255) algorithm.The corresponding edge directions are quantized into 72 bins of 5° each. Scale invariance is achieved by normalizing this histograms with respect to the number of edge points in the image.

Texture Feature: We extract texture features from the grey level co-occurrence matrix (GLCM) [21] of each image. In order to obtain efficient descriptors, the information contained in GLCM is traditionally condensed in a few statistical features. Four GLCM's for four different orientations (horizontal 0°,vertical 90°, and two diagonals 45° and 135°) are obtained and normalized to the entries [0,1] by dividing each entry by total number of pixels. Higher order features, such as energy, entropy, contrast, homogeneity and maximum probability are measured based on averaging features in GLCMs to form a 20-dimensional feature vector for an entire image.

Finally, two more features are extracted as Color Edge Direction Descriptor (CEDD) and Fuzzy Color Texture Histogram (FCTH) from the Lucene image retrieval (LIRE) library [22]. CEDD incorporates color and texture information into one single histogram and it requires low computational power in extracting comparing to MPEG7 descriptors. To extract texture information, CEDD uses a fuzzy version of the five digital filters proposed by the MPEG-7 EHD, forming 6 texture areas [23]. This descriptor is appropriate for accurately retrieving images even in distortion cases such as deformation, noise and smoothing. In contrast, FCTH uses the high frequency bands of the Haar wavelet Transform in a fuzzy system, to form 8 texture areas [24].

4. SIMILARITY FUSION

It is difficult to find a feature representation to compare images accurately for all types of queries. In other words, each feature representation along with its distance measure might be complementary in nature and will have its own limitations. In recent years, the category of work known as data fusion or multiple-evidence described a range of techniques in information retrieval (specially in text retrieval domain) whereby multiple pieces of information are combined to achieve improvements in retrieval effectiveness. These pieces of information can take many forms including different query representations, different document (image) representations, and different retrieval strategies used to obtain a measure of relationship between a query and a document (image). Many researchers have argued that better retrieval effectiveness may be gained by exploiting multiple query representations, retrieval algorithms or feedback techniques and combining the results of a varied set of techniques or representations. [25, 26].

CBIR also adopts some of the ideas from data fusion research, where the most commonly used approach is a linear combination of similarity matching scores of different features with pre-determined weights. In this framework, the similarity between a query image I_q and target image I_j is described as

$$\mathrm{Sim}(I_q, I_j) = \sum_F \omega^F \ \mathrm{Sim}^F(I_q, I_j) \tag{11}$$

where $F \in \{\mathrm{Concept, EHD, CLD, CCV, CEDD, FCTH, etc.}\}$ and ω^F are the weights within the different image representation.

However, there is a problem with the pre-determined hard-coded or fixed weight based approach. In this approach, for example, a color feature will have the same weight for the search of the microscopic pathology or X-ray images. Although, the importance of the color feature is negligible for many modalities, such as X-ray, CT, or MRI. We present an adaptive linear combination scheme based on the on-line category prediction of the unknown query images. In our approach, based on the on-line category prediction, precomputed category-specific feature weights (e.g., ω^F) are utilized in the linear combination similarity matching function instead of using the fixed weights for each query. Before performing any linear combination, the distance measure scores of each representation are normalized and converted to the similarity scores with a range of $[0, 1]$ as

$$\mathrm{Sim}(\mathbf{f}_q, \mathbf{f}_j) = 1 - \frac{\mathrm{Dis}(\mathbf{f}_q, \mathbf{f}_j) - \min(\mathrm{Dis}(\mathbf{f}_q, \mathbf{f}_j))}{\max(\mathrm{Dis}(\mathbf{f}_q, \mathbf{f}_j)) - \min(\mathrm{Dis}(\mathbf{f}_q, \mathbf{f}_j))} \tag{12}$$

where $\min(\cdot)$ and $\max(\cdot)$ are the minimum and maximum distance scores between query and database images (documents) for a particular feature vector $\mathbf{f}$. Generally, a similarity score is the converse of a distance score. So, when the similarity score is one (i.e. exactly similar), the distance score is zero and vice versa. The steps involved in the overall similarity matching approach are depicted in Algorithm 1.

