Medical Image Retrieval with Relevance Feedback via Pairwise Constraint Propagation

Menglin Wu[1,2], Qiang Chen[1], Quansen Sun[1]

[1]School of Computer Science and Engineering, Nanjing University of Science and Technology
Nanjing 210094, China
[e-mail: wumenglin8263@hotmail.com]

[2]College of Electronics and Information Engineering, Nanjing University of Technology
Nanjing 211816, China
[e-mail: wumenglin@njut.edu.cn]

*Corresponding author: Menglin Wu

Received November 8, 2013; revised December 25, 2013; accepted January 4, 2014; published January 29, 2014

Abstract

Relevance feedback is an effective tool to bridge the gap between superficial image contents and medically-relevant sense in content-based medical image retrieval. In this paper, we propose an interactive medical image search framework based on pairwise constraint propagation. The basic idea is to obtain pairwise constraints from user feedback and propagate them to the entire image set to reconstruct the similarity matrix, and then rank medical images on this new manifold. In contrast to most of the algorithms that only concern manifold structure, the proposed method integrates pairwise constraint information in a feedback procedure and resolves the small sample size and the asymmetrical training typically in relevance feedback. We also introduce a long-term feedback strategy for our retrieval tasks. Experiments on two medical image datasets indicate the proposed approach can significantly improve the performance of medical image retrieval. The experiments also indicate that the proposed approach outperforms previous relevance feedback models.

Keywords: Content-based medical image retrieval, manifold ranking, relevance feedback, pairwise constraint propagation, long-term feedback

http://dx.doi.org/10.3837/tiis.2014.01.015

1. Introduction

Nowadays medical images play an essential role in disease diagnosis, surgical planning, treatment evaluation, as well as medical research and education. With rapid advances in digital imaging technologies, images in diverse modalities, such as X-ray, computer tomography (CT), magnetic resonance imaging (MRI), and ultrasound, present important anatomical and functional information. Efficient searching of large image collections becomes a necessity due to the tremendous increase in the number of biomedical images.

Content-based medical image retrieval [1], which queries relevant medical images based on their visual content similarity, is one of the promising solutions to manage image datasets effectively. Recently, several prototypes have been proposed for searching biomedical images with different modalities. The boosting-based metric learning algorithm seeks to preserve both visual and semantic similarities in medical retrieval procedures [2]. Bag of words (BOW)-based approaches are adopted for effective X-ray image retrieval [3]. Considering related diseases have similar solutions, medical case retrieval is suggested by heterogeneous information fusion [4]. In addition, support vector machine (SVM)-based frameworks are also popular in medical image retrieval system during image filtering and dynamic features fusion [5]. The core of the aforementioned methods is the effectiveness of low-level features extracted from medical images, such as texture [6], BOW [3] or fusion features [5]. However, the low-level features may not be able to characterize the medical sense of the images, known as semantic gap.

To narrow down the semantic gap, relevance feedback is a powerful tool, which improves the performance of the image retrieval system [7]. This interactive approach allows users to mark the previously retrieval images as either relevant or irrelevant at each query iteration. Based on the feedback, the system can achieve more accurate searching and produce a re-ranking result for the next query. Various relevance feedback techniques have been proposed in the last decade. Most of short-term learning feedback strategies use the feedback information only within the current query session. Other methods are proposed for long-term feedback, which exploits past user feedback to refine future queries [8-9].

From a machine learning viewpoint, the relevance feedback can be treated as a classification problem, which classifies images as positive and negative samples based on user feedback. However, there are two problems in this learning scenario: small sample size and asymmetrical training. The small sample size is due to the fact that only a small number of images are labeled by user feedback at each query. Therefore, learning from a small sample size is difficult. The imbalance between positive and negative samples causes the asymmetrical training problem. In practice, the number of relevant images are often much smaller than that of the irrelevant images in the image set. Furthermore, the relevant images

belong to the same target class while the irrelevant images may belong to different semantic classes in various distributions. So the positive and negative feedbacks should not be treated equally.

In this paper, we proposed a novel interactive medical image retrieval framework with relevance feedback based on pairwise constraint propagation, which takes into account of the above small sample size and asymmetrical training. Motivated by the idea of pairwise constraint propagation [10-12], the proposed approach integrates pairwise constraint information from relevance feedback in manifold ranking-based retrieval. Specifically, we first convert user feedback to must-link and cannot-link constraints and propagate them to the entire image set. After the results of the propagation are applied to reconstruct the similarity matrix, the manifold learning-based image ranking can be implemented based on this new similarity matrix.

