



Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion

Md. Mahmudur Rahman ^{a,*}, Bipin C. Desai ^a, Prabir Bhattacharya ^b

^a Department of Computer Science & Software Engineering, Concordia University, Montreal, Canada

^b Institute for Information Systems Engineering, Concordia University, Montreal, Canada

Received 1 August 2006; received in revised form 30 September 2007; accepted 2 October 2007

Abstract

We present a content-based image retrieval framework for diverse collections of medical images of different modalities, anatomical regions, acquisition views, and biological systems. For the image representation, the probabilistic output from multi-class support vector machines (SVMs) with low-level features as inputs are represented as a vector of confidence or membership scores of pre-defined image categories. The outputs are combined for feature-level fusion and retrieval based on the combination rules that are derived by following Bayes' theorem. We also propose an adaptive similarity fusion approach based on a linear combination of individual feature level similarities. The feature weights are calculated by considering both the precision and the rank order information of top retrieved relevant images as predicted by SVMs. The weights are dynamically updated by the system for each individual search to produce effective results. The experiments and analysis of the results are based on a diverse medical image collection of 11,000 images of 116 categories. The performances of the classification and retrieval algorithms are evaluated both in terms of error rate and precision-recall. Our results demonstrate the effectiveness of the proposed framework as compared to the commonly used approaches based on low-level feature descriptors.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Medical imaging; Content-based image retrieval; Classification; Support vector machine; Classifier combination; Similarity fusion; Inverted file

1. Introduction

The digital imaging revolution in the medical domain over the past three decades has changed the way present-day physicians diagnose and treat diseases. Hospitals and medical research centers produce an increasing number of digital images of diverse modalities every day [1–4]. Examples of these modalities are the following: standard radiography (RX), computer tomography (CT), magnetic resonance imaging (MRI), ultrasonography (US), angiography, endoscopy, microscopic pathology, etc. These images of various modalities are playing an important role in detecting the anatomical and functional information about different body parts for the diagnosis, medical research, and education. Due to the huge growth of the World Wide Web, medical images are now available in large numbers in online repositories and atlases [1,5]. Modern medical information systems need to handle these valuable resources effectively and efficiently. Cur-

rently, the utilization of medical images is limited due to the lack of effective search methods; text-based searches have been the dominating approach for medical image database management [1,2]. Many hospitals and radiology departments nowadays are equipped with Picture Archiving and Communications Systems (PACS) [6,7]. In PACS, the images are commonly stored, retrieved and transmitted in the DICOM (Digital Imaging and Communication in Medicine) format [8]. Such systems have many limitations because the search for images is carried out according to the textual attributes of image headers (such as standardized description of the study, patient, and other technical parameters). The annotations available are generally very brief in the majority of cases as they are filled out automatically by the machine. Moreover, in a web-based environment, medical images are generally stored and accessed in common formats such as JPEG (Joint Photographic Experts Group), GIF (Graphics Interchange Format), etc. since they are easy to store and transmit compared to the large size of images in DICOM format. However, there is an inherent problem with the image formats other than DICOM, since there is no header information attached to the images and thus it is not possible to perform

* Corresponding author. Tel.: +1 514 932 0831.

E-mail address: mah.rahm@cs.concordia.ca (P. Bhattacharya).

a text-based search without any associated annotation information.

This explains the need for an effective way to retrieve relevant images automatically from such repositories using purely visual content, commonly known as content-based image retrieval (CBIR) [9,10]. In this case, CBIR systems are capable of carrying out a search for images based on the modality, anatomic region and different acquisition views [5]. Although it is desirable to carry out searches based on pathology, such searches have proven difficult without any associated annotation in the form of case or lab reports [11]. During the last decade, several image retrieval prototypes have been implemented in the medical domain [1]. For instance, the ASSERT system [12] is designed for high resolution computed tomography (HRCT) images of the lung, where a rich set of textural features and attributes that measure the perceptual properties of the anatomy are derived from the pathology-bearing regions (PBR). The WebMIRS¹ system [13] is an ongoing research project. The project aims at the retrieval of cervical spinal X-ray images based on automated image segmentation, image feature extraction, and organization along with associated textual data. I-Browse [14] is another prototype, aimed at supporting the intelligent retrieval and browsing of histological images of the gastrointestinal tract. Many other CBIR systems in the medical domain are currently available. See [1] for a brief introduction. However, the majority of current prototypes or projects concentrate mainly on a specific imaging modality [1]. These systems are task-specific and cannot be transferred to other domains or modalities. The characteristics of medical images (such as, the file format, size, spatial resolution, dimensionality, and image acquisition techniques) differ significantly between modalities. To date, only a few research projects aim at creating CBIR systems for heterogeneous image collections. For example, the IRMA (Image Retrieval in Medical Applications)² system [15] is an important project that can handle retrieval from a large set of radiological images obtained from hospitals based on various textural features. The medGIFT³ project [16] is based on the open source image retrieval engine GNU Image Finding Tool (GIFT). It aims to retrieve diverse medical images where a very high-dimensional feature space of various low-level features is used as visual terms analogous to the use of keywords in a text-based retrieval approach. In Ref. [17], we proposed a retrieval framework for images of diverse modalities by employing machine learning and statistical similarity matching techniques on low-level image features in a sub-space based on principal components analysis (PCA). In the ImageCLEFmed'06 competition [11], we successfully performed retrieval in a medical image collection based on a similarity fusion of different low-level image features [18]. In other general purpose medical CBIR systems, such as in *I²C* [19] or in COBRA [6], the low-level visual features are extracted either from the entire image or from a segmented image region. Although there exists a strong correlation

between the segmented regions and the regions of interest (ROI) in medical images, accurate and semantically valid automatic segmentation is an unsolved problem in image processing and computer vision. Using these low-level features directly without any learning-based classification schemes might also fail to distinguish images of different semantic categories due to limited description power.

Therefore, to enable a content-based search in a heterogenous medical image collection, the retrieval system must be able to recognize the current image class prior to any kinds of post-processing or similarity matching [20,21]. However, the manual classification and annotation of medical images is expensive and time consuming. It also varies from person to person. Hence, the automatic classification of medical images into different imaging modalities or semantic categories is essential to support further queries. So far, automatic categorization in the medical domain is mainly restricted to a specific modality with only a few exceptions [5,20,21]. In Ref. [5], the performances of two medical image categorization architectures with and without a learning scheme are evaluated on 10,322 images of 33 categories based on modality, body part, and orientation with a high accuracy rate of more than 95%. In Ref. [20], a novel similarity matching approach is described for the automatic and semantic-based categorization of diverse medical images, according to their modalities based on a set of visual features, their relevance, and/or generalization for capturing the semantics. In Ref. [21], the automatic categorization of 6231 radiological images into 81 categories is examined by utilizing a combination of low-level global texture features with low-resolution scaled images and a K-nearest-neighbors (KNN) classifier.

A successful categorization and indexing of images would greatly enhance the performance of CBIR systems by filtering out irrelevant images and thereby reducing the search space. As an example, for a query like “Find posteroanterior (PA) chest X-rays with an enlarged heart”, database images at first can be pre-filtered with automatic categorization according to modality (e.g., X-ray), body part (e.g., chest), and orientation (e.g., PA). The latter search could be performed on the pre-filtered set to find the enlarged heart as a distinct visual property. In addition, the automatic classification allows the labeling or annotation of unknown images up to certain axes. For example, a category could denote a code corresponding to an imaging modality, a body part, a direction, and a biological system, in order to organize images in a general way without limitation to a specific modality, such as the IRMA code [21]. Based on the image annotation, semantical retrieval might be performed by applying techniques analogous to the commonly used methods in many successful information retrieval (IR) systems [37]. This simple yet relatively effective solution has not been investigated adequately in the retrieval systems in the medical domain.

Motivated by the considerations above, we present a novel medical image retrieval framework based on image classification by supervised learning, an intermediate level image representation based on category membership scores, feature-level fusion by probabilistic classifier combinations and an adaptive similarity fusion scheme. In this framework, various low-level global, semi-global and low-resolution scale-specific image features

¹ <http://archive.nlm.nih.gov/proj/webmirs/>.

² <http://phobos.imib.rwth-aachen.de/irma/>.