Algorithm 1 Fusion-based Similarity Search Approach

1: (Off-line): Select a set of N training images of M categories with associated category label for each feature vector $\mathbf{f}^{\mathrm{F}}$ of where $F \in \{\mathrm{Concept, EHD, CLD, CCV, CEDD, FCTH, etc.}\}$.
2: (Off-line): Train a SVM classifier for each individual feature to generate the model files.
3: (Off-line): Store category specific weights (manually defined based on the importance of each category) for each feature to be used for similarity matching.
4: (On-line): For a query image I_q, extract both the concept and low-level global features and represent as vectors $\mathbf{f}_q^{\mathrm{F}}$.
5: For each feature, get a category prediction based on the probabilistic output of (2) by applying SVM.
6: Combine the outputs by applying any of the combination rules of (4).
7: Get the final category label as $C_k(j), k \in \{1, \cdots, M\}$ of the query image.
8: Consider the individual features weights ω^F for the specific query image category.
9: Finally, combine the similarity scores with the weights based on similarity fusion in (11).
10: Finally return the images based on the similarity matching values in descending order to obtain a final ranked list of images.

5. EXPERIMENTS

To evaluate the retrieval effectiveness, experiments are performed on a benchmark medical image collection under ImageCLEFmed'08 [7]. This collection contains more than 67,000 images of different modalities from the RSNA journals Radiology and Radiographics. The contents of this collection represent a broad and significant body of medical knowledge, which make the retrieval more challenging. The experimental results are generated based on the 30 query topics (e.g., a short sentence or phrase describing the search request in a few words with one to three relevant images) with ten topics in each of three categories: visual, mixed, and semantic [7]. The relevant sets of all topics were created by the CLEF organizers by considering top retrieval results of all submitted runs of the participating groups. Retrieval results are computed using the latest version of TREC -EVAL [3] software and evaluated using un interpolated (arithmetic) Mean Average Precisions (MAP) and Precision at rank 20 (P20) because most online image retrieval engines like Google, Yahoo, and Altavista display 20 images by default. Further measures considered include Geometric Mean Average Precision (GMAP) to test robustness, and the Binary Preference (BPREF) measure which is a good indicator for the completeness of relevance judgments [18].

5.1 Training for global classification by SVM

The training set for global classification contains 5000 images of 32 manually assigned and mutually exclusive categories, which is a subset of the ImageCLEFmed'08 collection. In this set, images are classified into three levels of details based on modalities (e.g., X-ray, CT, MRI, etc.), body parts (e.g., head, chest, knee, etc.), and orientation (e.g., frontal, coronal, sagittal, etc) or with distinct visual attributes as shown in Fig. 2.

For the training, we use the radial basis function (RBF), $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma ||\mathbf{x}_i - \mathbf{x}_j||^2), \gamma > 0$, as the kernel. There are two tunable parameters while using the RBF kernels: C and γ. The kernel parameter γ controls the shape of the kernel and the regularization parameter C controls the trade-offs between margin maximization and error minimization. Increasing C may decrease the training error but it can also lead to a poor generalization. It is not known beforehand which C and γ are the best choices for the classification problem at hand and are selected by cross-validation (CV). In the training stage, the goal is to identify the best (C and γ), so that the classifier can accurately predict the testing data. For the training set, a 10-fold CV is conducted where we first divide the training set into 10 subsets of equal size. Sequentially one subset is tested using the classifier trained

[3]http://trec.nist.gov/$trec - eval$/

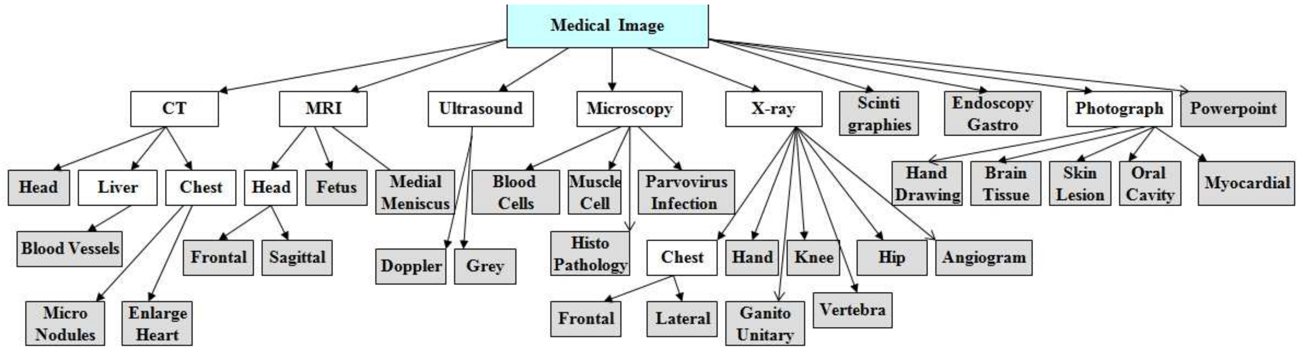

Figure 2: Global classification structure of the medical collection.