The proposed approach has the following three advantages. Firstly, unlike traditional manifold ranking methods concerning only manifold structures, the proposed approach obtains pairwise constraints from the feedback in order to improve the retrieval performance. In addition, it treats the relevance feedback procedure as a semi-supervised learning problem. Therefore, the small sample size is not an issue. Secondly, we introduce a biased image ranking scheme that processes relevant and irrelevant images differently for tackling the asymmetrical training in image ranking procedure. Thirdly, the proposed approach is easy to implement long-term learning feedback strategy. Future retrieval can be improved by logging the data relating to similar and dissimilar relations between images in past relevance feedback rounds.

The rest of this paper is organized as follows. Section 2 describes the summary of relate works. In section 3, we introduce the proposed interactive medical image retrieval framework and solutions to the two problems mentioned above in relevance feedback. Section 4 details the experiment results and analysis. The last section presents the conclusion.

2. Related Works

Relevance feedback is promising to improve the performance of content-based image retrieval. The idea is how to incorporate the positive and negative samples to refine the query and how to adjust the similar measures according to user feedback [13]. In this section, we summarize relevant previous works in this area.

The first is query point movement with the goal to construct a new point close to relevant results and far from irrelevant results. It can be illustrated by Rocchio's formula, implemented in many content-based image retrieval systems [14]. Feature re-weighting is the other method within this category [15] by adjusting the similarity weights in different feature spaces by user feedback. Although the above methods run relatively quickly and can

considerably improve the retrieval performance, they ignore the dependence of image features and treat features globally [16].

Other studies suggest relevance feedback is a binary-class classification problem. Many machine learning algorithms deal with this problem, among which, SVM is the most popular. SVM attempts to find the hyperplane that can achieve maximum separation between relevant and irrelevant images [17]. Biased-SVM and other methods overcome the limitations of standard SVM [18-19]. Another idea regarding relevance feedback is to create a subspace where the relevance images project closer together yet further away for the irrelevance images. Algorithms based on this idea include linear discriminant analysis (LDA) [20], biased discriminant analysis (BDA) [21] and manifold learning [22]. Probabilistic models are also commonly used in feedback procedure, which estimate posterior probability distribution according to user feedback. Bayesian approach [23], nearest neighbor methods [24-25], or mixture models [26] are within this field. Other related works include self-organizing maps (SOM) [16], fuzzy sets [27], or metric learning [28].

Recently, other research has proposed semi-supervised learning models for relevance feedback. In these models, investigation of both labeled images from feedback and unlabeled images results retrieval performance improvement. An enhancing relevance feedback framework based on the tri-training algorithm is another approach [29]. Biased maximum margin analysis integrating unlabeled samples is used to find the appropriate hyperplane for SVM-based feedback schemes [30]. Manifold ranking-based algorithms are applied to learn the manifold structure from user feedback and sort the relevant images by ranking score [31-32].

Manifold ranking has proven effective in the content-based image retrieval; however, it involves a short-term feedback strategy and ignores the pairwise constraints from feedbacks. Inspired by pairwise constraint propagation algorithms introduced for spectral clustering [10-12], we propose a novel relevance feedback method for interactive medical image retrieval. We convert user feedback to pairwise constraints and propagate these constraints to the whole image set to improve the manifold structure for manifold ranking-based image retrieval. Moreover, in the proposed approach we also resolve the problems in traditional feedback procedures and introduce a long-term feedback strategy.

3. The Proposed Approach

We propose a two-step relevance feedback for medical image retrieval. The first step is to convert user feedback to cannot-link and must-link constraints, and propagates these two category constraints to the whole image set. The result of the propagation is then applied to reconstruct the similarity matrix, which is the basis of the manifold ranking framework. The second step is to propose a biased image ranking scheme to tackle the asymmetrical training in feedback procedures. Furthermore, we introduce the proposed interactive medical image

retrieval framework with a long-term feedback strategy.