³ <http://www.dim.hcuge.ch/medgift/>.

are extracted, which represent different aspects of an image. Next, the SVM-based classification technique is investigated to associate these low-level features with their high-level semantic categories. The utilization of the probabilistic outputs of multi-class SVMs [26] and the classifier combination rules derived from Bayes's theory [30,31] are explored for the categorization and representation of the images in a feature space based on the probability or membership scores of image categories. It also presents a fusion-based similarity matching technique by using feedback information based on an adaptively weighted linear combination of individual similarity measures. Here, the feedback information is achieved based on predictions of the SVMs about relevant images compared to query image categories. Finally, an inverted file-based indexing scheme commonly used in the text retrieval domain, is also implemented for efficient organization and retrieval.

The rest of the paper is organized as follows. In Section 2, we briefly describe the multi-class classification approach based on the SVMs. Section 3 discusses the low-level feature extraction processes for the generation of the classifier's input. In Section 4, we present feature representation approaches in an intermediate level based on the probabilistic outputs of the SVMs and combination of the outputs derived from Bayes' theorem. In Section 5, a fusion-based similarity matching scheme, and in Section 6, an inverted file-based indexing technique are described. The experiments and the analysis of the results are presented in Sections 7 and 8, respectively, and finally Section 9 provides our conclusion.

2. Multi-class classification with SVMs

SVM is an emerging machine learning technology that has already been successfully used for image classification in both general and medical domain [5,11,17,23]. It performs the classification between two classes by finding a decision surface that is based on the most informative points of the training set [22]. Let $\{\mathbf{x}_1, \dots, \mathbf{x}_i, \dots, \mathbf{x}_N\}$ be a set of training examples that are vectors in space $\mathbf{x}_i \in \Re^d$ with associated labels $y_i \in (+1, -1)^N$. The set of training vectors are linearly separable if there exists a hyperplane for which the positive examples lie on one side and the negative examples on the other. This amounts to finding \mathbf{w} and b so that

$$y_i(\mathbf{w}^t \mathbf{x}_i + b) - 1 \geq 0 \quad \forall i. \quad (1)$$

Among the separating hyperplanes, the one for which the distance to the closest point is maximal is called *optimal separating hyperplane* (OSH). The OSH is found by minimizing $\|\mathbf{w}\|^2$ under constraints (1). If $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_N)$ be the N non-negative Lagrange multipliers associated with constraints (1), the primal form of the objective function is

$$L(\boldsymbol{\alpha}, \mathbf{w}, b) = \frac{1}{2} \mathbf{w}^t \mathbf{w} - \alpha_i(y_i(\mathbf{w}^t \mathbf{x}_i + b) - 1) \\ \text{subject to, } \alpha_i \geq 0 \text{ and } \sum_{i=1}^N y_i \alpha_i = 0. \quad (2)$$

The function $L(\boldsymbol{\alpha}, \mathbf{w}, b)$ is minimized with respect to \mathbf{w}, b and maximized with respect to $\boldsymbol{\alpha}$ and this can be achieved by the use of standard quadratic programming methods. Once the vector $\boldsymbol{\alpha}^0 = (\alpha_1^0, \dots, \alpha_N^0)$ solution of (2) has been found, the general form of the binary linear classification function can be written as [23]

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^N \alpha_i^0 y_i \mathbf{x}_i^t \mathbf{x} + b^0 \right) \quad (3)$$

where the *support vector* are the points for which $\alpha_i^0 > 0$. In the case when the training set is not linearly separable, slack variables ξ_i are defined as the amount by which each \mathbf{x}_i violates (1). Using the slack variables, the new constrained minimization problem becomes:

$$\min_{\mathbf{w}, b, \xi} \frac{1}{2} \mathbf{w}^t \mathbf{w} + C \sum_{i=1}^N \xi_i \quad (4)$$

subject to, $y_i(\mathbf{w}^t \mathbf{x}_i + b) \geq 1 - \xi_i, \quad \xi_i \geq 0 \quad \forall i$

Here C is a penalty term related to misclassification errors.

In SVM training, the global framework for the non-linear case consists in mapping the training data into a high-dimensional space where linear separability will be possible. Here training vectors \mathbf{x}_i are mapped into a high-dimensional Euclidean space by the non-linear mapping function $\phi : \Re^d \rightarrow \Re^h$, where $h > d$ or h could even be infinite. Both the optimization problem and its solution can be represented by the inner product. Hence,

$$\mathbf{x}_i \cdot \mathbf{x}_j \rightarrow \phi(\mathbf{x}_i)^t \phi(\mathbf{x}_j) = K(\mathbf{x}_i, \mathbf{x}_j) \quad (5)$$

where the symmetric function K is referred to as a kernel under Mercer's condition. Under non-linear case, the SVM classification function is given by Vapnik [22]

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^N \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \right) \quad (6)$$

Thus the membership of a test element \mathbf{x} is given by the sign of $f(\mathbf{x})$. Hence, input \mathbf{x} is classified as 1, if $f(\mathbf{x}) \geq 0$, and as -1 otherwise.

The SVMs were originally designed for binary classification problems. However, when dealing with several classes, as in general medical image classification, one need an appropriate multi-class method. As two-class or binary classification problems are much easier to solve, a number of methods have been proposed for its extension to multi-class problems [25,27]. They essentially separate L mutually exclusive classes by solving many two-class problems and combining their predictions in various ways. For example, *one against one* or pairwise coupling (PWC) method [24,26] constructs binary SVMs between all possible pairs of classes. Hence, this method uses $L(L - 1)/2$ binary classifiers for L number of classes, each of which provides a partial decision for classifying a data point. During the testing of a feature \mathbf{x} , each of the $L(L - 1)/2$ classifiers votes for one class. The winning class is the one with the largest number of accumulated votes. On the other hand, *one against the others* method compares a given class with all the others put together. It

basically constructs L hyperplanes where each hyperplane separates one class from the other classes. In this way, it generates L decision functions and an observation \mathbf{x} is mapped to a class with the largest decision function. In Ref. [25], it was shown that the PWC method is more suitable for practical use than the other methods, such as *one against the others*. Hence, we use the *one against one* multi-class classification method [26] based on the LIBSVM [38] tool by combining all pairwise comparisons of binary SVM classifiers.

3. Low-level image feature representation

The performance of a classification or retrieval system depends on the underlying image representation, usually in the form of a feature vector. Numerous low-level features (e.g., color, texture, shape, etc.) are described in the existing literature both for general and domain specific systems [1]. Most systems utilize the low-level visual features without any semantic interpretation of the images. However, in a heterogeneous medical image collection the semantic categories of which are reasonably well defined, it might be possible to utilize the low-level features to depict the semantic contents of each image with a learning-based classifier. To generate the feature vector as an initial input to the classification system, low-level color, texture and edge specific features are extracted from different levels of image representation. Based on previous experiments [18], we have found that the image features at different levels are complementary in nature. Together they can contribute to distinguishing the images of different categories effectively.

In the present work, the MPEG (Moving Picture Experts Group)-7 based Edge Histogram Descriptor (EHD) and Color Layout Descriptor (CLD) are extracted for image representation at the global level [28]. 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 feature vector, \mathbf{x}_{ehd} having a dimension of 80 [28].

The CLD represents the spatial layout of the images in a very compact form [28]. Although CLD is created for color images, we experimentally found it equally suitable for gray-level images (such as images in our collection) with proper

choice of coefficients. It is obtained by applying the discrete cosine transformation (DCT) on the two-dimensional 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 three channels is averaged separately for the 8×8 image blocks. The scalable representation of CLD is allowed in the MPEG-7 standard format. So, we select the number of coefficients to use from each channel of the DCT output. In the present research, a CLD with only 64 Y is extracted to form a 64-dimensional feature vector \mathbf{x}_{cld} since the collection contains only grey-level images.

To retain the spatial information, some fixed grid-based image partitioning techniques have been proposed with moderate success in the general CBIR domain [10]. However, an obvious drawback of this approach is that it is sensitive to shifting, scaling, and rotation because images are represented by a set of local properties and the fixed partitioning scheme might not match with the actual semantic partitioning of the objects. On the other hand, this approach might be suitable in the general medical domain as the images from different modalities are generally captured in a fixed viewpoint and shifting or scaling are less frequent than the images in general domain.

Many medical images of different modalities can be distinguished via their texture characteristics [21]. Hence, the texture features are extracted from sub-images based on the grey-level co-occurrence matrix (GLCM) [29]. For this, a simple grid-based approach is used to divide the images into five overlapping sub-images. These sub-images are obtained by first dividing the entire image space into 16 non-overlapping sub-images. From there, four connected sub-images are grouped to generate five different clusters of overlapping sub-regions as shown in Fig. 1. GLCM is defined as a sample of the joint probability density of the gray levels of two pixels separated by a given displacement. Second order moments (such as energy, maximum probability, entropy, contrast and inverse difference moment) are measured based on the GLCM. These moments are normalized and combined to form a 5-dimensional feature vector for each sub-region and finally concatenated to form a 25-dimensional (5×5) texture moment-based feature vector, $\mathbf{x}_{\text{t-moment}}$.

