

Supervised Machine Learning based Medical Image Annotation and Retrieval

Md. Mahmudur Rahman, Bipin C. Desai, Prabir Bhattacharya
CINDI Group, Concordia University
1455, De Maisonneuve Blvd. West,
Montreal, QC, H3G 1M8, Canada
mah_rahm@cs.concordia.ca

Abstract

This paper presents the approaches and experimental results of image annotation and retrieval in our first participation of ImageCLEFmed 2005. In this work, we investigate a supervised learning approach to associate low-level global image features with their high level visual and/or semantic categories for image annotation and retrieval. For automatic image annotation, we represent input images through a large dimensional feature vector of texture, edge and shape features. A multi-class classification system based on pairwise coupling of several binary support vector machine (SVM) is trained on this input to predict the categories of test images, which will be effective for later annotation. For visual only retrieval, we utilize a low dimensional feature vector of color, texture and edge features based on principal component analysis (PCA) and category specific feature distribution information in a statistical similarity measure function. Based on the online category prediction of query and database images by the multi-class SVM classifier, pre-computed category specific first and second order statistical parameters are utilized in Bhattacharyya distance measure on the assumption that distributions are multivariate Gaussian. Experimental results of both image annotation and retrieval are reported in this paper.

Categories and Subject Descriptors

H.3 [Information Storage and Retrieval]: H.3.1 Content Analysis and Indexing; H.3.3 Information Search and Retrieval; H.3.7 Digital Libraries; I.4.8 [Image Processing and Computer Vision]: Scene Analysis—*Object Recognition*

General Terms

Algorithms, Machine learning, Performance, Experimentation

Keywords

Image annotation, Classification, Content based image retrieval, Support vector machine, Statistical distance measure

1 Introduction

During the last decade, there have been an overwhelming research interests in medical image retrieval and classification from different communities [11, 10]. Medical images of various modalities (X-ray, CT, MRI, Ultrasound etc.) constitute an important source of anatomical and functional

information for the diagnosis of diseases, medical research and education. Effectively and efficiently searching and annotating these large image collections poses significant technical challenges.

ImageCLEFmed is part of the Cross Language Evaluation Forum (CLEF), a benchmarking event for multilingual information retrieval. The main goal of ImageCLEFmed is to improve the retrieval of medical images from heterogeneous and multilingual document collections containing images as well as text. This year (2005) in ImageCLEFmed, there are mainly two task: image annotation and retrieval to be performed.

In automatic image annotation task, the main aim is to find out how well current techniques can identify image modality, body orientation, body region, and biological system examined based on the images. Here, we utilized a database of 9,000 fully classified images taken randomly from medical routine to train a classification system. 1,000 images for which classification labels are not available are used as test images and have been classified by a multi-class classification system. The results of the classification step can be used for multilingual image annotations as well as for DICOM header corrections.

In image retrieval task, we have experimented with a visual only approach, where an example image is used to perform a search against a medical image database to find similar images based on visual attributes (color, texture, etc.). Each medical image or a group of images represents an illness, and case notes in English or French are associated with each illness to be used for diagnosis. This task is based on the Casimage, MIR, PEIR, and PathoPIC datasets, containing about 50,000 images of different modalities (CT, MRI, X-ray etc.).

2 Image annotation

Automatic image annotation or image classification is an area of active research in the field of machine learning and pattern recognition. Retrieval systems have traditionally used manual image annotation for indexing and then later retrieving their image collections. However, manual image annotation is an expensive and labor intensive procedure [7]. Here, we investigate an automatic approach to categorize or annotate images based on a supervised learning technique. A classification task usually involves with training and testing data which consist of some data instances. In supervised classification, we are given a collection of labeled images (or priori knowledge), and the problem is to label a newly encountered, yet unlabeled image. Each instance in the training set contains category or class specific labels and several image feature descriptors in the form of a combined feature vector.

In this work, we present effective texture, edge and shape descriptors to represent image content at global level and input in to a multi-class classification system based on several binary support vector machine (SVM) classifier. The goal of SVM is to produce a model which will predict target value or category of images with highest probability or confidence in the testing set which are given as input feature vectors to the classification system.