3.1 Relevance Feedback based on Pairwise Constraint Propagation

During the relevance feedback process of medical image retrieval, experts label several images according to their relevance to a query in the medical sense. Assume that $\chi = \{x_1, ..., x_n\}$ is the medical image set. Let P and N be the relevant sample set and irrelevant sample set marked by the user in relevance feedback. We then define

$$\begin{aligned} M &= \{(x_i, x_j) \mid x_i, x_j \in P\} \\ C &= \{(x_i, x_j) \mid x_i \in P \vee x_j \in N\} \end{aligned} \tag{1}$$

The data pair (x_i, x_j) in M can be seen as a must-link since the two images are the same query concept, while in C can be treated as a cannot-link because they belong to different classes.

After we obtain the pairwise constraints from user feedback, we can propagate these two category constraints to the whole image set and refine the similarity matrix for image ranking. As shown in **Fig. 1**, the similarity between two image samples will increase if two images have a must-link constraint and decrease if they have a cannot-link constraint. The pairwise constraint propagation then influences the whole manifold structure denoted by similarity matrix. It is worth noting that this feedback approach exploits the unlabeled images in dataset and pairwise constraints on whole manifold structure, so it can well deal with the small sample size.

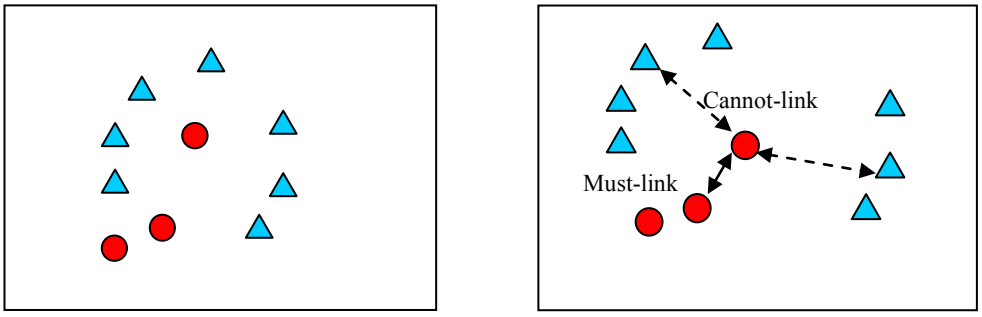


Fig. 1. An example of pairwise constraint propagation in one image query. (a) Red dots represent relevant images and blue triangles represent irrelevant images. (b) The whole manifold structure is influenced by the pairwise constraints.

For the purpose of pairwise constraint propagation, we first introduce some basic notions. Assume $G = (V, W)$ be an undirected weighted graph with its vertex set $V = \chi$ and

weight matrix $W=\{w_{ij}\}_{n\times n}$, where the edge w_{ij} denotes the similarity between x_i and x_j. To simplify the calculation, $w_{ij}=\exp(-\| x_i-x_j \|/(2\sigma^2))$ if x_j is the k-nearest neighborhood of x_i, and otherwise $w_{ij}=0$. We set $w_{ii}=0$ for $1\le i\le n$ to avoid self-reinforcement and let $W=(W+W^T)/2$ to ensure W is nonnegative symmetric. The normalized graph Laplacian L is represented as:

$$L=I-D^{-1/2}WD^{1/2} \tag{2}$$

where I is the identity matrix and $D=\{d_{ii}\}_{n\times n}$ is the diagonal matrix with $d_{ii}=\sum_j w_{ij}$.

We then represent an initial relation matrix $Y=\{y_{ij}\}_{n\times n}$:

$$y_{ij}=\begin{cases}+1 & (x_i,x_j)\in M\\ -1 & (x_i,x_j)\in C\\ 0 & otherwise\end{cases} \tag{3}$$

Furthermore, the matrix $F=\{f_{ij}\}_{n\times n}$ with $|f_{ij}|\le 1$ is defined as the final result of the pairwise constraint propagation. $f_{ij}>0$ means (x_i,x_j) is a must-link constraint while $f_{ij}<0$ means a cannot-link constraint. This problem can be solved by vertical and horizontal propagation procedure over the graph G [11].

Let $Y_{.i}$ be the *i*-th column of the initial matrix Y, while $F_{.i}$ is the *i*-th column of the result matrix F. It can be observed that $Y_{.i}$ is the initial label vector of a two-class semi-supervised learning problem on the graph G with respect to x_i, where the "positive class" denotes the must-link constraint and the "negative class" denotes the cannot-link constraint.