Images in the database may vary in size for different modalities or within the same modality and may undergo translations. Resizing them into a thumbnail of a fixed size can reduce the translational error and some of the noise due to the artifacts

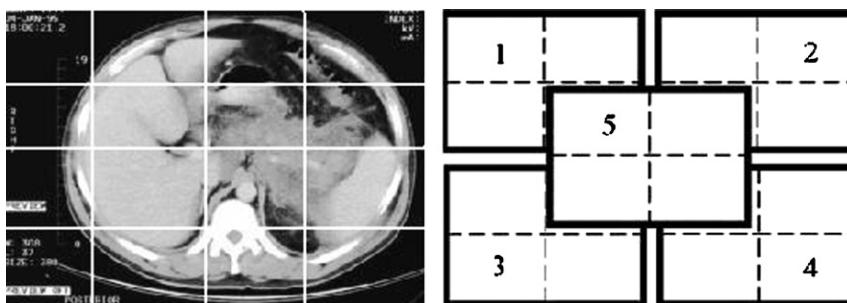


Fig. 1. Region generation from sub-images.

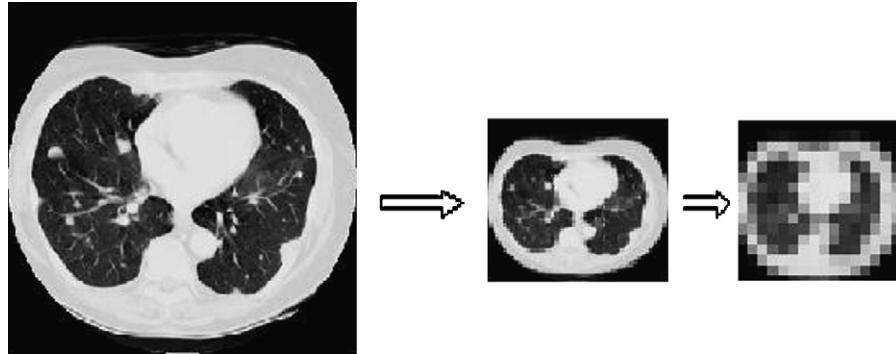


Fig. 2. Feature extraction from scaled image.

present in the images, but it might also introduce distortions. This approach is extensively used in face or fingerprint recognition and has proven to be effective. We have used a similar approach for the feature extraction from the low-resolution scaled images where each image is converted to a gray-level image (one channel only) and scaled down to the size 64×64 regardless of the original aspect ratio. Next, the down-scaled image is partitioned further with a 16×16 grid to form small blocks of (4×4) pixels. The average gray value of each block is measured and concatenated to form a 256-dimensional feature vector, $\mathbf{x}_{\text{avg-gray}}$. By measuring the average gray value of each block we can partially cope with global or local image deformations and can add robustness with respect to translations and intensity changes. An example of this approach is shown in Fig. 2 where the left image is the original one, the middle image is the down-scaled version (64×64 pixels), and the right image shows the average gray values of each block (4×4 pixels). All the above feature descriptors (e.g., \mathbf{x}_{ehd} , \mathbf{x}_{cld} , $\mathbf{x}_{\text{t-moment}}$, and $\mathbf{x}_{\text{avg-gray}}$) are used separately as inputs to the multi-class SVMs for training and categorizing the test images for the annotation and indexing purposes described in the following sections.

4. Feature representation as probabilistic output

This section describes how the low-level feature vectors in Section 3 are converted to a feature vector in an intermediate level based on the probabilistic output of the SVMs. Although the voting procedure for multi-class classification based on *one against one* or pairwise coupling (PWC) method [24,26] requires just pairwise decisions, it predicts only a class label. However, to represent each image with category-specific confidence scores, a probability estimation approach would be useful. A related work in Ref. [27] is also investigated to generate category-specific label vectors for image annotation in general domain. It performs annotation using global features, and it uses a Bayes Point Machines (BPMs)-*One against the others* ensemble to provide multiple semantical labels for an image. However, the main difference between this and our work is that, we extend it further by using probabilistic output-based label vectors directly in similarity and feature level fusion schemes for an effective retrieval purpose instead of performing only image annotation.

In the present work, a probability estimation approach described in Ref. [26] for multi-class classification by PWC

is used. For the SVM-based training, the initial input to the retrieval system is a feature vector set of training images in which each image is manually annotated with a single semantic label selected out of M labels or categories. So, a set of M labels are defined as $\{C_1, C_2, \dots, C_M\}$, where each $C_i, i \in \{1, \dots, M\}$ characterizes the representative semantics of an image category. In the testing stage of the probabilistic classifier, each non-annotated or unknown image is classified against the M categories. The output of the classification produces a ranking of the M categories. Each category would assign a probability (confidence) score to each image. The confidence score represents the weight of a category label in the overall description of an image. The probabilities or confidence scores of the categories form an M -dimensional vector for a feature $\mathbf{x}_m, m \in \{\text{cld}, \text{ehd}, \text{t-moment}, \text{avg-gray}\}$ of image i as follows:

$$\mathbf{p}_i(\mathbf{x}_m) = [p_{i1}(\mathbf{x}_m) \cdots p_{ik}(\mathbf{x}_m) \cdots p_{iM}(\mathbf{x}_m)]^T \quad (7)$$

Here, $p_{ik}(\mathbf{x}_m)$, $1 \leq k \leq M$, denotes the *posterior* probability that an image i belongs to category C_k in terms of input feature vector \mathbf{x}_m . Finally, an image i belongs to a category $C_l, l \in \{1, \dots, M\}$ using feature vector \mathbf{x}_m where the category label is determined by

$$l = \operatorname{argmax}_k [p_{ik}(\mathbf{x}_m)] \quad (8)$$

that is the label of the category with the maximum probability score. In this context, given the feature vector \mathbf{x}_m , the goal is to estimate

$$p_k = P(y = k | \mathbf{x}_m), \quad k = 1, \dots, M \quad (9)$$

To simplify the presentation, we drop the terms i and \mathbf{x}_m in $p_{ik}(\mathbf{x}_m)$.

Following the setting of the *one against one* (i.e., PWC) approach for multi-class classification, the pairwise class probabilities r_{kj} are estimated as an approximation of μ_{kj} as [26]

$$r_{kj} \approx P(y = k | y = k \text{ or } j, \mathbf{x}_m) \approx \frac{1}{1 + e^{A\hat{f} + B}} \quad (10)$$

where A and B are the parameters estimated by minimizing the negative log-likelihood function, and \hat{f} are the decision values of the training data (fivefold cross-validation (CV) is used to form an unbiased training set). Finally, p_k is obtained from all these r_{kj} 's by solving the following optimization problem based

on the second approach in Ref. [26]:

$$\begin{aligned} \min_{\mathbf{p}} & \frac{1}{2} \sum_{k=1}^M \sum_{j:j \neq k} (r_{jk} p_k - r_{kj} p_j)^2 \\ \text{subject to } & \sum_{k=1}^M p_k = 1, p_k \geq 0 \quad \forall k. \end{aligned} \quad (11)$$

where \mathbf{p} (e.g., $\mathbf{p}_i(\mathbf{x}_m)$) is a M -dimensional vector of multi-class probability estimates. See [26] for detailed implementation of the solution.

Similarly, the label vector of a query image q can be found online by applying its feature descriptors as inputs at different levels to the SVM classifiers as

$$\mathbf{p}_q(\mathbf{x}_m) = [p_{q1}(\mathbf{x}_m) \cdots p_{qk}(\mathbf{x}_m) \cdots p_{qM}(\mathbf{x}_m)]^T \quad (12)$$

In the vector-space model of IR [37], one common measure of similarity is the cosine of the angle between the query and document vectors. In many cases, the direction or angle of the vectors are a more reliable indication of the semantic similarities of the objects than the Euclidean distance between the objects in the term-document space [37]. The proposed feature representation scheme closely resembles the document representation where a category is analogous to a keyword. Hence, we adopt the cosine similarity measure between feature vectors of query image q and database image i for a particular feature input m as follows:

$$S_m(q, i) = \frac{\sum_{k=1}^M p_{qk}(\mathbf{x}_m) p_{ik}(\mathbf{x}_m)}{\sqrt{\sum_{k=1}^M (p_{qk}(\mathbf{x}_m))^2} \sqrt{\sum_{k=1}^M (p_{ik}(\mathbf{x}_m))^2}} \quad (13)$$

4.1. Feature level fusion as 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. The combination of ensembles of classifiers has been studied intensively and evaluated on various image classification data sets involving the classification of digits, faces, photographs, etc. [30,32–34]. It has been realized that combination approaches can be more robust and more accurate than the systems using a single classifier alone.

