A novel content-based medical image retrieval method based on query topic dependent image features (QTDIF)

Wei Xiong [(1)], Bo Qiu [(1)], Qi Tian [(1)], Henning Müller[(2)], Changsheng Xu [(1)]
[(1)]Institute for Infocomm Research, Singapore
[(2)]University Hospitals of Geneva, Service of Medical Informatics, Switzerland
Email: {wxiong, visqiu, tian, xucs}@i2r.a-star.edu.sg, henning.mueller@sim.hcuge.ch

ABSTRACT

Medical image retrieval is still mainly a research domain with a large variety of applications and techniques. With the ImageCLEF 2004 benchmark, an evaluation framework has been created that includes a database, query topics and ground truth data. Eleven systems (with a total of more than 50 runs) compared their performance in various configurations. The results show that there is not any one feature that performs well on all query tasks. Key to successful retrieval is rather the selection of features and feature weights based on a specific set of input features, thus on the query task. In this paper we propose a novel method based on query topic dependent image features (QTDIF) for content-based medical image retrieval. These feature sets are designed to capture both inter-category and intra-category statistical variations to achieve good retrieval performance in terms of recall and precision. We have used Gaussian Mixture Models (GMM) and blob representation to model medical images and construct the proposed novel QTDIF for CBIR. Finally, trained multi-class support vector machines (SVM) are used for image similarity ranking. The proposed methods have been tested over the Casimage database with around 9000 images, for the given 26 image topics, used for imageCLEF 2004. The retrieval performance has been compared with the medGIFT system, which is based on the GNU Image Finding Tool (GIFT). The experimental results show that the proposed QTDIF-based CBIR can provide significantly better performance than systems based general features only.

1. INTRODUCTION

Content-based image retrieval or visual multimedia retrieval is one of the most active research areas closely related to the fields of computer vision, image processing and information retrieval [1]. Reasons for the large number of developed systems include the exploding amount of visual data being produced in digital form and thus high demand for more efficient and effective access to these images.

Massive amounts of visual medical images are being produced in hospitals today and it is expected that the amounts of medical images will further increase in the future. Several hospitals start having completely digital archives and at least CT and MRI pictures are almost always stored in digital archives as film archives are extremely expensive to handle. A typical university hospital currently produces several thousand images a day and the yearly production is in the range of terabytes. Still the access to these visual data is still almost exclusively done by patient and study identification. This does not use the available knowledge up to its full potential. Visual access methods can be a missing piece to create a medical visual knowledge management. Still, most currently available techniques for varied datasets perform unreliably even for relatively easy tasks. Only very specialized applications on limited datasets are available. There are great needs for improved content-based visual image search, navigation and browsing tools to make efficient and effective image access possible so that we can make use of these data to their full potential, including the textual information that the images are attached to. Content-based image retrieval is an advanced method to search, navigate and browse large medical image databases that can help as a diagnostic aid [2], for research studies and to manage teaching file systems and similar image collections [3].

Many visual descriptors exist to describe grey level distributions, textures and shapes in medical images [1]. Large amounts of literature exist but only few comparisons of such feature sets for image retrieval tasks that state reproducible performance measures. Partly, this is due to a lack of commonly accessible medical databases. In this study we will use the Casimage database [16] to benchmark the performance of the proposed query dependent feature selection methods. The Casimage database is freely available and is also being used for the image retrieval benchmark imageCLEF [4]. It contains almost 9000 images, and 26 query topics were chosen for the imageCLEF competition (see Fig. 1). Ground truth for these query topics exist so we can evaluate retrieval results with several subsets of the features and compare them. As the dataset is freely available, the results are reproducible and can be compared with other techniques.

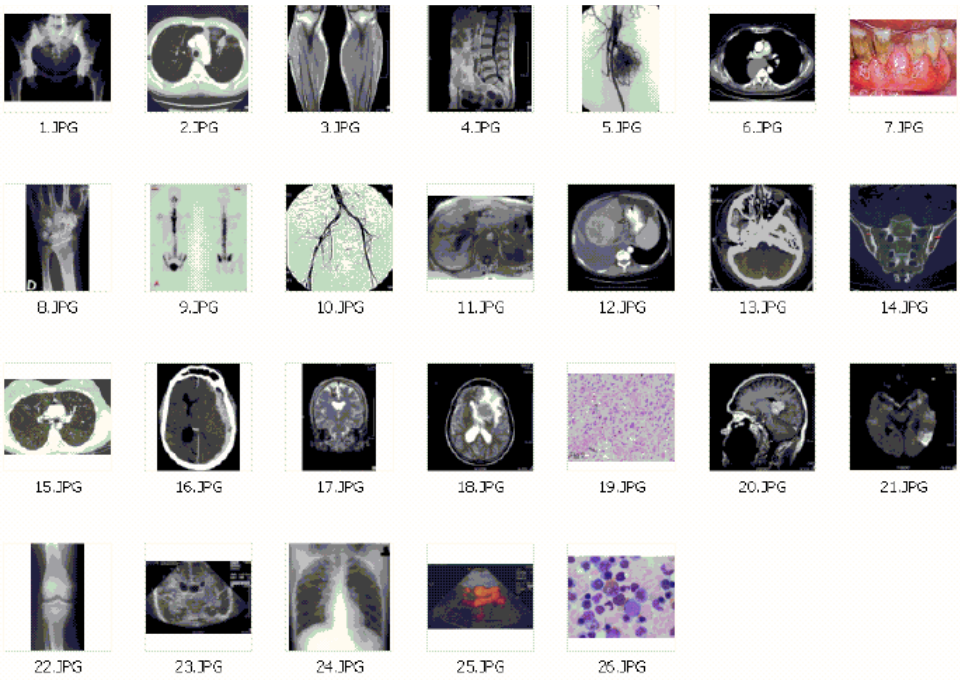

Figure 1. 26 image topics defined for the ImageCLEF medical image database.