Based on regularized energy function minimization of semi-supervised learning in [33], the vertical pairwise constraint propagation with respect to x_i can be formulated as:

$$F_{.i}^{*}=\arg\min_{F_{.i}}\frac{1}{2}F_{.i}^{T}LF_{.i}+\frac{\mu}{2}\| F_{.i}-Y_{.i} \|^2 \tag{4}$$

where $\mu > 0$ is the regularization parameter. We also can rewrite (4) as a matrix form:

$$F_v^* = \arg\min_{F_v} \frac{1}{2}\mathrm{tr}(F_v^T L F_v) + \frac{\mu}{2} \| F_v - Y \|_F^2 \tag{5}$$

where $\mathrm{tr}(.)$ stands for the trace of a matrix , $\| . \|_F$ is the Frobenius norm, and F_v is the result matrix after vertical propagation. To solve the optimization problem, we differentiate with respect to F_v and set it to zero:

$$\mu(F_v^* - Y) + LF_v^* = 0 \tag{6}$$

So the solution is $F_v^* = \mu(\mu I + L)^{-1} Y$.

In horizontal propagation step, we calculate the result matrix F_h^* with using F_v^* as the initial matrix, so the graph regularization can be formulated similarly as:

$$F_h^* = \arg\min_{F_h} \frac{1}{2}\mathrm{tr}(F_h L F_h^T) + \frac{\mu}{2} \| F_h - F_v^* \|_F^2 \tag{7}$$

By the derivation of (7) and set it to zero, we can obtain the final constraint propagation result F :

$$\begin{aligned} F = F_h^* &= \mu F_v^* (\mu I + L)^{-1} \\ &= \mu^2 (\mu I + L)^{-1} Y (\mu I + L)^{-1} \end{aligned} \tag{8}$$

So F represents the pairwise constraints on the whole image set. According to [11], the reconstructed similarity matrix is defined as:

$$w_{ij}^* = \begin{cases} 1 - (1 - f_{ij})(1 - w_{ij}) & f_{ij} > 0 \\ (1 + f_{ij}) & f_{ij} < 0 \end{cases} \tag{9}$$

where $W^* = \{w_{ij}^*\}_{n \times n}$ is the similarity matrix after the reconstruction. Equation (9) shows the similarity between x_i and x_j , which will increase if two images have a must-link constraint and decrease if they have a cannot-link constraint. It is easy to find out that W^* is symmetric since both W and F are symmetric. Furthermore, $w_{ij} \in [0,1]$ and

$|f_{ij}| \leq 1$ ensures that W^* is nonnegative.

3.2 Image Ranking

In this subsection, we rank the images based on the reconstructed similarity matrix. This procedure can be modeled as a manifold ranking problem. Assume $G^* = (V, W^*)$ is the graph represented by the reconstructed similarity matrix. Let L denote the images labeled by the user and $y_L = \{y_1, y_2, ..., y_l\}^T$ represent their initial ranking score, while U denote the unlabeled images. $S = \{s_1, s_2, ..., s_n\}^T$ denotes the ranking score for each of the images in manifold ranking. According to graph-based semi-supervised learning theory [34], image ranking becomes the problem of minimizing the following energy function:

$$\begin{aligned} S^* &= \arg\min_S \sum_{i,j} (s_i - s_j)^2 w_{ij}^* \\ s.t. \quad & s_i \equiv y_i \quad (1 \leq i \leq l) \end{aligned} \tag{10}$$

This function here means that nearby images will have similar ranking score according to the graph represented by similarity matrix. The constraint requires that the initial score labeled by the user will not change in manifold ranking procedure.

Then we can rewrite formula (10) in matrix form as:

$$\begin{aligned} S^* &= \arg\min_S S^T P S \\ s.t. \quad & S_L \equiv y_L \end{aligned} \tag{11}$$

where $P = D^* - W^*$ and $D^* = \{d_{ii}^*\}_{n \times n}$ is the diagonal matrix with $d_{ii}^* = \sum_j w_{ij}^*$. Split the matrix P by image set L and U as follows:

$$P = \begin{bmatrix} P_{LL} & P_{LU} \\ P_{UL} & P_{UU} \end{bmatrix} \tag{12}$$

Differentiating with respect to S in (11) and setting it to zero, we obtain the ranking score for unlabeled images as $S_U^* = (I - P_{UU})^{-1} P_{UL} y_L$, which can be represented as iteration form in [34]:

$$S_U^{t+1} = P_{UU} S_U^t + P_{UL} y_L \tag{13}$$

where t is the iteration time.