In general, a classifier combination is defined as the instances of the classifiers with different structures trained on the dis-

tinct feature spaces [30,32]. In the present research, we consider combination strategies of the SVMs with different low-level features as inputs, based on five fundamental classifier combination rules derived from Bayes's theory [30]. These combination rules, namely product, sum, max, min, and median, and the relations among them have been theoretically analyzed in depth in Ref. [30]. These rules are simple to use but require that the classifiers output posterior probabilities of classification. This is exactly the kind of output the probabilistic SVMs produce as described in Section 4. Let us assume that we have m classifiers as experts, each representing the given image i with a distinct feature vector. Hence, the m -th classifier utilizes the feature vector \mathbf{x}_m as input for initial training. Each classifier measures the posterior probability $p_i(C_k|\mathbf{x}_m)$ of i belonging to class C_k , $k \in \{1, \dots, M\}$ using feature vector \mathbf{x}_m . Here, the posterior probability $p_i(C_k|\mathbf{x}_m)$ is equivalent to $p_{ik}(\mathbf{x}_m)$, which also denotes the probability that an image i belongs to class C_k . In these combination rules, a priori probabilities $P(C_k)$ are assumed to be equal. The decision rules for the product, sum, max, min and median are made by using the following formulas in terms of the *a posteriori* probabilities yielded by the respective classifiers as

$$l = \operatorname{argmax}_k [p_{ik}^r], r \in \{\text{prod,sum, max, min ,med}\} \quad (14)$$

where for the product rule

$$p_{ik}^{\text{prod}} = P^{-(m-1)}(C_k) \prod_m p_i(C_k|\mathbf{x}_m) \quad (15)$$

similarly, for the sum, max, min and median rules

$$p_{ik}^{\text{sum}} = (1-m)P(C_k) + \sum_m p_i(C_k|\mathbf{x}_m) \quad (16)$$

$$p_{ik}^{\text{max}} = (1-m)P(C_k) + \max_m p_i(C_k|\mathbf{x}_m) \quad (17)$$

$$p_{ik}^{\text{min}} = P^{-(m-1)}(C_k) + \min_m p_i(C_k|\mathbf{x}_m) \quad (18)$$

and

$$p_{ik}^{\text{med}} = \frac{1}{m} \sum_m p_i(C_k|\mathbf{x}_m) \quad (19)$$

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 [30]. Classifier combination based on these two rules often performs better than the other rules, such as min, max and median [30,18].

The SVMs with different feature descriptors are combined with the above rules based on Eqs. (15)–(19) and finally represent an image i as an M -dimensional feature vector of confidence scores as

$$\mathbf{p}_i^r = [\hat{p}_{i1}^r \cdots \hat{p}_{ik}^r \cdots \hat{p}_{iM}^r]^T \quad (20)$$

Here, the element $\hat{p}_{ik}^r = (p_{ik}^r) / (\sum_{k=1}^M p_{ik}^r)$, $1 \leq k \leq M$ denotes the normalized membership score according to which an image i belongs to class C_k in terms of the combination rule $r \in \{\text{prod,sum, max, min ,med}\}$, where $\sum_{k=1}^M \hat{p}_{ik}^r = 1$ and $\hat{p}_{ik}^r \geq 0$.

5. Similarity fusion with adaptive weighting

It is difficult to find a unique feature representation or distance function to compare images accurately for all types of queries. In other words, different feature representations might be complementary in nature and will have their own limitations. In Section 4, we represented images in a feature space based on the probability score of each category. For each low-level feature descriptor \mathbf{x}_{i_m} of image i as input to the multi-class SVMs, we obtained $\mathbf{p}_i(\mathbf{x}_m)$ as the new descriptors in the probabilistic feature space. Since these feature descriptors were generated by different low-level input features, the question is how to combine them in a similarity matching function for retrieval purposes. The most commonly used approach in this direction is the linear combination. In this model, the similarity between a query image q and a database image i is defined as

$$S(q, i) = \sum_m \alpha_m S_m(q, i) \quad (21)$$

where the α_m are certain weights for different similarity matching functions and subject to $0 \leq \alpha_m \leq 1$, $\sum \alpha_m = 1$. The effectiveness of the linear combination depends mainly on the choice of weights α_m . The weights may be set manually according to prior knowledge, learned off-line based on the optimization techniques, or may be adjusted on-line by user feedback [35,36].

The present section describes a simple and effective method of weight adjustment based on both the first and the third options. Since the training data is available, we would like to take advantage of it. At the beginning, each individual similarity measure $S_m(q, i)$ is weighted by using the normalized value based on the best k -fold cross-validation (CV) accuracy of the associated feature space obtained while training the SVMs. The CV is an established technique for estimating the accuracy of a classifier, where a pool of labeled data is partitioned into k equally sized subsets. Each subset is used as a test set for a classifier trained on the remaining $k - 1$ subsets. The final value of the accuracy is given by the average of the accuracies of these k classifiers [24]. After the initial retrieval result, the weights are adjusted on-line for each query, based on the SVMs prediction as feedback of the relevant images on the top K retrieved images. If the predicted category label of an image matches with the category of the query image, then it is considered as a relevant image. Based on the relevant information obtained from the SVMs prediction, the performance or effectiveness of each feature space $\mathbf{p}(\mathbf{x}_m)$ on each query is calculated by using the following formula:

$$E(\mathbf{p}(\mathbf{x}_m)) = \frac{\sum_{i=1}^K \text{Rank}(i)}{K/2} P(K) \quad (22)$$

where $\text{Rank}(i) = 0$, if the image in the rank position i is not relevant based on the user's feedback and $\text{Rank}(i) = (K - i)/(K - 1)$ for the relevant images. Hence, the function $\text{Rank}(i)$ monotonically decreases from one (if the image at rank position 1 is relevant) down to zero (e.g., for a relevant image at rank position K). On the other hand, $P(K) = R_K/K$ is the precision at

the top K , where R_k is the number of relevant images in the top K retrieved result. Hence, the Eq. (22) is basically the product of two factors, rank order and precision. The rank order factor takes into account the position of the relevant images in the retrieval set, whereas the precision is a measure of the retrieval accuracy, regardless of the position. Generally, the rank order factor is heavily biased by the position in the ranked list over the total number of relevant images, and the precision value totally ignores the rank order of the images. To balance both criteria, we use a performance measure that is the product of the rank order factor and precision. If there is more overlap between the relevant images of a particular retrieval set and the one from which a user provides the feedback, then the performance score will be higher. Both terms on the right side of Eq. (22) will be 1, if all the top K returned images are considered as relevant. The raw performance scores obtained by the procedure above are then normalized by the total score as $\hat{E}(\mathbf{p}(\mathbf{x}_m)) = \hat{\alpha}_m = E(\mathbf{p}(\mathbf{x}_m))/(\sum_m E(\mathbf{p}(\mathbf{x}_m)))$ to yield numbers in $[0, 1]$ where $\sum_m \hat{E}(\mathbf{p}(\mathbf{x}_m)) = 1$. After the individual similarity measures of each representation are determined in the previous iteration for query q and for target image i , we can linearly combine them into a single similarity matching function as follows:

$$S(q, i) = \sum_m S_m(q, i) = \sum_m \hat{\alpha}_m S(\mathbf{p}_q(\mathbf{x}_m), \mathbf{p}_i(\mathbf{x}_m)) \quad (23)$$

where $\sum_m \hat{\alpha}_m = 1$. Thus, the steps involved in this process are as follows:

- Step 1: Initially, consider the top K images by applying similarity fusion $S(q, i) = \sum_m \alpha_m S_m(q, i)$ based on the normalized CV accuracy.
- Step 2: For each top retrieved image $j \in K$, determine its category label as $C_k(j)$, $k \in \{1, \dots, M\}$ by using Eq. (14) of any of the combination rules $r \in \{\text{prod}, \text{sum}, \text{max}, \text{min}, \text{med}\}$.
- Step 3: Consider image j as relevant to query q , if $(C_k(j) = C_k(q))$, e.g., j and q are in the same category.
- Step 4: For each ranked list based on individual similarity matching, consider top K images and measure the performance as $E(\mathbf{p}_q(\mathbf{x}_m))$ by utilizing Eq. (22).
- Step 5: Finally, utilize the updated weight-based similarity fusion of Eq. (23) for rank-based retrieval.
- Step 6: Continue steps 2–5 until no changes are noticed, i.e., the system converges.