2.1 Feature Selection & Extraction

This section describes how our system characterizes images for efficient representation. The collection contained mainly monochrome or gray level medical images with specific layout. Hence, we characterize images by texture, edge and shape features, whereby ignoring color information totally. We know that only one type of low-level features does not work well. So, we incorporate these features in a combined feature vector and use it as input to the classification system. The accuracy of our classification system depends greatly on the representation of these low-level visual features. These image features help propagate annotation from trained images to unlabeled images. The more discriminative the low-level features, the more accurate the content-based classification or annotation.

2.1.1 Texture Feature

Medical images of different categories can be distinguished via their homogeneousness or texture characteristics. Therefore, it may be useful to extract texture features for image classification.

We include a texture feature set approach to extract spatially localized texture information from the gray level co-occurrence matrix [1]. A gray level co-occurrence matrix is defined as a sample of the joint probability density of the gray levels of two pixels separated by a given displacement d and angle θ . The $G \times G$ gray level co-occurrence matrix p for a displacement vector $d = (dx, dy)$ is defined as [16]:

$$p(i, j) = |\{(r, s), (t, v) : (I(r, s) = i, I(t, v) = j)\}|, \quad (1)$$

where $(r, s), (t, v) \in N \times N, (t, v) = (r + dx)(s + dy)$.

Likewise image histograms, which can be thought of as an estimate of the probability distribution of gray values or color levels in an image, a co-occurrence matrix is an estimate of the joint probability distribution of pairs of gray-levels at some fixed relative position in the image. Typically, the information stored in a co-occurrence matrix is sparse. Also, it is often useful to consider a number of co-occurrence matrices, one for each relative position of interest, in order to grasp different textual cues or the same textual cur at different scale. In order to obtain efficient descriptors, the information contained in co-occurrence matrices is traditionally condensed in a few statistical features. We obtained four co-occurrence matrices for four different orientations (horizontal 0° , vertical 90° , and two diagonals 45° and 135°) and normalize the entries [0,1] by dividing each entry by total number of pixels.

The co-occurrence matrix reveals certain properties about the spatial distribution of the gray levels in the texture image. For example, if most of the entries in the co-occurrence matrix are concentrated along the diagonals, then the texture is coarse with respect to the displacement vector d . Haralick has proposed a number of useful texture features that can be computed from the co-occurrence matrix. Higher order features, such as energy, entropy, contrast, homogeneity and maximum probability are measured based on each gray level co-occurrence matrix to form a five dimensional feature vector as follows [16]:

$$Energy = \sum_i \sum_j p^2(i, j) \quad (2)$$

$$Entropy = - \sum_i \sum_j p(i, j) \log p(i, j) \quad (3)$$

$$Contrast = \sum_i \sum_j (i - j)^2 p(i, j) \quad (4)$$

$$Homogeneity = \sum_i \sum_j \frac{p(i, j)}{1 + |i - j|} \quad (5)$$

$$Maximum \ probability = \max p(i, j) \quad (6)$$

Finally, we obtained a twenty dimensional feature vector by concatenating the feature vectors of each co-occurrence matrix.

2.1.2 Edge Feature

To represent the edge feature on a global level, a histogram of edge direction is constructed. 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 [4]. 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. Hence, the dimension of the edge feature vector is 72.

2.1.3 Shape Feature

Image classification for medical images with gray scale images have to use shape feature for efficient representation. We represent the global shape of an image in terms of seven invariant moments [8]. These features are invariant under rotation, scale, translation, and reflection of images and have been widely used in a number of applications due to their invariance properties. For a 2-D image, $f(x, y)$, the central moment of order $(p + q)$ is given by

$$\mu_{pq} = \sum_x \sum_y (x - \bar{x})^p (y - \bar{y})^q f(x, y). \quad (7)$$

Seven moment invariants ($M_1 - M_7$) based on the second and third order moments are given as follows [8]:

$$\begin{aligned} M_1 &= (\mu_{20} + \mu_{02}) \\ M_2 &= (\mu_{20} - \mu_{02})^2 + \mu_{11}^2 \\ M_3 &= (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2 \\ M_4 &= (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2 \\ M_5 &= (\mu_{30} + \mu_{12}) + (\mu_{30} - 3\mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ &\quad + (3\mu_{21} - \mu_{03}) \times (\mu_{21} + 3\mu_{03})[3(\mu_{03} + \mu_{21})^2 - (\mu_{21} - \mu_{03})^2] \\ M_6 &= (\mu_{20} + \mu_{02})[(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2] + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}) \\ M_7 &= (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})[(\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2] \\ &\quad - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})[3(\mu_{03} + \mu_{21})^2 - (\mu_{21} - \mu_{03})^2] \end{aligned} \quad (8)$$

$M_1 - M_6$ are invariant under rotation and reflection. M_7 is invariant only in its absolute magnitude under a reflection.