Table 1: Statistics of the Training Set for Local Concepts

Concept	#	Concept	#
CT-Tissue-Brain	400	CT-Tissue-Abdomen	380
CT-Tissue-Lung	350	EC-Tissue-Gastro	300
CT-Bone-Whole	350	CT-Bone-Corner	300
Skin-Normal	420	Skin-Melanoma	370
Dental-White	330	Tissue-Mouth	310
MRI-Tissue-Leg	320	MRI-Tissue-Brain	350
Xray-Tissue-Lung	400	Xray-Hepatic-Veins	450
Xray-Chest-bone	260	Micro-Infection	200
US-Grey-Texture	445	Doppler-color	400
Micro-Blood-Blue	490	Histo-Blue	480
Micro-Pink-Bacteria	520	Micro-Tissue-Muscle	500
Photo-Tissue-Brain	520	Photo-Tissue-Cardiac	500
Backgrnd-White	500	Backgrnd-Grey	510
Backgrnd-Blue	350	Backgrnd-Brown	550
Backgrnd-Black	350	Line-Drawing	550

on the remaining 9 subsets. Thus, each instance of the whole training set is predicted once so the cross-validation accuracy is the percentage of data which are correctly classified. We perform a *grid-search* on C and γ by using cross-validation. Basically, pairs of (C , γ) are tried and the ones with the lowest CV error rates are picked for the different feature representations.

5.2 Training for local concept classification by SVM

For the visual concept model generation based on the SVM learning, 30 local concept categories are manually defined, as shown in Table 1. The training set used for this purpose consist of only 1% images of the entire data set of 67000 images. To generate the local patches, each image in the training set is at first partitioned into an 8×8 grid generating 64 non-overlapping regions. Only the regions that conform to at least 80% of a particular concept category are selected and labeled with the corresponding category label. For the SVM training, we again use RBF kernel with a 10-fold CV to find the best values of tunable parameters C and γ. After finding the best values of the parameters $C = 200$ and $\gamma = 0.02$ of the RBF kernel with a CV accuracy of 81.01%, they are utilized for the final training to generate the concept model. We utilized the *LIBSVM* software package [27] for implementing the multi-class SVM classifiers for both global and local concept classification.

Table 2: Retrieval Results

Method	MAP	R-prec	B-Perf
EHD	0.0032	0.0088	0.0259
CLD	0.0014	0.0043	0.0107
Visual Concept	0.025	0.0092	0.0208
Fusion (Equal Weight)	0.0175	0.0291	0.0529
Fusion (CV)	0.0239	0.0430	0.0732
Fusion (Category)	0.0265	0.0485	0.0806

6. RESULTS

This section presents the experimental results of the retrieval approaches with or without using similarity fusion. The performances of the different search schemes are shown in Table 2 for the retrieval of medical collection based on the query image set as discussed previously. The proposed category-specific similarity fusion approach (e.g., Fusion (Category)) is compared with similarity fusion with equal weighting (e.g., Fusion (Equal Weight)) and fusion with weights based on the CV accuracies of different features (e.g., Fusion (CV)). The weights are normalized based on the CV accuracies of the features, which subject to $0 \leq \omega^F \leq 1$ and $\sum \omega^F = 1$ for $F \in \{\text{Concept}, \text{EHD}, \text{CLD}, \text{CCV}, \text{etc.}\}$. In addition, we consider the MPEG-7 based EHD and CLD, and Concept features to compare our fusion search approaches with these single features based on the Euclidean distance measure. For the category-based similarity fusion, we applied the product rule of classifier combination as it provides the best classification accuracy in this collection when compared to the other rules. It is clear from Table 2 that performance is improved for the category-specific similarity search in terms of the MAP, R-prec and B-Perf scores when compared to the other similarity fusion approaches or approaches based on the search only in a single feature space.

Fig. 6 shows the precision values at different rank position (e.g., 5, 10, 20, 30, 100) for the photographic and medical collections respectively. From Fig. 6, we can conjecture that the precision in different rank positions are comparatively much better for the category specific similarity fusion approaches (e.g., Fusion (Category) and Fusion (CV)) when compared to the similarity fusion based on equal weighting and using a single feature only (e.g., EHD and CLD). In ad-