The main idea of this algorithm is to give more weight to the similarity matching functions that are more consistent across the example set chosen by the user. There is a trade-off between the automatic and the interactive weight updating approaches. For the first one, the user's semantic perception can be incorporated directly into the system. However, it might take longer to select the relevant images at each iteration and user might not provide enough feedback to perform the system better. On the other hand, the automated method depends solely on the classification accuracy without any user involvement. Hence, it can execute faster compared to the interactive feedback method. However, if the prediction goes wrong for a query image at first,

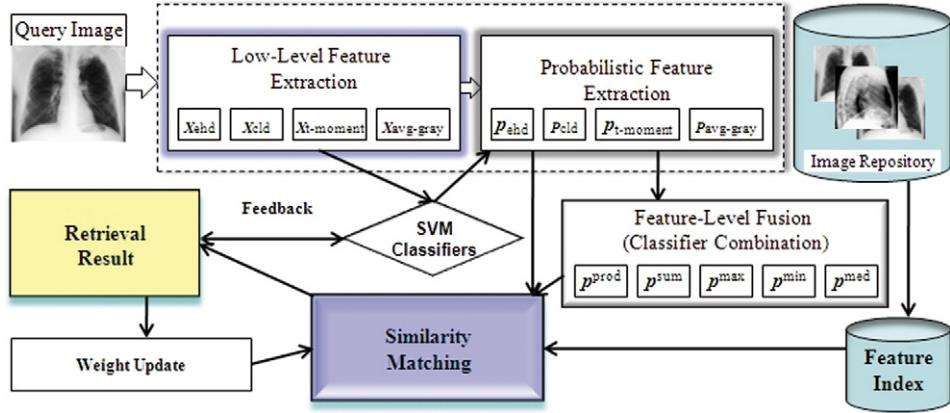


Fig. 3. Block diagram of the retrieval framework.

then the performance worsens from the previous iteration due to the incorrect selection of relevant images. A similar kind of problem is observed in the cases of pseudo or blind relevance feedback in information retrieval [37]. The block diagram of the proposed image retrieval framework based on the feature and the similarity-level fusion is shown in Fig. 3 from a query image perspective.

6. Inverted file-based indexing

The majority of CBIR systems in the medical as well as in the general domain typically focus on the quality of the feature representations and the similarity matching functions with little or no emphasis on the efficiency of the retrieval. In large database applications, the indexing strategies, i.e., pre-filtering, are required to avoid exhaustive searches in the entire collection. Some multi-dimensional tree-based and cluster-based indexing structures have been proposed for general domain image indexing [10]. However, the accuracy of these indexing techniques largely depends on the feature dimensionality, which degrades rapidly as the dimension increases.

In the present paper, an inverted file-based indexing scheme [37], commonly used in the text retrieval domain, is implemented by exploiting the discrete nature of the probabilistic feature space. A similar technique is also proposed in the medGIFT system [16], where an inverted file contains an entry for each possible low-level feature and a list of the images containing that feature. However, the features that are used are at a much lower semantic level than the terms or keywords in the text documents. The inverted file might be used efficiently if the image representation closely resembles the keyword-based representation of the documents. In the proposed image representation schemes, each category or label can be considered as a keyword, where the weight of the keyword is the probability or the membership score for that category. However, there is a subtle difference between this representation and that for the documents. Instead of obtaining a sparse representation where many keywords contain a zero weight for a particular document out of all possible keywords, the proposed feature vector is continuous with a non-zero weight in each category due to the nature of the probabilistic algorithm. Although the semantic behind this

is very straightforward. A high probability score for a category gives a high probability of an image being in that particular category and vice-versa. Hence, each category of the images the weight of which is below a certain threshold value τ (that is obtained based on experimental result on the training samples) is transformed to zero, and the rest of the categories (keywords) are considered only in the indexing process to generate the inverted file. In other words, for each category or label, the pointers or references to images that have the category weight above the threshold value τ are stored in a list and the rest are ignored. If the query image q is represented as $\mathbf{p}_q(\mathbf{x}_m)$ based on (12) or \mathbf{p}_q^r based on (20), then the images that appear in the lists of the categories with a probability score $p_{qk}(\mathbf{x}_m) \geq \tau$ or $\hat{p}_{qk}^r \geq \tau$ for $k \in \{1, \dots, M\}$ are regarded as candidates for further calculations. Finally, only these candidate images are compared with the query image using the cosine similarity measure and are sorted to provide the ranked-based retrieval result. In this way, the response time is reduced at the expense of the storage whereas the retrieval accuracy is still maintained as shown in Section 7. Generally, this search algorithm is very efficient in terms of the time complexity. Most of the time, only very few categories or labels are qualified (based on the τ value) to search in the inverted list or posting files. However, there is always a trade-off between the time and the space complexities in the inverted file because the inverted lists grow larger for bigger collections with additional space requirements in exchange for faster access time.

7. Experiments

To measure the effectiveness and efficiency of the proposed image retrieval framework, exhaustive experiments were performed in a diverse medical image collection under the ImageCLEFmed [11] benchmark for automatic annotation. This collection contains 11,000 radiographs grouped into 116 categories. It was made available by the IRMA group from the University Hospital, Aachen, Germany [15]. The images in the collection are in grey level and in PNG (Portable Network Graphics) format. All the images are classified manually by reference coding with respect to a mono-hierarchical coding scheme [15]. In this scheme, the technical code (T) describes



Fig. 4. Intra-class variability within the category label 14.

the imaging modality or technique used, the directional code (D) denotes the body orientation, the anatomical code (A) refers to the body region examined, and the biological code (B) describes the biological system examined. The entire code results in a character string of not more than 13 characters (TTTT-DDD-AAA-BBB). Based on this code, 116 distinct categories are defined. The images have a high intra-class variability and inter-class similarity, which make the classification and retrieval task more difficult. For example, Fig. 4 shows that a great deal of intra-class variability exists in images of category label 14, which is encoded as “1121-120-413-7**” in IRMA code and annotated as “X-ray, plain radiography, analog, overview image, coronal, anteroposterior (AP, coronal), upper extremity (arm), hand, carpal bones, musculoskeletal system”, mostly due to the illumination changes and to small amounts of position and scale differences. On the other hand, Fig. 5 exhibits an example of inter-class similarities between two different categories (e.g., 5 and 52). Images in the upper row of Fig. 5 belong to category label 52 (“1121-127- 700-500” in IRMA code) and are annotated as “X-ray, plain radiography, analog, overview image-coronal, anteroposterior (AP, coronal), supine-abdomen-uropoietic system”, whereas the images in the lower row belong to category label 5 (“1121-115-700-400” in IRMA code) and are annotated as “X-ray, plain radiography, analog, overview image-coronal, posteroanterior (PA), upright-abdomen-gastrointestinal

system”. Although the images in both categories are hard to distinguish with the untrained eye, they basically differ in the orientation and the biological systems.

7.1. Training

For the training of the multi-class SVMs, 10,000 images are used as the training set. The remaining 1000 images are used as the test set, which conforms to the experimental setting of ImageCLEFmed’06 [11]. Fig. 6 shows the number of images in each category in the training set. The categories in the database are not uniformly distributed (for example, category 111 has 1927 images, whereas four categories have only 10 images), making the training and consequently the classification task more difficult. The test set is used to measure the error rate of the classification systems and to generate the feature vectors for evaluating the precision–recall of the retrieval approach. 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 . It is not known beforehand which values of C and γ are the best for the classification problem at hand. Hence, a fivefold cross-validation (CV) is conducted, where various pairs of (C, γ) are tried and the one with the lowest CV error rate is picked. The best values for the parameters C and γ that are obtained for



Fig. 5. Inter-class similarity (the images in the top row belong to category 52 and the images in the bottom row belong to category 5).