In the past 10 years, extensive work has been done for general content-based image retrieval. Fig. 2 shows the general CBIR framework, which consists of four main components: feature extraction, similarity measures, a database of pre-analyzed image collections, and a relevance feedback loop [3]. Many features have been proposed for CBIR, including global and local features based on color, texture and the shape of objects, or region-based features, etc. All these features are general features and *independent of query image categories or topics.* This is due to the very nature of content-based image retrieval where we do not know how many categories of images there are in the image collections. In this article we will use the two terms interchangeably mainly depending the context. In the medical image retrieval, in particularly, ImageCLEF, the term, *topic*, 26 image topics are defined to benchmark the performance of various CBIR systems. In the general CBIR literatures the term *categories* are used instead.

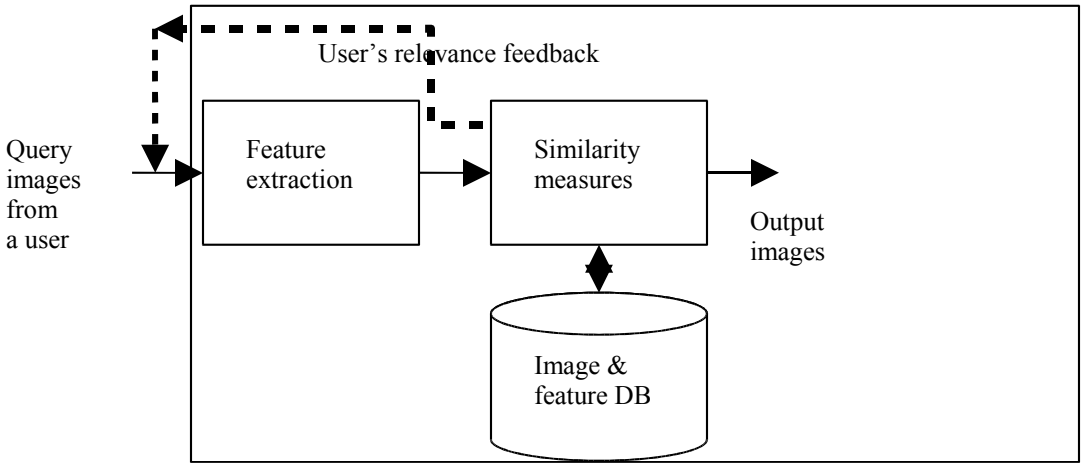

Figure 2. General architecture for CBIR with relevance feedback where image features are query-independent.

Different from general image collections, medical images have its own unique special characterization. Among other characteristics one unique feature of medical images is that in general there are a limited numbers of image *categories or topics* which depend on several factors: body parts, view positions and image modalities. Based on this very fact, we propose a novel image feature for content-based medical image retrieval: the proposed features are dependent on specific

image query topics. Therefore, these feature sets capture both inter-category and intra-category statistical variations to achieve good performance of recall and precision. Conceptually, we can consider that general CBIR works on open image collections where there are unlimited image categories. Even for home photos there are so many different situations for photo taking, it is almost impossible to count how many categories we may end up with if we try to find out how many categories we could have for home photo image collections. However for medical images used in hospitals the above is not completely true because of a limited number of modalities of medical image acquisition devices, limited body parts or organs. Based on these medical categories it is possible to develop query topic dependent features and associate similarity measures to achieve better retrieval performance.

In the next Section, the methodology of the proposed approach is elaborated. An implementation of such an approach is introduced in the following Sections. Specifically, we will elaborate the Gaussian Mixture Model, local regional feature extraction and representation in Section 3. In Section 4, multiple-class SVM for image retrieval will be explained briefly. Experimental results are presented in Section 5. At last a discussion and our conclusions are presented.

2. METHODOLOGY

The proposed general system architecture is shown in Fig. 3 for using query topic dependent image features (QTDIF) for content-based medical image retrieval (CBMIR). What is different from the general architecture for CBIR shown in Fig.2 is that the extracted features depend upon the query topics.