Now, let us consider f_t , f_e , and f_s be the texture, edge and shape feature vector respectively of an image. Now the composite feature vector is formed by simple concatenation of each individual feature vector as $F_{\text{combined}} = (f_t + f_e + f_s)$, where the dimension R^d of F_{combined} is the sum of individual feature vector dimension and $d = (20 + 72 + 7) = 99$. We use this vector to represent the images. Thus, the input space for our SVM classifiers is a 99-dimensional space, and each image in our database corresponds to a point in this space.

2.2 Category prediction with multi-class SVM

Support vector machine (SVM) is an emerging machine learning technology which has been successfully used in content based image retrieval [5].

Given training data $(\vec{x}_1, \dots, \vec{x}_n)$ that are vectors in some space $\vec{x}_i \in R^n$ and their labels (y_1, \dots, y_n) where $y_i \in \{+1, -1\}^n$, the general form of the binary linear classification function is

$$g(x) = \vec{w} \cdot \vec{x} + b \quad (9)$$

which corresponds to a separating hyper plane

$$\vec{w} \cdot \vec{x} + b = 0 \quad (10)$$

where \vec{x} is an input vector, \vec{w} is a weight vector, and b is a bias. The goal of SVM is to find the parameters \vec{w} and b for the optimal hyper plane to maximize the geometric margin $\frac{2}{\|\vec{w}\|}$ between the hyper planes, subject to the solution of the following optimization problem [3]:

$$\min_{\vec{w}, b, \xi} \quad \frac{1}{2} \vec{w}^T \vec{w} + C \sum_{i=1}^n \xi_i \quad (11)$$

subject to

$$y_i(\vec{w}^T \phi(\vec{x}_i) + b) \geq 1 - \xi_i \quad (12)$$

where $\xi_i \geq 0$ and $C > 0$ is the penalty parameter of the error term. Here training vectors \vec{x}_i are mapped into a high dimensional space by the non linear mapping function $\phi : R^n \rightarrow R^f$, where $f > n$ or f could even be infinite. Optimization problem and its solution can be represented by the inner product. Hence,

$$\vec{x}_i \cdot \vec{x}_j \rightarrow \phi(\vec{x}_i)^T \phi(\vec{x}_j) = K(\vec{x}_i, \vec{x}_j) \quad (13)$$

where K is a kernel function. The SVM classification function is given by [5]:

$$f(\vec{x}) = \text{sign} \left(\sum_{i=1}^n \alpha_i y_i K(\vec{x}_i, \vec{x}) + b \right) \quad (14)$$

A number of methods have been proposed for extension to multi-class problem to separate L mutually exclusive classes essentially by solving many two-class problems and combining their predictions in various ways [5]. One technique, commonly known as Pairwise Coupling (PWC) or “one-vs.-one” is to construct SVMs between all possible pairs of classes. This method trains $L * (L - 1)/2$ binary classifiers, each of which provides a partial decision for classifying a data point. PWC then combines the output of all classifiers to form a class prediction. During testing, each of the $L * (L - 1)/2$ classifier votes for one class. The winning class is the one with the largest number of accumulated votes. We use this technique for the implementation of our multi-class SVM by using the LIBSVM software package [6].

Multi-class image classification systems work by classifying an image into one of many pre-defined categories. There are 57 categories of images provided in the training set for the image annotation task. So, we define a set of 57 labels where each label characterizes the representative semantics of an image category. Class labels along with feature vectors are generated from all images at the training stage. In testing stage, each unannotated image is classified against the 57 categories using PWC or “one-vs.-one” technique. This produces a ranking of the 57 categories, with each category assigned a confidence or probability score to each image. The labels of these 57 categories, along with their probabilities, become the annotation for the testing images as the following format:

$\langle \textit{imageno} \rangle \langle \textit{confidence class1} \rangle \langle \textit{confidence class2} \rangle \dots \langle \textit{confidence class57} \rangle$
--

The confidence represents the weight of a label or category in the overall description of an image and the class with the highest confidence is considered to be the class of the image. This ranking of categories based on confidence score can be very useful in future annotation or search purposes.