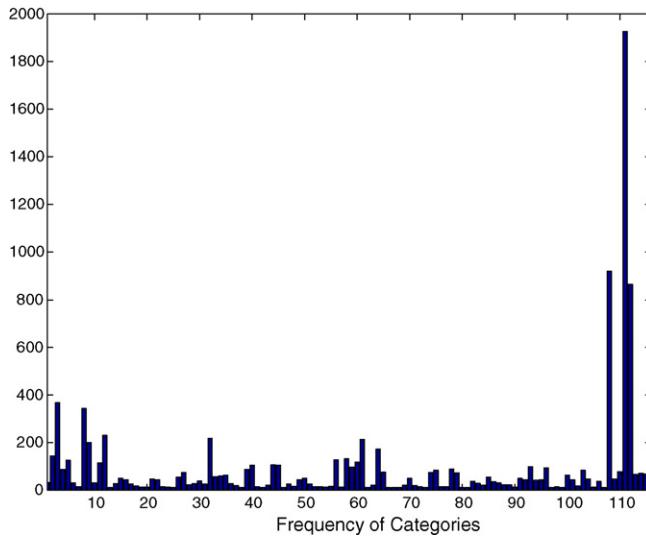


Fig. 6. Frequency of images in each category in the training set.

Table 1
CV (fivefold) error rate (training set)

Image feature	Kernel	C	γ	Error rate (%)
x_{ehd}	RBF	100	.005	22.1
x_{cld}	RBF	10	.025	24.5
$x_{t\text{-moment}}$	RBF	20	.05	23.4
$x_{avg\text{-gray}}$	RBF	20	.05	26.4

the different feature representations are shown in Table 1. After finding the best values for the parameters C and γ , the values are used to train the set to generate the SVM model files. We utilize the LIBSVM software package [38] for the implementation of the SVMs.

8. Results

To compare the performances of individual classifiers based on distinct feature inputs and combined classifiers based on the combination rules, we measure the error rate of the test image set. The error rate in the test set for the individual SVMs with low-level features as inputs is shown in Table 2. We observe that the lowest error rate (e.g., 25.5%) is achieved with the EHD feature (e.g., x_{ehd}) as input. Hence, it conforms to the importance of edge structure in diverse medical images. However, when we applied the classifier combination rules involving the SVMs with different features as the input, lower error rates were observed in three out of five combination rules as shown in Table 3. The classifier combination improves the results in some cases because each of the single SVM classifier evaluates different aspects of

Table 2
Error rate of the individual classifiers (test set)

SVM (x_{ehd})	25.5%
SVM (x_{cld})	26.2%
SVM ($x_{t\text{-moment}}$)	25.7%
SVM ($x_{avg\text{-gray}}$)	26.1%

Table 3
Error rate of different classifier combinations (test set)

Prod	24.8%
Sum	24.1%
Max	26.1%
Min	26.5%
Med	25.2%

the image representation. The error rates achieved by the best seven groups in ImageCLEFmed’06 competition for the automatic annotation task based on the same image collection and experimental setting (e.g., 10,000 training images and 1000 test images) is shown in Table 4. We have submitted our runs based on the group name of “CINDI” and achieved an error rate of 24.1% for our best run (e.g., sum rule) as shown in Table 4[18]. Although the main motivation of the present paper is not about how to improve classification accuracy. Instead, it is about how to utilize the probabilistic classification information effectively for retrieval based on feature and similarity level fusion. If we utilize more low-level features as inputs for the classifier training, classification accuracy would obviously be improved as shown in ImageCLEFmed [11] automatic annotation results.

For a quantitative evaluation of the retrieval results, we selected all the images in the test set as query images and used “Query-by-Example” query-by-example as the search method, where the query is specified by providing an example image to the system. A retrieved image is considered as a correct match if it is in the same category (based on the ground truth) as the query image. Fig. 7 presents the precision-recall curves of the different feature spaces. Fig. 7(a) shows the precision-recall curves based on the cosine similarity matching in the proposed probabilistic feature spaces and the Euclidean similarity measures for the low-level feature spaces. As shown in Fig. 7(a), the better performances in terms of precision are always achieved when the search is performed in the proposed probabilistic feature spaces. There is a clearly visible large gap in the performances of low-level features when compared to the probabilistic feature spaces. Such results are expected since the proposed feature space retains better information related to the semantic categories as compared to the low-level features generally used in many CBIR systems. Fig. 7(b) shows the precision-recall curves for the feature spaces based on the five classifier combination rules. Like the improved classification accuracies in three out of five cases, here also we can observe improved performances in terms of precision-recall in all three cases. Hence,

Table 4
Error rate in ImageCLEFmed’06 evaluation (group wise)

Group	Runtag	Error rate (%)
RWTHi6	SHME	16.2
UFR	UFR-ns-1000-20 × 20 × 10	16.7
MedIC-CISMeF	Local + global-PCA335	17.2
MSRA	WSM-msra-wsm-gray	17.6
RWTHmi	Opt	21.5
CINDI	Cindi-svm-sum	24.1
OHSU	OHSU-iconGLCM2-tr	26.3

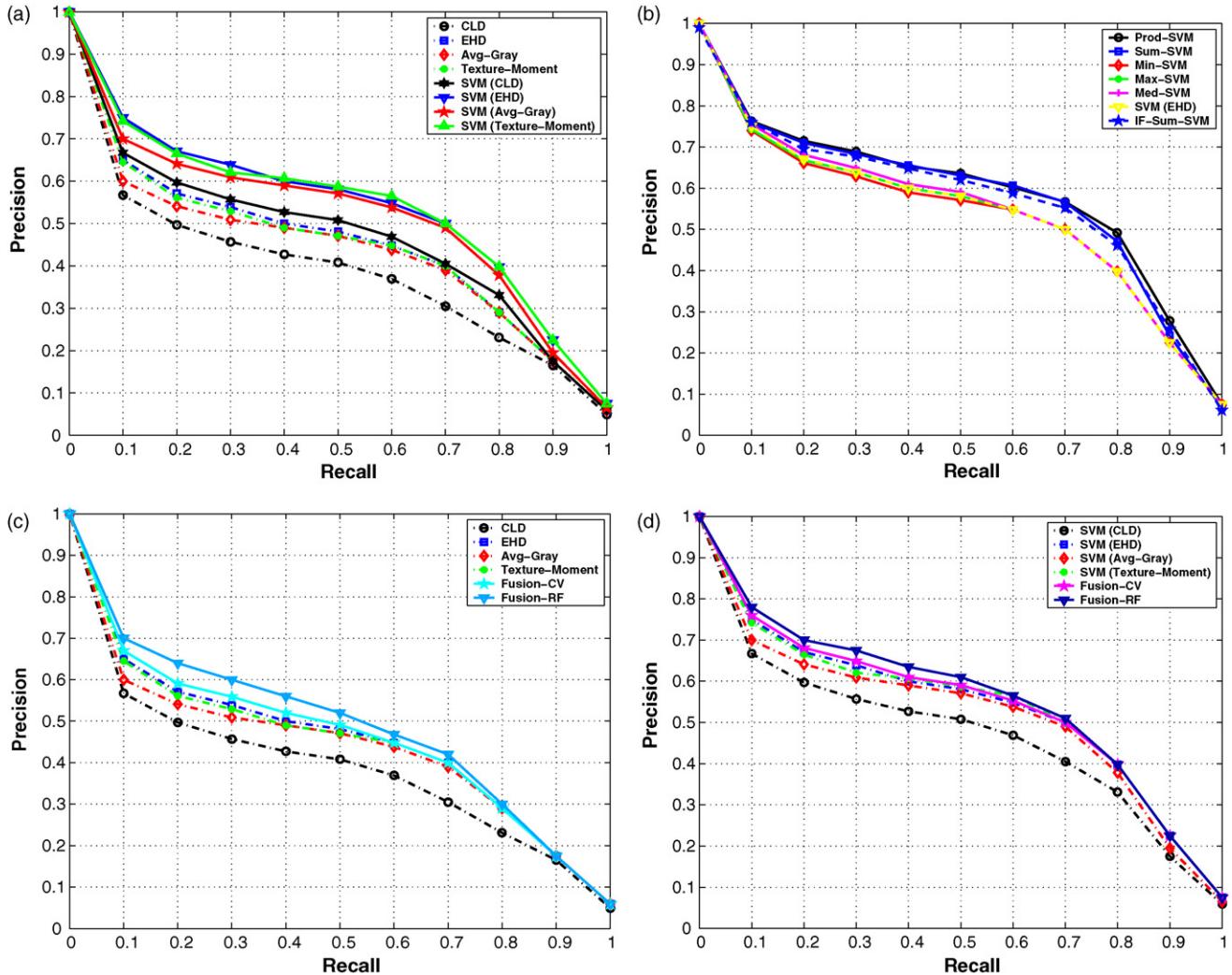


Fig. 7. Precision–recall curves. (a) Comparison in case of low-level and SVM generated feature spaces. (b) Comparison in case of feature spaces generated by combination rules. (c) Comparison in case of similarity fusion and Euclidean similarities in individual low-level feature spaces. (d) Comparison in case of similarity and cosine similarities in individual SVM generated feature spaces.