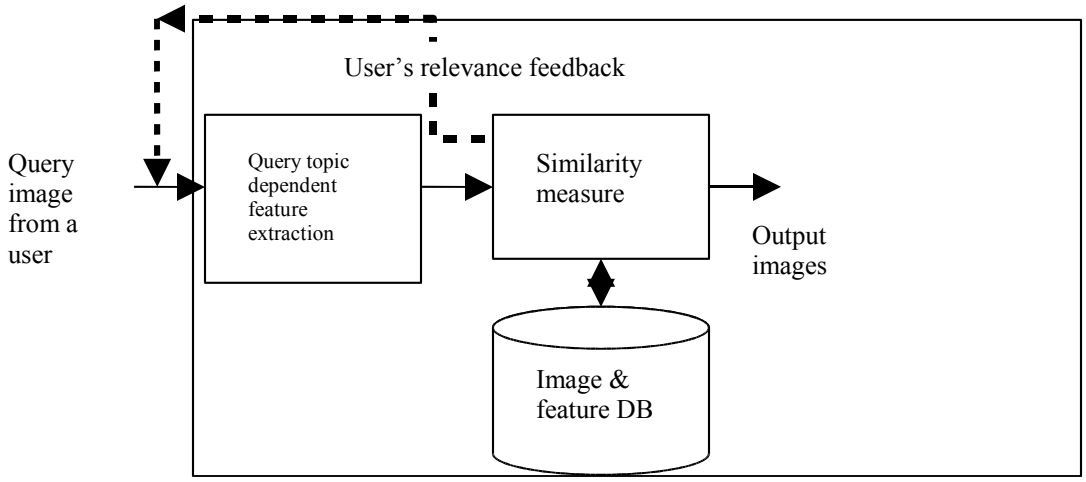

Figure 3. A novel general architecture for using QTDIF for CBMIR.

It is important to point out that for a pre-defined topic set, for instance, the 26 topics used in the ImageCLEF, it is relatively easy for people to determine which topic a given query image belongs to, at least for most of the topics (see Fig. 1 showing the images). Therefore, a user can inform a CBIR system to use better features for retrieval based on known topic information. We can consider the CBIR process actually performed in two steps: the first step is done by humans to identify a given query's topic class; the second step is done by computers to use QTDIF and similarity measures for image retrieval from the image DB. Effectively, the above process makes best use of both human and computer capabilities. This can be seen in Fig. 4.

For the proposed QTDIF-based medical image retrieval the key design considerations include how to design good features for given topics and also how to perform similarity measures, either based on clustering method or trained classifiers to determine the most similar images for a given query topic. In our current work we have chosen to use a blob representation to represent regional features of medical image samples, and then to use multi-class support vector machine (SVM) to perform similarity measuring of the given query image and the image collection stored in the DB. Please note that we assume that we have a pre-defined image topic set, in our experiments on the ImageCLEF medical image DB there are 26 topics defined as shown in Fig. 1.

There are several ways to design and implement the proposed QTDIF based CBIR system. One way is to design different features for several groups of topics based on the image characteristics such as presence of structural features, or smoothness of main textures, or distribution of edges along main directions, etc. In the end, the given topics can be divided into several groups and each group has one corresponding set of features. Another possible way is that we

experimentally decide one sub-set of the given topics and corresponding topic-dependent features, then the rest of the topics will only use a general image feature set. In general, both methods require some experiments to design the feature sets to make the proposed method effective and efficient.

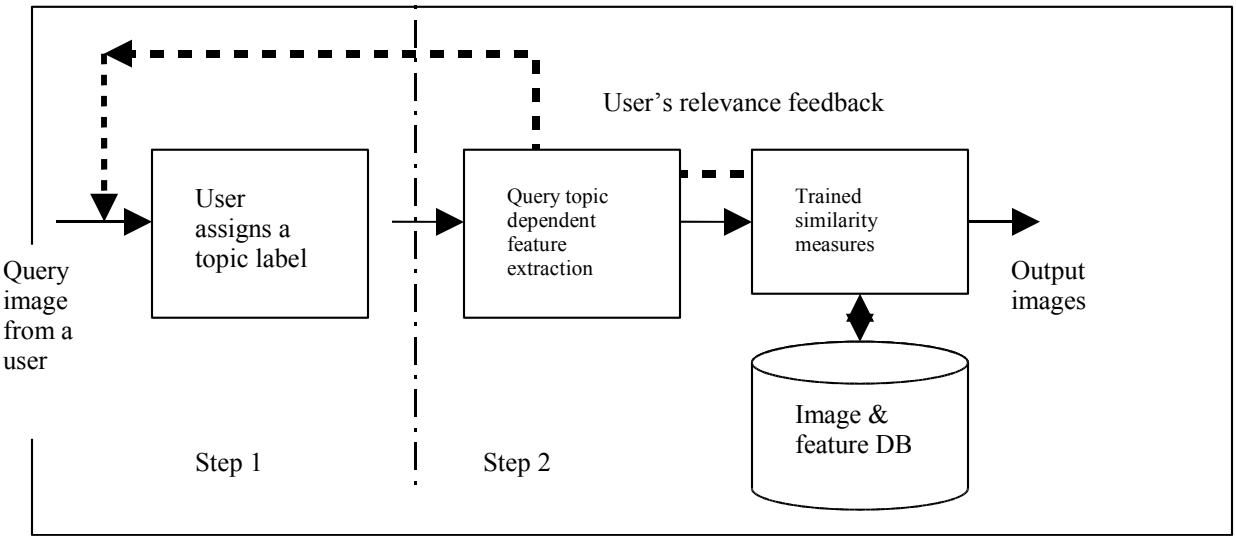

Figure 4. QTDIF-based medical image retrieval is performed in two steps: a human determines a topic class for a given query image, then computer makes use of the topic label and QTDIF feature to perform image retrieval

3. LOCAL REGIONAL FEATURES AND THEIR REPRESENTATIONS

As the medical collection contains many images and a large variety of image types, we need to employ color, texture and shape information as features to represent images. Besides the imaging modality, one of the most important characteristics of medical images is the anatomic structure. Hence, in the current work local regional features instead of global features are extracted and utilized.