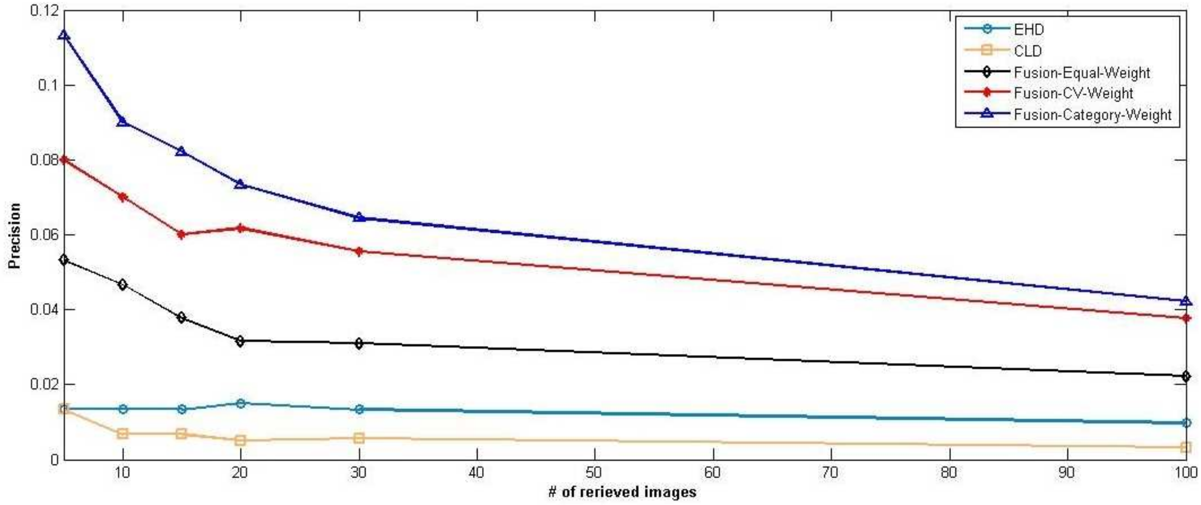

Figure 3: Average Precision curves for different similarity matching schemes

dition, the precisions in initial rank positions (up to 30) are significantly better for fusion approaches as compared to the single low level color and edge features. This improvement in performance is important as users are usually interested only on the first few top retrieved images. In general, we achieved around 30-40% increases in MAP scores and precision at different ranks as shown in Table 2 and Fig. 6 when compared to the searches without using any category information of the collection.

7. CONCLUSION

In this paper, a novel image retrieval framework based on image categorization, concept feature representation and retrieval is proposed for the diverse medical image collections of different modalities. Unlike few other approaches where image categorization is the very first step of image processing for filtering out irrelevant images, we have taken a different approach. In this framework, the category information is utilized directly to adjust the feature weights in a linear combination of similarity matching. Specially, we explore the utilization of the probabilistic multi-class SVM and various classifier combination rules in different aspects of the image feature spaces for the categorization, representation and similarity matching of the images. Overall, this framework might be useful as a front-end for medical databases where a search can be performed in diverse images for teaching, training and research purposes. In future, we will investigate to incorporate other learning methodologies such as, boosting and relevance feedback (RF) and will integrate the textual modality in the framework.

Acknowledgment

This research is supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). We would like to thank the CLEF [7] organizers for making the database available for the experiments.

8. REFERENCES

[1] A. Smeulder, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 12, pp. 1349–1380, 2000.

[2] H. Müller, N. Michoux, D. Bandon, A. Geissbuhler, "A Review of Content-Based Image Retrieval Systems in Medical Applications Clinical Benefits and Future Directions,"*Int J Med Inform.*, vol.73, pp. 1–23, 2004.

[3] C.R. Shyu, C.E. Brodley, A.C. Kak, A. Kosaka, A.M. Aisen, and L.S. Broderick, "ASSERT: a physician-in-the-loop content-based image retrieval system for HRCT image databases," *Comput Vis Image Understanding*, vol. 75, pp. 111–132, 1999.

[4] W. Hsu, S. Antani, L.R. Long, L. Neve, and G.R. Thoma, "SPIRS: a Web-based Image Retrieval System For Large Biomedical Databases," *International Journal of Medical Informatics*, vol. 78, pp. 13–24, 2008.

[5] T.M. Lehmann, B.B. Wein, J. Dahmen, J. Bredno, F. Vogelsang, and M. Kohnen, "Content–based image retrieval in medical applications-A novel multi-step approach," *Proc SPIE*, vol. 3972, pp. 312-320, 2000.

[6] H. Müller, A. Rosset, J. Vallee, and A. Geissbuhler, "Integrating content-based visual access methods into a medical case database," *Proc Med Inform Europe (MIE 2003)*, St Malo, France, pp. 480–485, 2003.