3 Image retrieval

Currently, most content based image retrieval (CBIR) systems are similarity-based, where similarity between query and target images in a database is measured by some form of distance metrics in feature space [15]. However, retrieval systems generally conduct this similarity matching on a very high-dimensional feature space without any semantic interpretation or paying enough attention to the underlying distribution of the feature space [14]. However, high dimensional feature vectors not only increase the computational complexity in similarity matching and indexing, but also increase the logical database size. For many frequently used visual features in medical images, often their category specific distributions are also available. In this case, it is possible to extract a set of low-level features (e.g., color, texture, shape, etc.) to predict semantic categories of each image by identifying its class assignment using a classifier. Thus, an image can be best characterized by exploiting the information of feature distribution from its semantic category.

In the image retrieval task, we have investigated a category based adaptive statistical similarity measure technique on the low-dimensional feature space. For this, we utilized principal component analysis (PCA) for dimension reduction and multi-class support vector machine (SVM) for online

category prediction of query and database images. Hence, category specific statistical parameters in low-dimensional feature space can be exploited by the statistical distance measure in real time similarity matching.

3.1 Feature extraction & representation in PCA sub-space

The performance of a CBIR system mainly depends on the particular image representation and similarity matching function employed. Images in four datasets contain both color and gray level images for retrieval evaluation. Hence, we have extracted color, texture and edge features for our image representation at global level.

Color is the most useful low-level feature and its histogram based representation is one of the earliest descriptors, which is widely used in CBIR [15]. For color feature, a 108 dimensional color histogram is created in vector form on HSV (Hue, Saturation, Value) color space. In HSV space, the colors correlates well and can be matched in a way that is consistent with human perception. In this work, we uniformly quantized HSV space into 12 bins for hue(each bin consisting of a range of 30°), 3 bins for saturation and 3 bins for value, which results in 108 bins for color histogram. Many medical images of different categories can be distinguished via their homogeneous texture and global edge characteristics. Hence, we have extracted the same global texture and edge features as measured for image annotation.

As the dimensions of color, texture and edge ($108+20+72 = 200$) feature vectors are high, we need to apply some dimension reduction technique to reduce the computational complexity and logical database size. Moreover, if the training samples used to estimate the statistical parameters are smaller compare to the size of feature dimension, then inaccuracy or singularity may arise for second order (covariance matrix) parameter estimation. The problem of selecting most representative feature attributes commonly known as dimension reduction, has been examined by principal component analysis (PCA) [12] in our experiment. The basic idea of PCA is to find m linearly transformed components so that they explain the maximum amount of variances in the input data and mathematical steps used to describe the method is as follows:

given a set of N feature vectors (training samples) $\vec{x}_i \in R^d | i = (1 \cdots N)$, where the mean vector(μ) and covariance matrix (C) is estimated as

$$\mu = \frac{1}{N} \sum_{i=1}^N \vec{x}_i \quad \& \quad C = \frac{1}{N} \sum_{i=1}^N (\vec{x}_i - \mu)(\vec{x}_i - \mu)^T \quad (15)$$

Let ν_i and λ_i be the eigenvectors and the eigenvalues of C , then they satisfy the following:

$$\lambda_i = \sum_{i=1}^N (\nu_i^T (\vec{x}_i - \mu))^2 \quad (16)$$

Here, $\sum_{i=1}^N \lambda_i$ accounts for the total variance of the original feature vectors set. Now, PCA method tries to approximate the original feature space using an m dimensional feature vector, that is using m largest eigenvalues account for a large percentage of variance, where typically $m \ll \min(d, N)$. These m eigenvectors span a subspace, where $\vec{V} = [\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_m]$ is the $d \times m$ -dimensional matrix that contains orthogonal basis vectors of the feature space in its columns. The $m \times d$ transformation \vec{V}^T transforms the original feature vector from $R^d \rightarrow R^m$ ones. That is

$$\vec{V}^T (\vec{x}_i - \mu) = \vec{y}_i, i = 1 \cdots N \quad (17)$$

where $\vec{y}_i \in R^m$ and k th component of the \vec{y}_i vector is called the k th principal component (PC) of the original feature vector \vec{x}_i . So, the feature vector in the original R^d space for query and database images can be projected on to the R^m space via the transformation of \vec{V}^T [12]. We have applied this PCA technique to our composite feature vector and reduced the feature dimension for subsequent SVM training and similarity matching.