Blob representation [5] is a robust approach to representing local coherent regions in color and texture. In this approach, pixels in each image are assumed to obey a GMM in a joint color-texture-spatial feature vector space. The Expectation Maximization (EM) is used to estimate the model parameters. Homogeneous regions with similar feature properties are then grouped together and segmented out. The contour of each dominant region is decomposed into a series of harmonic ellipses through the ellipse Fourier transformation [6]. The geometrical parameters of these ellipses are used for image matching and retrieval. The first order harmonics are often used to visualize the structure of the regions inside the image centered at their ellipse centers with their mean colors rendered respectively.

3.1. Texture Features

Give any image $I(x,y)$, its gradient $\nabla I(x,y)$ at (x,y) is a column vector $\nabla I = (\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y})^T$ measuring the local changes in color intensities. For color images, the L*a*b* color space is chosen and only the L* channel is treated. We consider that the noise in the image $I(x,y)$, $\nabla I(x,y)$ will be smoothed by a Gaussian kernel $G_\sigma(x,y)$ with variance σ^2. In order to process the image adaptively, the Gaussian kernel will be different at each pixel, i.e., $\sigma = \sigma(x,y)$. The optimal value of $\sigma(x,y)$ at pixel (x,y) can be found by examining the convergence of the local polarity with respect to monotonically increasing scales. The polarity is to measure the extension of points in a certain neighborhood whose gradient vectors are in the same directions [5]. With respect to the local dominant orientation within the window, and for a fixed scale and a given pixel position (x,y), it can be defined by

$$p_\sigma = \frac{|E_+ - E_-|}{E_+ + E_-} \quad (1)$$

where

$$E_+ = \sum_{(x,y)} G_\sigma(x,y)\left[\nabla I \cdot \vec{n}\right]_+ ,$$

(2)

and

$$E_- = \sum_{(x,y)} G_\sigma(x,y)\left[\nabla I \cdot \vec{n}\right]_- .$$

(3)

Here, $[.]_+$ and $[.]_-$ are the rectified positive and negative parts of their argument and $\vec{n}$ is the unit vector perpendicular to the dominant orientation [5]. The sums in the above equations are made over the given window defined by $G_\sigma(x,y)$. p_σ is positive and ranges from 0 to 1.

Now, we link $\sigma(x,y)$ with the scale order k by $\sigma_k = s/2$, $s = 0,1,...,7$, and produce a series of polarity images which are smoothed by a Gaussian with standard deviation $2\sigma_k$ to yield $p_{\sigma_k}(x,y)$. For each pixel (x,y), the scale $\hat{\sigma}(x,y)$ is selected as the first value of $\sigma_k(x,y)$ for which the difference between successive polarities is less than 2%. The local dominant orientation is the same as that of the eigenvector corresponding to the larger eigenvalueλ_1 of the two eigenvalues (λ_1 and λ_2) defined by the 2-by-2 smoothed second moment matrix

$$M_\sigma(x,y) = G_\sigma(x,y) * (\nabla I)(\nabla I)^T . \quad (4)$$

Once the particular scale $\hat{\sigma}(x,y)$ is selected, the anisotropy and the contrast are calculated by$\tau = 1 - \lambda_2/\lambda_1$ and $c = 2\sqrt{\lambda_1 + \lambda_2}$, respectively. Here $0 \le \tau \le 1$ and $0 \le c \le 1$. If the mean contrast within a region across scales is less than 0.05, the region is considered uniform.

3.2. Feature Probabilities

Assuming there is a total of J pixels within the image $I(x,y)$. Of each pixel, indexed by j in an image, $j = 1,...,J$, its position (x,y), its three color components and its texture descriptors p_σ, τ, c form a $d = 8$ dimensional feature vector z_j in R^d. A random variable Z used to describe these feature vectors z_j over the image is assumed to obey the K-component mixture Gaussian model

$$f(Z \mid \Theta) = \sum_{k=1}^{K} \beta_k \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_k)}} \exp[-\frac{1}{2}(Z-\mu_k)^T \Sigma_k^{-1}(Z-\mu_k)]$$

(5)

Here, for the k-th component, $\beta_k \ge 0$ is the weight subject to $\sum_k \beta_k = 1$, μ_k and Σ_k are the mean vector and the covariance matrix respectively, $\Theta = \{\theta_k\}_{k=1}^K$ is the collective parameter set with $\theta_k = (\beta_k, \mu_k, \Sigma_k)$. The EM algorithm is then used to obtain the maximum likelihood estimation ofθ :

$$\hat{\Theta} = \arg\max_{\Theta} \{\log \prod_{j=1}^{J} f(z_j \mid \Theta)\} . \quad (6)$$

We use the Minimum Description Length principle [7] to choose the number of mixture components K.

3.3. Region Grouping and Image Segmentation

Given θ_i, the conditional probability to which a feature vector z_j (corresponding to (x, y) in the image) belongs is

$$f_k(z_j|\theta_k) = \beta_k \frac{1}{\sqrt{(2\pi)^d \det(\Sigma_k)}} \exp[-\frac{1}{2}(z_j - \mu_k)^T \Sigma_k^{-1}(z_j - \mu_k)],\ k = 1,...,K\ . \quad (7)$$

The labeling of z_j in the feature space (i.e., pixel (x, y) in the image $I(x, y)$) to one of the K classes, $\hat{i}(z_j)$, is a naïve Bayesian decision problem:

$$\hat{k}(z_j) = \arg\max_k \{f_k(z_j|\theta_k)\}\ . \quad (8)$$

Grouping together those spatially connected pixels with the same respective labels $\hat{k}(z_j)$, $\hat{k}(z_j) = 1,...,K$, yields segmented homogeneous regions in $I(x, y)$.