[7] H. Müller, J. Kalpathy-Cramer, C.E. Kahn Jr., W. Hatt, S. Bedrick, and W. Hersh "Overview of the ImageCLEFmed 2008 Medical Image Retrieval Task," *Working Notes for the CLEF 2008 Workshop*, September 17-19, Aarhus, Denmark.

[8] T.M. Lehmann, M.O. Güld, T. Deselaers, D. Keysers, H. Schubert, K. Spitzer, H. Ney, and B.B. Wein, "Automatic categorization of medical images for content-based retrieval and data mining," *Computerized Medical Imaging and Graphics*, vol. 29, pp. 143–155, 2005.

[9] F. Florea, H. Müller, A. Rogozan, A. Geissbuhler, S. Darmoni, "Medical image categorization with MedIC

and MedGIFT,"*Proc Med Inform Europe (MIE 2006)*, Maastricht, Netherlands, pp. 3–11, 2006.

[10] M.M. Rahman, B.C. Desai, and P. Bhattacharya, "Medical Image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion," *Computerized Medical Imaging and Graphics*, vol. 32, pp. 95–108, 2008.

[11] A. Mojsilovic, and J. Gomes, "Semantic based image categorization, browsing and retrieval in medical image databases," *Proceedings of IEEE International Conference of Image Processing*, vol. 3, pp. 145–148, 2002.

[12] V. Vapnik, *Statistical Learning Theory*, New York, NY, Wiley; 1998.

[13] J. Kittler, M. Hatef, R.P.W. Duin, and J. Matas, "On combining classifiers," *IEEE Trans Pattern Anal Machine Intell.*, vol. 20(3), pp. 226–239, 1998.

[14] M.M. Rahman, S.K. Antani, and G.R. Thoma, "A Medical Image Retrieval Framework in Correlation Enhanced Visual Concept Feature Space," *22nd IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, August 3-4, 2009, Albuquerque, New Mexico, USA.

[15] R.O. Duda, P.E. Hart, and D.G. Stork, *Pattern Classification*, 2nd ed. Canada: John Wiley & Sons Ltd., 2001.

[16] K. Duan, and S.S. Keerthi, "Which is the best multiclass SVM method? An empirical study," *Proc of the Sixth International Workshop on Multiple Classifier Systems*, pp. 278–285, 2005.

[17] T.F. Wu, C.J. Lin, R.C. Weng, "Probability Estimates for Multi-class Classification by Pairwise Coupling," *Journal of Machine Learning Research*, vol. 5, pp. 975–1005, 2004.

[18] R.B. Yates and B.R. Neto, *Modern Information Retrieval*, Addison Wesley, 1999.

[19] S.F. Chang, T. Sikora, and A. Puri, "Overview of the mpeg-7 standard," *IEEE Trans Circ Syst Video Technology*, vol. 11, pp. 688–695, 2001.

[20] G. Pass, R. Zabih, and J. Miller, "Comparing images using color coherence vectors," *Proc of the fourth ACM international conference on Multimedia*, pp. 65–73, 1997.

[21] R.M. Haralick Shanmugam, I. Dinstein, "Textural features for image classification," *IEEE Trans Syst Man Cybernetics*, vol. 3, pp. 610–21, 1973.

[22] M. Grubinger, P. Clough, A. Hanbury, and H. Müller, "Lire: lucene image retrieval: an extensible java CBIR library," *Proc of the 16th ACM international conference on Multimedia* , Vancouver, British Columbia, Canada, pp. 1085–1088, 2008.

[23] S.A. Chatzichristofis and Y.S. Boutalis, "Cedd: Color and edge directivity descriptor. a compact descriptor for image indexing and retrieval," *Proc of the 6th International Conference on Computer Vision Systems, ICVS 2008*, vol. 5008 of LNCS, pp. 312Ű-322, 2008.

[24] S.A. Chatzichristofis and Y.S. Boutalis, "Fcth: Fuzzy color and texture histogram a low level feature for accurate image retrieval," *Proc of the 9th International Workshop on Image Analysis for Multimedia Interactive Services*, WIAMIS 2008.

[25] E.A. Fox and J.A. Shaw, "Combination of Multiple Searches," *Proceedings of the 2nd Text Retrieval Conference (TREC-2)*, NIST Special Publication 500-215, pp. 243–252, 1994.

[26] J.H. Lee, "Combining Multiple Evidence from Different Properties of Weighting Schemes," *Proceedings of the 18th Annual ACM-SIGIR*, pp. 180–188, 1995.

[27] C.C. Chang and C.J. Lin, "LIBSVM : a library for support vector machines," Software available at http://www.csie.ntu.edu.tw/ cjlin/libsvm, 2001.