3.2 Adaptive statistical distance measure

Statistical distance measure, defined as the distances between two probability distributions, finds its uses in many research areas in pattern recognition, information theory, and communication. It captures correlations or variations between attributes of the feature vectors and provides bounds for probability of retrieval error of a two way classification problem. Recently, CBIR community also adopted statistical distance measures for similarity matching [2, 14]. In this scheme query image q and target image t are assumed to be in different classes and their respective density as $p_q(\vec{x})$ and $p_t(\vec{x})$, both defined on R^d . When these densities are multivariate normal, they can be approximated by mean vector μ and covariance matrix C as $p_q(\vec{x}) = N(\vec{x}; \mu_q, C_q)$ & $p_t(\vec{x}) = N(\vec{x}; \mu_t, C_t)$ where,

$$N(\vec{x}; \mu, C) = \frac{1}{\sqrt{(2\pi)^d |C|}} \exp^{-\frac{1}{2}(\vec{x}-\mu)^T C^{-1}(\vec{x}-\mu)} \quad (18)$$

here, $\mathbf{x} \in R^d$ and $|\cdot|$ is matrix determinant [9]. A popular measure of similarity between two Gaussian distributions is the Bhattacharyya distance, which is equivalent to an upper bound of the optimal Bayesian classification error probability [9] [13]. Bhattacharyya distance (D_{Bhatt}) between query image q and target image t in the database is given by:

$$D_{\text{Bhatt}}(q, t) = \frac{1}{8}(\mu_q - \mu_t)^T \left[\frac{(C_q + C_t)}{2} \right]^{-1} (\mu_q - \mu_t) + \frac{1}{2} \ln \frac{\left| \frac{(C_q + C_t)}{2} \right|}{\sqrt{|C_q| |C_t|}} \quad (19)$$

where μ_q and μ_t are the mean vectors, and C_q and C_t are the covariance matrices of query image q and target image t respectively. Equation (19) is composed of two terms, the first one being the distance between mean vectors of images, while the second term gives the class separability due to the difference between class covariance matrices. In the retrieval experiment, we used the Bhattacharyya distance measure for similarity matching. Here, it is called adaptive due to the nature of online selection of mean (μ) and covariance matrix (C) by the multi-class SVM as discussed in the next section.

3.3 Category prediction & parameter estimation

To utilize category specific distribution information in similarity matching, we utilize the multi-class SVM classifier as described in section 2.2 to predict the category of query and database images. Based on the online prediction, distance measure functions will be adjusted to accommodate category specific parameters for query and reference images of database. To estimate the parameters of the category specific distributions, feature vectors are extracted and reduced in dimension as mentioned in section 3.1, from N selected training image samples. It is assumed that feature of each category will have distinguishable normal distribution. Computing the statistical distance measures between two multivariate normal distributions requires first and second order statistics in the form of mean (μ) and covariance matrix (C) or parameter vector $\theta = (\mu, C)$ as described in previous section. Suppose that there are L different semantic categories in the database, each assumed to have a multivariate normal distribution with mean vector μ_i and covariance matrix C_i , for $i \in L$. However, the true values of μ and C of each category usually are not known in advance and must be estimated from a set of training samples N [9]. The μ_i and C_i of each category are estimated as

$$\mu_i = \frac{1}{N_i} \sum_{j=1}^{N_i} \vec{x}_{i,j} \quad \& \quad C_i = \frac{1}{N_i - 1} \sum_{j=1}^{N_i} (\vec{x}_{i,j} - \mu_i)(\vec{x}_{i,j} - \mu_i)^T \quad (20)$$

where $\vec{x}_{i,j}$ is sample j from category i , N_i is the number of training samples from category i and $N = (N_1 + N_2 + \dots + N_L)$.

4 Experimentation & Results

For annotation experiment, we perform the following procedure at the training stage:

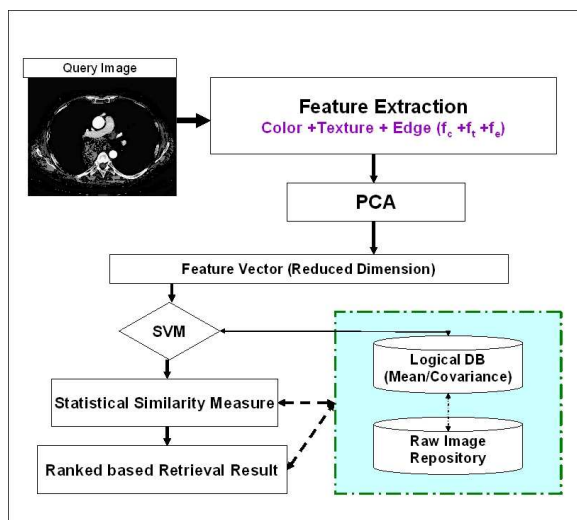


Figure 1: Block diagram of the retrieval technique

- Consider the radial basis kernel function (RBF) $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, $\gamma > 0$.
- Use cross-validation to find the best parameter C and γ
- Use the best parameter C and γ to train the whole training set