The relevant and irrelevant images should be treated differently, due to the problematic

asymmetrical training in relevance feedback; the relevant images always belong to the same target class while the irrelevant images may belong to different semantic classes. To tackle this, we introduce a biased image ranking scheme for positive and negative feedback. Consequently, we first define the initial ranking score as the combination of the positive and negative score:

$$y_L = y_L^+ + \gamma y_L^- \tag{14}$$

The elements of y_L^+ are set to 1 if corresponding images are the query or relevant images, and other elements are set to 0; while the elements of y_L^- are set to -1 if corresponding images are irrelevant images and 0 otherwise. The parameter γ is defined as:

$$\gamma = \exp(-\frac{|N|}{|P|}) \tag{15}$$

where $|P|$ and $|N|$ are the number of relevant and irrelevant images, respectively. The parameter $\gamma \in (0,1]$ means that the relevant images are always more important than irrelevant images in manifold ranking procedure due to the asymmetric and unbalanced distribution, and thus the unlabeled images obtain more ranking score by positive score spreading than negative score. The number of irrelevant and relevant images also influences the importance of positive and negative feedback. For example, when the number of irrelevant images increases, the value of γ decreases and it leads to less ranking score spreading by each negative feedback.

For clarity, we summarize the interactive medical image retrieval approach in algorithm 1:

Algorithm 1. Medical image retrieval with relevance feedback

Input: query images, medical image dataset

1. Create graph $G = (V, W)$ by query images and medical images in the dataset.
2. Given a query image, get top K images by image ranking using (13).
3. Get user feedback and Update the relation matrix Y as (1) and (3).
4. Propagate pairwise constraints to the whole image set and reconstruct the similarity matrix W^* using (8) and (9).
5. Ranking the images according to the new graph $G^* = (V, W^*)$ by biased manifold

ranking strategy (13)-(15), and return to step 3 until the user is satisfied with the retrieval result.

Output: the retrieval result in descending order of ranking score.

3.3 Retrieval Framework with Long-term Feedback

Although the medical image retrieval system takes advantage of the relevance feedback from users, it ignores the information from past user feedback. Log data of accumulated user feedback can be used to improve the performance of relevance feedback in image query for a long-term learning purpose [8]. The log data is preserved as a matrix $R = \{r_{ij}\}_{m \times z}$. Each row of R denotes a query session, and each column of R denotes an image in the image dataset [8]. When an image labeled as relevant in a query, the corresponding element of R is set to 1. Similarly, the element is set to -1 when the image is judged as irrelevant and 0 when the image is not judged in a query session. We then obtain a relation matrix $Y_R = \{y_{R(ij)}\}_{n \times n}$ from R, whose size is as same as initial relation matrix Y:

$$y_{R(ij)} = \left(\sum_{k=1}^{m} \beta_{ki'j'} \cdot r_{ki'} \cdot r_{kj'} \right) / m_{ij} \tag{16}$$

where image i and j in Y_R are corresponding to i' and j' in R. m_{ij} is the number of the rows that satisfy $r_{ki'} \cdot r_{kj'} \neq 0$ and $\beta_{ki'j'} \neq 0$. The parameter $\beta_{ki'j'}$ is defined as:

$$\beta_{ki'j'} = \begin{cases} 1 & r_{ki'} + r_{kj'} \geq 0 \\ 0 & r_{ki'} + r_{kj'} < 0 \end{cases} \tag{17}$$

We do not consider the case that both of $r_{ki'}$ and $r_{kj'}$ are -1 since it is difficult to judge whether they belong to the same class when they both are labeled as irrelevant in a query session.

We then can rewrite the initial relation matrix as:

$$Y' = Y + \delta \bullet Y_R \tag{18}$$

where '$\bullet$' stands for Hadamard product. The element of δ is denoted as:

$$\delta_{ij} = \begin{cases} 0 & y_{ij} \neq 0 \\ 1 & y_{ij} = 0 \end{cases} \tag{19}$$

which means the relation of image i and j is determined by current query if the user has

labeled both of them, otherwise it is determined by historical relation matrix Y_R.