3.4. Local Regional Features

The segmented regions are processed and their boundaries are followed to obtain the respective contours. For each region Ω its contour is decomposed by the ellipse Fourier transformation [6]. This processing produces a series expansion of the ordered positions (x, y) composed by the first levels of harmonics (ellipses) with lowest frequencies. For each level of ellipse, its center $(\bar{x}, \bar{y})$, the lengths of its semi- major and minor axes, and the orientation of the major axis are recorded. If we only consider those of the first level decomposition of Ω, together with the 8 mean values of the respective 8 components of z_j over Ω, there will be 11 local regional features used to represent Ω. We denote them as a feature vector φ, which can be considered as a middle level feature approximately representing both locations and shapes of uniform regions of the image.

4. MULTICLASS SUPPORT VECTOR MACHINES FOR IMAGE RETRIEVAL

SVM is a method widely used for statistical learning, classifiers and regression model design. It is used here to define QTDIF, the associated "models" and to produce similarity measures for medical image retrieval. Primarily SVM tackles the binary classification problem. The objective is to find an optimal separating hyper-plane (OSH) that correctly classifies feature data points as much as possible and separates the points of two classes as far as possible. The approach is to map the training data into a higher dimensional (possibly infinite) space and formulate a constrained quadratic programming for the optimization. Different mappings construct different SVMs.

Generally speaking, for a linear problem, let $g(\boldsymbol{\eta}) = \mathbf{w}^T \boldsymbol{\eta} + b$ be a discriminant function to separate the two classes in question. Here $\boldsymbol{\eta}$ is a feature vector, $\mathbf{w}$ is the weight vector, and b is the bias. Finding the OSH is to maximize $2/\|\mathbf{w}\|$, the margin of separation of nearest samples between the two classes. The margin can be seen as a measure of the generalization ability: the larger the margin, the better the generalization is expected to be. Given training examples $\boldsymbol{\eta}_i \in R^n$ from the two classes, $i = 1,...,N$, and a class label $d_i \in \{-1, 1\}$ for each $\boldsymbol{\eta}_i$, the OSH corresponds to the boundary that minimizes $\Phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i$ subject to $d_i(\mathbf{w}^T\boldsymbol{\eta}_i + b) \geq 1 - \xi_i,\ \xi_i \geq 0$. Here ξ_i is a measure of deviation of data from the margin and parameter C controls the tradeoff between the minimization of classification errors and maximization of margin. To solve this, the Lagrange function can be constructed as

$$J(\mathbf{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\mathbf{w}^T\mathbf{w} + C\sum_{i=1}^{N}\xi_i - \sum_{i=1}^{N}\alpha_i\left[d_i(\mathbf{w}^T\boldsymbol{\eta}_i + b) - 1 + \xi_i\right] - \sum_{i=1}^{N}\mu_i\xi_i\ , \quad (9)$$

where α_i, μ_i are positive Lagrange multipliers. The optimization is achieved to minimize $J(\mathbf{w}, b, \xi)$ with respect to $\mathbf{w}$ and b, and to maximize $J(\mathbf{w}, b, \xi)$ with respect to $\boldsymbol{\alpha}$. For the minimization of $J(\mathbf{w}, b, \xi)$, we have

$$\mathbf{w} = \sum_{i=1}^{N} \alpha_i d_i \eta_i \tag{10}$$

subject to $\sum_{i=1}^{N} \alpha_i d_i = 0$, and $C - \alpha_i - \mu_i = 0$. For the maximization of $J(\mathbf{w}, b, \xi)$, using these results just obtained, it can be rewritten as,

$$Q(\boldsymbol{\alpha}) = \sum_{i=1}^{N} \alpha_i - \sum_{i=1}^{N} \alpha_i \left[d_i (\mathbf{w}^T \eta_i + b) - 1 + \xi_i \right] - \sum_{i=1}^{N} \mu_i \xi_i \tag{11}$$

subject to the constraints $\sum_{i=1}^{N} \alpha_i d_i = 0$, and $0 \le \alpha_i \le C$. Solving the above equations we find that $\alpha_i = 0$ for most feature vectors $\boldsymbol{\eta}_i$, except that those feature vectors closest to the OSH, called support vectors, contribute a non-zero α_i to the equations.

As for the non-linear SVM, a so-called inner-product kernel is introduced to map input feature vectors into a higher-dimensional feature space nonlinearly. The mapped features are then put into the above linear SVM method and the OSH will be calculated based on this mapping.

The above SVM is for two-class problems. SVM for multiple-class classification are still a research problem. So far, several approaches have been proposed. One type has been to incorporate multiple class labels directly into the quadratic solving algorithm [8,9,10]. Another more popular type is to combine several binary classifiers: One vs. One (OVO) applies pair-wise comparison between classes [11]; and in One vs. All (OVA), one class is compared with the others [12]; Directed Acyclic Graph (DAG) is similar to OVO in the training stage, but in the testing stage it uses a negative logic and a tree structure [13].