For image classification, recent work [5] shows that the RBF kernel works well with high-dimensional image data and can handle the case when the relation between class labels and attributes is nonlinear. Therefore, we use RBF kernel as a reasonable first choice. There are two tunable parameters while using RBF kernels and soft-margin SVMs in the version space: C and γ . The γ in the RBF kernel controls the shape of the kernel and C controls the trade-offs between margin maximization and error minimization. Increasing C may decrease training error, but it can also lead to poor generalization [7].

It is not known beforehand which C and γ are the best for our classification problem. In the training stage, the goal is to identify good (C and γ), so that the classifier can accurately predict testing data. It may not be useful to achieve high training accuracy (i.e., classifiers accurately predict training data whose class labels are indeed known). Therefore, we use a 10 fold cross-validation with various combinations of γ and C to measure the classification accuracy.

In 10-fold cross-validation, we first divide the training set into 10 subsets of equal size. Sequentially one subset is tested using the classifier trained on the remaining 9 (10-1) 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 and using cross-validation. Basically pairs of (C , γ) are tried and the one with the best cross-validation accuracy is picked. We find the best (C , γ) is (200, 0.01) with the cross-validation rate 54.65%. After the best (C , γ) is found, the whole training set of 9,000 images is trained again to generate the final classifiers. Finally, the generated classifiers are tested on the 1000 image testing set of unknown labels to generate the image annotation.

In ImageCLEFmed 2005 automatic image annotation experiment, we have submitted only one run with the above parameters and the classification error rate is 43.3%. Which means, 433 images were misclassified out of 1000 or accuracy of our system is 56.7% at this moment.

In image retrieval experiment, for statistical parameter estimation and SVM training, we observed the images of four datasets closely and selected 33 different categories based on perceptual and modality specific differences, each with 100 images for generating the training samples. However, for actual evaluation of similarity measure functions, we conducted our experiments on the entire database (Around 50,000 images from four different collections). For SVM training, we used

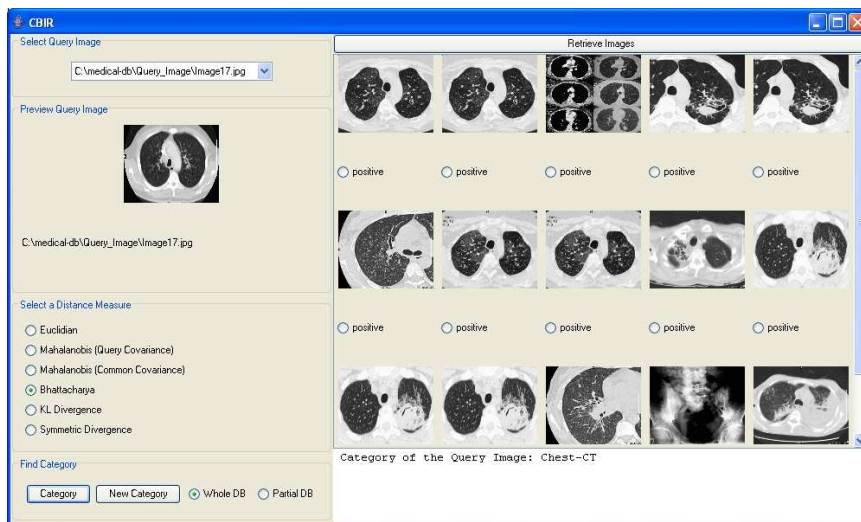


Figure 2: A snapshot of the retrieval result (most similar 15 images) of a Chest-CT (Image17.jpg) query image

the reduced feature vector with radial basis kernel (RBF) function. After 10 fold cross validation, we found the best parameters $C = 100$ and $\gamma = .03$ with an accuracy of 72.96% in our current setting and finally trained the whole training set with these parameters. The dimensionality of the feature vector is reduced to $R^d \rightarrow R^m$, where $d=200$ and $m=25$ and accounted for 90.0% of the total variances. In the ImageCLEFmed 2005 evaluation, we have submitted only one run with the above parameters and achieved a mean average precision of 0.0072 across all queries, which is very low at this moment compare to other systems.