we can observe that the retrieval performance is linearly dependent on the classification accuracy. Moreover, the performances are almost identical with the inverted file (based on the threshold value $\tau = 0.15$) in the feature space obtained by the sum combination rule compared to the linear search on the same feature space. To test the efficiency of the indexing scheme, we also compare the average retrieval time (in ms) with and without indexing schemes for the query set. It is clear from Table 5 that the search is about four times faster with the indexing scheme compared to the linear search in the test set. Fig. 7(c) shows improvement in the performance with the similarity fusion (one with CV accuracies as weights (e.g., Fusion-CV) and the other with the adaptive weights (e.g., Fusion-RF) compared to the Euclidean similarities in the low-level feature spaces. For the dynamic weight adjust-

ment, the top $K = 20$ images, and two iterations of feedback information as classifier’s prediction based on the sum rule (e.g., the best performed combination approach) are considered for experimental purposes. Similarly, we can observe improved performances in probabilistic feature spaces as shown in Fig. 7(d). Hence, the results justify the assumption of the complementary nature of the feature spaces and the dynamic weight update requirement for individual searches.

For a qualitative evaluation of the performance improvement, Figs. 8 and 9 show the snapshots of the CBIR retrieval interface for a query image. In Fig. 8, for a query image (the image in the left panel) that belongs to the category label 111 (e.g., X-ray, plain radiography, analog, high beam energy, coronal, anteroposterior, supine, and chest), the system returns 5 images of the same category out of the top 10 images by applying the Euclidean similarity measure on the EHD. The five relevant images are located in rank position 1, 3, 5, 7 and 9 where ranking goes from left to right and from top to bottom. By contrast, in Fig. 9, the system returns 7 (in position 1–4, 6, 8, and 9) images from the same category by applying

Table 5
Retrieval time with and without indexing schemes

Linear search (ms)	442
Inverted index (ms)	119

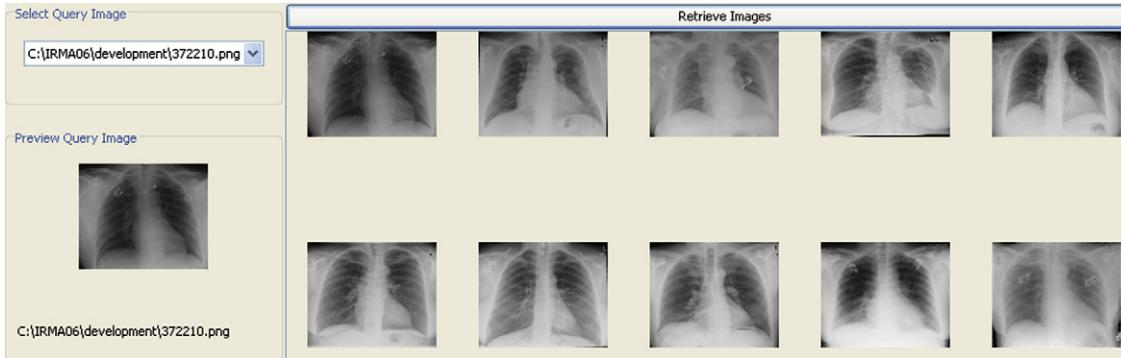


Fig. 8. A snapshot of the image retrieval based on EHD (x_{ehd}) feature.

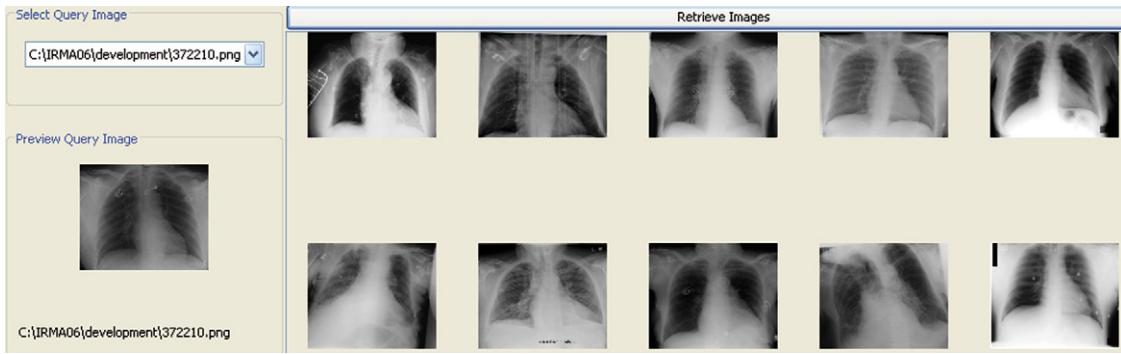


Fig. 9. A snapshot of the image retrieval based on SVM-EHD ($p(x_{ehd})$) feature.

the cosine similarity on the feature space generated by probabilistic output of SVMs when EHD is used as input feature. In both cases, the irrelevant returned images are from category label 108 (e.g., X-ray, plain radiography, analog, high beam energy, coronal, posteroanterior (PA), and chest), where the main difference between these two categories is in the orientation (e.g., anteroposterior (AP) vs. posteroanterior (PA)). There is thus clear improvement in the performance in the probabilistic feature space for this particular query in terms of finding the correct categories in the proposed image retrieval framework.

9. Summary

In the present paper, a novel image retrieval framework based on feature and similarity level fusions is proposed for diverse medical image collections. We explore the utilization of probabilistic output of multi-class SVMs and various classifier combination rules in different aspects of the image feature spaces for the representation and the feature level fusion. Hence, the images are represented in a feature space based on the probabilistic outputs of the multi-class SVMs and the outputs based on several classifier combination rules. This technique eventually provides us with a basic semantic knowledge about the image and serves as a semantical descriptor for a more meaningful image retrieval by narrowing down the semantical gap that is prevalent in current CBIR systems. Since different feature spaces are generally complementary in nature, we also exploit this observation in a similarity-level fusion scheme based on

prior knowledge and a dynamic weight adjustment by using the feedback of the SVMs.

The experiments and analysis of the results are performed on a diverse medical collection. The results indicate that the proposed probabilistic feature spaces are effective in terms of precision–recall compared to low-level features. It is shown that the feature vectors from the different representations contain valuable information due to their complementary nature. They could be fused together by combining the classifiers or similarity matching functions for the improvement of accuracy. The proposed retrieval approach also finds relevant images effectively and efficiently based on the similarity level fusion and inverted file-based indexing in the semantical feature space. The analysis of the precision–recall curves in different feature spaces confirms the improved performance of the proposed method. Overall, this framework will be useful as a front-end for medical databases where the search can be performed in diverse images for teaching, training, and research purposes. Later, it might be extended for diagnostic purposes by selecting the appropriate parameters and matching functions in the category-specific search process.

Acknowledgments

This work was partially supported by IDEAS, NSERC, and Canada Research Chair grants. We would like to thank TM Lehmann, Department of Medical Informatics, RWTH Aachen, Germany [15], for making the database available for the experiments and C.C. Chang and C.J. Lin for the LIBSVM software

tool [38] that is used for the SVM-related experiments. We would also like to thank the anonymous reviewer for the valuable and constructive comments that helped us improve the presentation.