In our case, 26 classes are needed but in fact more classes are contained in the dataset. So a multi-class SVM will be employed. For the simplicity of computation, we use the logic of OVA. Besides this, as the feature vectors based on the regional properties are in a high dimensional space, non-linear mapping with the Gaussian kernel is introduced. In practice, 26 models are constructed in the SVM training procedure. Given an unknown class feature vector, all of the 26 models are compared with it and the respective 26 prediction values (corresponding to the signed distance of the feature from the OSH) are generated. These prediction values are similarities between the features and the model (topic) compared. In this way, all images in the database are examined. For a particular model, which contains the QTDIFs and the classifier information. Feature vectors compared to it are sorted according to the predication. The 1000 feature vectors corresponding to the 1000 largest predication values are chosen and considered as the 1000 best retrieved results.

5. EXPERIMENTAL RESULTS

This section describes the performances of the results for the gift system as well as our new approach and compares the two very different techniques.

5.1. Results of QTDIF

To test the performance of QTDIF, images in the Casimage database are preprocessed. For all images, its non-trivial local regions are segmented. In terms of the areas of the regions, they are sorted and only the 10 largest regions are considered. Their blob representations are obtained; see Fig. 5 for a few example images and their respective blob representation visualized. Take the example of the hand bone shown in Fig. 5, the thumb is represented by three connected ellipses; the wrist, one; the palm, two; and the regions between fingers by three (dim) ellipses as well. The structure of the blobs resembles the hand bone quite well. It shows that, the blob approximation is a good middle level representation of images. Following is the computation of regional features of each selected region Ω_m, $m = 1,...,10$, its local feature vector $\boldsymbol{\varphi}_m$ is computed. Thus, there are $n = 10 \times 11$ features in the ten selected regions. They form a single feature vector $\boldsymbol{\eta}$ in the 110-dimensional vector space R^n where the SVM works.

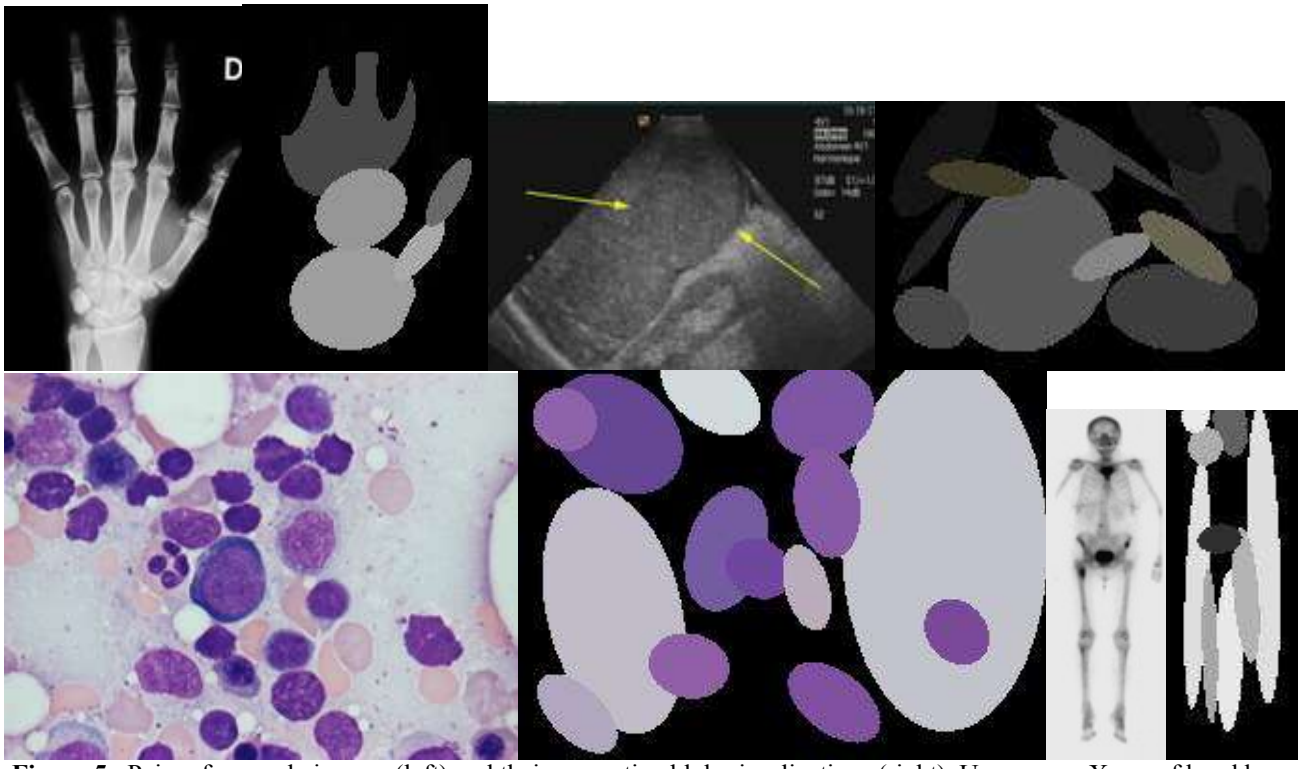

Figure 5. Pairs of example images (left) and their respective blob visualizations (right). Upper row: X-ray of hand bone and ultrasound scan; lower row: cells and skeleton scintigraphy.