5 Conclusion and future work

This is the first year for the CINDI group taking part in the ImageCLEF campaign and specially in the ImageCLEFmed track. This year, our main goal was to participate and do some initial experiment with the databases provided by the organizer.

This report presents our approaches to automatic image annotation and retrieval based on global image feature contents and multi-class SVM classifier. Despite having 57 categories for annotation, many of them closely-related, our classification system is still able to provide moderate classification performance. In future, we will investigate with region based local image features and statistical methods that can deal with the challenges of an unbalanced training set, as provided in the current experimental setting.

The retrieval performance of our system is not good enough due to the following main reasons. First of all, it is very difficult to select a reasonable training set of images with predefined categories from four different datasets with a huge amount of variability in image size, resolution, modality etc. The performance of our system is critical to the appropriate training of multi-class SVM as parameter selection of statistical distance measure is depended on the online category prediction. Secondly, we have only used global image feature, which may not suitable for medical images as large unwanted background dominated in many of these images. In future we will try to resolve these issues and incorporate text based search approach in addition with the visual based approach.

References

- [1] Aksoy, S., Haralick, R. M.: Texture Analysis in Machine Vision., Chapter Using Texture in Image Similarity and Retrieval, Series on Machine Perception and Artificial Intelligence., World Scientific (2000)
- [2] Aksoy, S., Haralick, R.M.: Probabilistic vs. geometric similarity measures for image retrieval., Proceedings. IEEE Conference on Computer Vision and Pattern Recognition, **2**(2000) 357–362
- [3] Burges, C.: A tutorial on support vector machines for pattern recognition., Data Mining and Knowledge Discovery, **2**, (1998) 121–167
- [4] Canny, J. : A computational approach to edge detection., IEEE Trans. Pattern Anal. Machine Intell., **8** (1986) 679–698
- [5] Chapelle, O., Haffner, P., Vapnik, V.: SVMs for histogram-based image classification., IEEE Transaction on Neural Networks (1999)
- [6] Chang, C. C., Lin, C.J.: LIBSVM : a library for support vector machines., (2001) Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
- [7] Chang, E. Kingshy, G. Sychay, G. Gang, W. : CBSA: Content-based Soft Annotation for Multimodal Image Retrieval Using Bayes Point Machines., IEEE Transactions on Circuits and Systems for Video Technology., **13**(2003) 26–38
- [8] Dudani, S. A., Breeding, K. J. and McGhee, R. B. : Aircraft identification by moment invariants., IEEE Trans. Comput. **C-26** (1977) 39–45
- [9] Fukunaga, K.: Introduction to Statistical Pattern Recognition., Second ed. Academic Press, (1990)
- [10] H. D. Tagare, C. Jafe, J. Duncan. :Medical image databases: A content-based retrieval approach., Journal of the American Medical Informatics Association., **4** (**3**) (1997) 184–198
- [11] H. Muller, N. Michoux, D. Bandon, A. Geissbuhler :A review of content-based image retrieval applications -clinical benefits and future directions., International Journal of Medical Informatics, **73:1**(2004) 1–23
- [12] Jain, A.K., Bhandrasekaran, B. : Dimensionality and sample size considerations in pattern recognition practice., Handbook of Statistics, **2** (1987) 835–855
- [13] Kailath, T.: The divergence and Bhattacharyya distance measures in signal selection., IEEE Trans. Commun. Technol, **COM-15**(1967) 52–60
- [14] Puzicha, J., Buhmann, J., Rubner, Y., Tomasi, C.: Empirical evaluation of dissimilarity measures for color and texture., Intern. Conf. on Computer Vision, (1999)
- [15] Smeulder, A., Worring, M., Santini, S., Gupta, A., Jain, R.: Content-Based Image Retrieval at the End of the Early Years., IEEE Trans. on Pattern Anal. and Machine Intell., **22**, (2000) 1349–1380
- [16] Tuceryan, M., Jain, A. K. : The Handbook of Pattern Recognition and Computer Vision (2nd Edition), by C. H. Chen, L. F. Pau, P. S. P. Wang (eds.), World Scientific Publishing Co., (1998) pp. 207–248