References

- [1] Müller H, Michoux N, Bandon D, Geissbuhler A. A review of content-based image retrieval systems in medical applications clinical benefits and future directions. *Int J Med Inform* 2004;73:1–23.
- [2] Wong TC. Medical image databases. New York: Springer-Verlag; 1998.
- [3] Tang LHY, Hanka R, Ip HSS. A review of intelligent content-based indexing and browsing of medical images. *J Health Inform* 1999;5:40–9.
- [4] Tagare HD, Jaffe CC, Duncan J. Medical image databases: a content-based retrieval approach. *J Am Med Informat Assoc* 1997;4:184–98.
- [5] Florea F, Müller H, Rogozan A, Geissbuhler A, Darmoni S. Medical image categorization with Medic and MedGIFT. *Proc Med Inform Europe (MIE 2006)*. p. 3–11.
- [6] Lehmann TM, Güld MO, Thies C, Fischer B, Keysers D, Kohnen M. Content-based image retrieval in medical applications for picture archiving and communication systems. *Proc SPIE* 2003;5033:109–17.
- [7] El-Kwae EA, Xu H, Kabuka MR. Content-based retrieval in picture archiving and communication systems. *IEEE Trans Knowledge Data Eng* 2000;13(2):70–81.
- [8] Güld MO, Kohnen M, Schubert H, Wein BB, Lehmann TM. Quality of DICOM header information for image categorization. *Proc SPIE* 2002;4685:280–7.
- [9] Smeulders AWM, Worring M, Santini S, Gupta A, Jain R. Content-based image retrieval at the end of the early years. *IEEE Trans Pattern Anal Mach Intell* 2000;22(12):1349–80.
- [10] Rui Y, Huang TS, Chang SF. Image retrieval: current techniques, promising directions and open issues. *J Vis Commun Image Rep* 1999;10(4):39–62.
- [11] Müller H, Deselaers T, Lehmann TM, Clough P, Kim E, Hersh W. Overview of the ImageCLEFmed 2006 medical retrieval and annotation Tasks. In: Evaluation of multilingual and multi-modal information retrieval—seventh workshop of the cross-language evaluation forum (CLEF 2006); 2007. *Proc LNCS* 2006;4730:595–608.
- [12] Shyu CR, Brodley CE, Kak AC, Kosaka A, Aisen AM, Broderick LS. ASSERT: a physician-in-the-loop content-based image retrieval system for HRCT image databases. *Comput Vis Image Understand* 1999;75:111–32.
- [13] Long LR, Thoma GR. Landmarking and feature localization in spine X-rays. *J Elect Imaging* 2001;10(4):939–56.
- [14] Tang LHY, Hanka R, Ip HHS, Lam R. Extraction of semantic features of histological images for content-based retrieval of images. *Proc SPIE* 2000;3662:360–8.
- [15] Lehmann TM, Wein BB, Dahmen J, Bredno J, Vogelsang F, Kohnen M. Content-based image retrieval in medical applications—a novel multi-step approach. *Proc SPIE* 2000;3972:312–20.
- [16] Müller H, Rosset A, Vallee J, Geissbuhler A. Integrating content-based visual access methods into a medical case database. *Proc Med Inform Europe (MIE 2003)*; 2003. p. 480–5.
- [17] Rahman MM, Bhattacharya P, Desai BC. A framework for medical image retrieval using machine learning & statistical similarity matching techniques with relevance feedback. *IEEE Trans Inform Tech Biomed* 2007;11(1):59–69.
- [18] Rahman MM, Sood V, Desai BC, Bhattacharya P. CINDI at ImageCLEF 2006: image retrieval & annotation tasks for the general photographic and medical image collections. In: Evaluation of multilingual and multi-modal information retrieval—seventh workshop of the cross-language evaluation forum (CLEF 2006); 2007. *Proc LNCS* 2006;4730:715–24.
- [19] Orphanoudakis SC, Chronaki C, Kostomanolakis S. I^2C —a system for the indexing, storage and retrieval of medical images by content. *Med Inform* 1994;19(2):109–22.
- [20] Mojsilovic A, Gomes J. Semantic based image categorization, browsing and retrieval in medical image databases. *Proc IEEE Int Conf Image Process* 2002;3:145–8.
- [21] Lehmann TM, Güld MO, Deselaers T, Keysers D, Schubert H, Spitzer K. Automatic categorization of medical images for content-based retrieval and data mining. *Comput Med Imaging Graph* 2005;29:143–55.
- [22] Vapnik V. Statistical learning theory. New York, NY: Wiley; 1998.
- [23] Chapelle O, Haffner P, Vapnik V. SVMs for histogram-based image classification. *IEEE Trans Neural Networks* 1999;10(5):1055–64.
- [24] Krebel U. Pairwise classification and support vector machines. In: *Adv in kernel methods: support vector learning*. Cambridge, MA: MIT Press; 1999. p. 255–68.
- [25] Hsu CW, Lin CJ. A comparison of methods for multi-class support vector machines. *IEEE Trans Neural Networks* 2002;13(2):415–25.
- [26] Wu TF, Lin CJ, Weng RC. Probability estimates for multi-class classification by pairwise coupling. *J Mach Learn Res* 2004;5:975–1005.
- [27] Chang E, Kingshy G, Sychay G, Gang W. CBSA: content-based soft annotation for multimodal image retrieval using bayes point machines. *IEEE Trans Circ Syst Video Technol* 2003;13:26–38.
- [28] Chang SF, Sikora T, Puri A. Overview of the MPEG-7 standard. *IEEE Trans Circ Syst Video Technol* 2001;11:688–95.
- [29] Haralick RM, Shanmugam, Dinstein I. Textural features for image classification. *IEEE Trans Syst Man Cybern* 1973;3:610–21.
- [30] Kittler J, Hatef M, Duin RPW, Matas J. On combining classifiers. *IEEE Trans Pattern Anal Mach Intell* 1998;20(3):226–39.
- [31] Fukunaga K. Introduction to statistical pattern recognition. 2nd ed. Boston: Academic Press; 1990.
- [32] Xu L, Krzyzak A, Suen CY. Methods of combining multiple classifiers and their applications to handwriting recognition. *IEEE Trans Syst Man Cybern* 1992;23(3):418–35.
- [33] Hansen LK, Salamon P. Neural network ensembles. *IEEE Trans Pattern Anal Mach Intell* 1990;12(10):993–1001.
- [34] Cho SB, Kim JH. Combining multiple neural networks by fuzzy integral for robust classification. *IEEE Trans Syst Man Cybern* 1995;25(2):380–4.
- [35] Chen Z, Liu WY, Zhang F, Li MJ, Zhang HJ. Web mining for web image retrieval. *J Am Soc Informat Sci Technol* 2001;52(10):831–9.
- [36] Zhou XS, Huang TS. Relevance feedback for image retrieval: a comprehensive review. *Multimedia Syst* 2003;8(6):536–44.
- [37] Baeza-Yates R, Ribiero-Neto B. Modern information retrieval. Boston, MA: Addison-Wesley; 1999.
- [38] Chang CC, Lin CJ. LIBSVM: a library for support vector machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>, 2001.

Md. Mahmudur Rahman received the M.Sc. degree in Computer Science from California State Polytechnic University, Pomona, California, USA in 2002. He is currently pursuing the Ph.D. degree in the Department of Computer Science & Software Engineering at Concordia University, Montreal, Canada. His research interests include Content-based Image Retrieval, Medical Image Annotation and Retrieval, Statistical and Interactive Learning in Multimedia Systems.

Bipin C. Desai is a professor in the Department of Computer Science & Software Engineering at Concordia University, Montreal, Canada. He is the General chair of IDEAS (International Database Engineering & Applications Symposium). His research interests include application of AI and Intelligent Systems, Database Engineering and Applications, Virtual Library, Web and its applications.

Prabir Bhattacharya received the D.Phil. degree in Mathematics in 1979 from the University of Oxford, UK specializing in group theory. He received his undergraduate education from St. Stephens College, University of Delhi, India. He is currently a Full Professor at the Concordia Institute for Information Systems Engineering, Concordia University, Montreal, Quebec, Canada where he holds a Canada Research Chair, Tier 1. During 1986–2001, he served at the Department of Computer Science and Engineering, University of Nebraska, Lincoln, USA where he joined as an Associate Professor and then became a Full Professor in 1994. During 1999–2004, he worked at the Panasonic Technologies Laboratory in Princeton, NJ, USA as a Principal Scientist (during 1999–2001 he took a leave of absence from the University of Nebraska). He also served as a Visiting Full Professor at the Center for Automation Research, University of Maryland, College Park, USA for extended periods. He is a Fellow of the IEEE, the International Association for Pattern Recognition (IAPR), and the Institute for Mathematics and Its Applications (IMA), UK. He is currently serving as the Associate Editor-in-Chief of the IEEE Transactions on Systems, Man and Cybernetics, Part B

(Cybernetics). Also, he is an associate editor of the Pattern Recognition, Pattern Recognition Letters, International Journal of Pattern Recognition and Artificial Intelligence, and Machine Graphics and Vision. During 1996–1998, he was on the editorial board of the IEEE Computer Society Press. He was a Distinguished Visitor of the IEEE Computer Society during 1996–1999, and a National Lec-

turer of the Association for Computing Machinery (ACM) during 1996–1999. He has co-authored over 190 publications including 91 journal papers, and co-edited a book on Vision Geometry (Oxford University Press). Also, he holds four US Patents and 7 Japanese Patents.