As shown in Fig. 3, given a topic label q (i.e., an image representing the topic), visually similar images from the database can be picked up manually to form a training set for q. As there are a limited number of topics in the database, one can always form their respective training sets according to the assigned topics. To examine the performance of QTDIF proposed in this work, the same 26 topics used in imageCLEF 2004 are chosen. The feature vector $\boldsymbol{\eta}$ of each image of all topics is fed into the multi-class SVM algorithm for *topic* classification, where, in all our work with SVM, the Gaussian kernel is selected for non-linear feature mapping. Note that this *topic* classification is different from the *image* classification in general because a single image can be assigned to more than one topic training set. Therefore, this classification is done in a non-separable feature space in essence. The training results in 26 models containing query topic dependent features (those support vectors) as well as the decision boundaries for each of the 26 topics.

Given a topic from the 26 topics in the test phase, the feature $\boldsymbol{\eta}$ generated from every image I in the database will be compared with the model associated with this topic by SVM. This is done in the SVM testing phase where the OVA logic is applied. The signed distance of $\boldsymbol{\eta}$ from the decision hyperplane is associated with a prediction confidence parameter ν. A positive prediction strength corresponds to a test sample being assigned to a single class rather than to the ''all other'' class. The larger the prediction, the further the sample from the boundary is. Thus sorting the distances $\nu(I)$ of all images I in the database generates the top expected retrieval results. The retrieval performance is evaluated by using the standard evaluation program in the imageCLEF medical track [14]. The mean average precision (MAP) [17], which is used as the major performance measure in imageCLEF is defined as the mean of the average precision scores of each of the individual topics in the run. Geometrically, MAP is the area underneath a non interpolated recall-precision curve. The resulting MAPs are plotted in Fig. 6. The average MAP is 0.336. Some of them are tabulated in Table 1.

Table 1. MAP and breakeven values of the 26 topics resulting from the current approach.

Topic	3	4	7	8	9	10	11
MAP	0.2346	0.2649	0.8342	0.2159	0.4107	0.1347	0.4200
Topic	13	14	16	20	21	23	
MAP	0.2354	0.3239	0.6854	0.3191	0.2747	0.4381	

5.2. Results of medGIFT

medGIFT [18] participated in the task in various submissions and using a large variety of techniques from simple one-shot queries to automatic query expansion and manual relevance feedback, and all this also in combinations of visual and textual retrieval. In this part we only take into account the one-shot query of the medGIFT base configuration, which is actually the best one-shot configuration. Note that, features used in the medGIFT are not chosen based on the topics. The MAP for this system averaged over all topics is .0.3757, which is among the best five systems. When analyzing the single queries, it can be seen that depending on the topic, the results were extremely varied. The best query had a MAP of 0.9556 whereas the worst query had a MAP of 0.0038. For comparison, together with those of the QTDIF, Fig. 6 also shows the distribution of the MAP results with respect to the topics.

5.3. Comparison

Comparing our results with those of medGIFT, we find that some of our results are much better. For instance, for topic 11 (human skeleton scintigraphy shown in Fig. 5 lower right), medGIFT only retrieves it with MAP=0.37%, whereas QTDIF presents MAP=42%. On topic 23, the ultrasound images without colored parts in the middle (see Fig. 5 upper right), we achieve MAP=43.81% while medGIFT MAP=9.46%. This is perhaps due to the fact that images in both topics are of poor contrast and very noisy. However, on some topics, the current system is not as good as medGIFT, for example, topics 2, 15, 24, 25, 26. Topics 2 and 15 would be represented quite alike; both are axial CT images of the lung. For inexperienced people, it is highly possible to put them in wrong training sets. On topics 24, 25 and 26, we could still achieve MAP>40%. This shows that the parameters in blob features could be further tuned to improve the results.

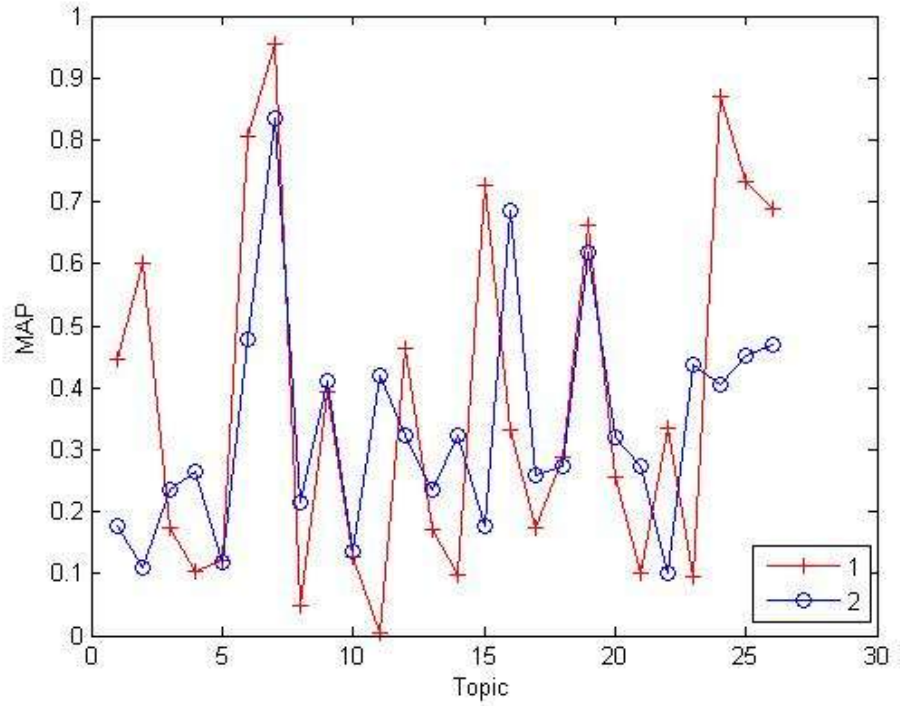

Figure 6. The MAPs of all topics resulted from both the medGIFT system (curve 1) and QTDIF (curve 2).

6. DISSCUSSION AND CONCLUSION

Differently from general image collections, medical images have their own unique characterization. Among other characteristics one feature of medical images is that in general there are a limited number of image categories, which depend on several factors: body parts, view positions and image modalities. Based on this very fact, we have proposed a novel image feature for content-based medical image retrieval: the proposed features are dependent on specific image query topics. Therefore, these feature sets capture both inter-category and intra-category statistical variations to achieve good performance of recall and precision.

Preliminary experimental results show that the Blob representation is a powerful middle level feature to capture structural characteristics of medical images. It provides good performance for several topics.
In the future we will investigate possible grouping methods to cluster several topics together and then for each group to choose appropriate sets of features for CBIR in order to achieve good performance. One possibility is to group the features together for a classification over several layers. This can mean to first identify the modality of the images (xray, CT, MRI, …) and then, within the modalities identify anatomic region and viewing angle. In connection with a confidence score for this sort of "annotation" it can help us to further train our classifiers and achieve better performance through a multi-step approach. The limit of our current approach is that the database contains many more topics than were used for evaluation in imageCLEF. Only with high quality training data good results can be obtained and so we need to test our system on a larger number of classes. Participation in the 2005 imageCLEF is planned to be able to judge the system performance with unknown class images and to be able to compare the approach with other systems under the same conditions.
Another goal is to integrate relevance feedback into the system as only relevance feedback queries can well describe the user's information need. In connection with our classifiers, an optimal relevance feedback strategy will need to be found.

REFERENCES

1. A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, R. Jain, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **22**(12), pp. 1349-1380, 2000.
2. A. M. Aisen, L. S. Broderick, H. Winer-Muram, C. E. Brodley, A. C. Kak, C. Pavlopoulou, J. Dy, C. R. Shyu, and A. Marchiori. "Automated storage and retrieval of thin-section CT images to assist diagnosis: System description and preliminary assessment," *Radiology*, **228,** pp. 265-270, 2003.
3. T. M. Lehmann, M. O. Güld, C. Thies, B. Fischer, K. Spitzer, D. Keysers, H. Ney, M. Kohnen, H. Schubert, B. B. Wein, "Content-based image retrieval in medical applications, *Methods of Information in Medicine*,**43**, pp 354-361, 2004.
4. H. Müller, A. Rosset, J.-P. Vallée, F. Terrier, A. Geissbuhler, "A reference data set for the evaluation of medical image retrieval systems," *Journal on Computerized Medical Imaging and Graphics*, **28**, pp. 295-305, 2004.
5. C. Carson, S. Belongie, H. Greenspan, J. Malik, "Recognition of images in large databases using color and texture," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **24**(8), pp. 1026-1038, 2002.
6. Frank P. Kuhl, Charles R. Giardina, "Elliptic Fourier Features of a Closed Contour," *Computer Graphics and Image Processing*, **18**, pp. 236-258, 1982.
7. R. O. Duta, P. E. Hart, and D.G. Stork, *Pattern Classification, 2nd ed.,* John Wiley & Sons, Inc., 2001.
8. J. Weston and C. Watkins, *Multi-class support vector machines*. Technical Report CSD-TR-98-04, Department of Computer Science, Royal Holloway, University of London, Egham, TW20 0EX, UK, 1998.
9. J. Weston and C. Watkins, "Support vector machines for multiclass pattern recognition," in *Proceedings of the Seventh European Symposium On Artificial Neural Networks*, April 1999.
10. K. Crammer, and Y. Singer, "On the learnability and design of output codes for multi-class problems," *Computational Learning Theory*, **47**(2-3) pp. 201-233, 2002.
11. K.S. Goh, E. Chang, K.T. Cheng, "Support vector machine pairwise classifiers with error reduction for image classification," *Proceedings of the 2001 ACM workshops on Multimedia: multimedia information retrieval*, Ottawa, Ontario, Canada, 2001.
12. T. Joachims, "Text Categorization with Support Vector Machines: Learning with Many Relevant Features," *Proceedings of the European Conference on Machine Learning*, Springer, 1998.
13. J.C. Platt, N. Cristianini, J. Shawe-Taylor, "Large margin DAG's for multi-class classification," *Advances in Neural Information Processing Systems 12*, Cambridge, MA: MIT Press, **12**, pp. 547-553, 2000.
14. Paul Clough, Mark Sanderson and Henning Müller, "The CLEF Cross Language Image Retrieval Track (ImageCLEF) 2004," *Springer Lecture Notes in Computer Science,* - to appear.
15. A. Pinhas and H. Greenspan, "A continuous and probabilistic framework for medical image representation and categorization," *Proceedings of SPIE* **5371**, pp. 230-238, 2004.
16. Casimage web site: http://www.casimage.com/
17. E. M. Voorhees, D. Harman, "Overview of the Sixth Text Retrieval Conference (TREC-6), *Information Processing and Management*, **36**, pp. 3–25, 2000.
18. H. Müller, A. Rosset, J.-P. Vallée, A. Geissbuhler, Integrating Content-Based Access Methods into a Medical Case Database, Proceedings of *Medical Informatics Europe* (MIE 2003), pp. 480-485, St. Malo, France, 2003.