We have replaced the initial relation matrix Y with Y' in algorithm 1 (18). Therefore, our retrieval system can benefit from both past feedback and current user feedback in this long-term learning strategy. The diagram of the retrieval framework is shown in **Fig. 2**.

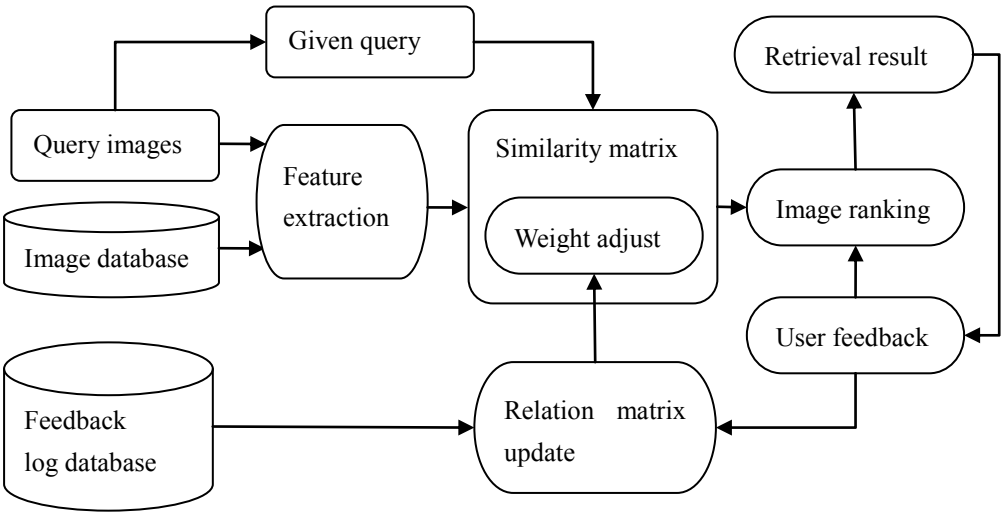


Fig. 2. The diagram of the retrieval framework with long-term feedback

4. Experiments and Analysis

4.1 Experiment Setup

To evaluate the performance of the proposed framework, we carried out exhaustive experiments on the Mammographic patches dataset and ImageCLEFmed2009 dataset. Aachen University of Technology provided both of the manually IRMA coded datasets [35]. This IRMA code describes image modality, the body orientation, body region examined and biological system examined of a medical image according to a mono-hierarchical coding scheme. A summary of the details of each dataset is follows:

- ImageCLEFmed2009: is a popular benchmark dataset for evaluating medical image annotation and retrieval. It consists of diverse models of X-ray images, which represent different ages, genders, view positions and pathologies from plain radiography. In the experiments, we reorganize 6110 images of 50 semantic classes from this dataset to ensure the number of each class is close and of sufficient size. X-ray images in this dataset are described by patch-based visual words [3].
- Mammographic patches: consists of texture patches from screen mammography composed of the DDSM, MIAS, LLNL and RWTH databases [36]. It includes 2796 patches of 12 categories classified by 4 kinds of tissue density and 3 kinds of tumor. We extract Gabor features, widely used in mammogram retrieval and classification [6], to obtain textual

information.

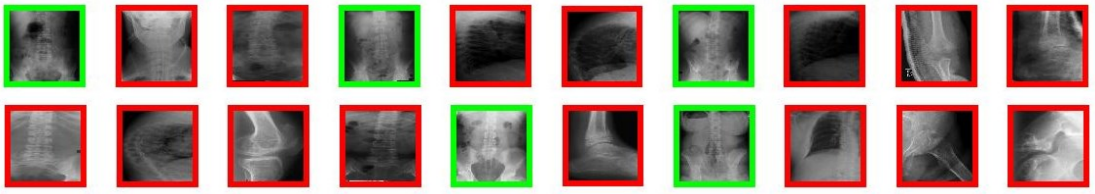
4.2 Comparative Results of Short-term Feedback

To evaluate the performance of the medical image retrieval system, we compare the proposed approach in this paper with two manifold ranking-based algorithms and four other relevance feedback methods. The manifold ranking-based algorithms were: (a) manifold ranking-based image retrieval algorithm (MRBIR) [31]; (b) random walker (RW)-based approach [32]. The relevance feedback methods were: (c) query point movement (QPM) using Rocchio's formula [14]; (d) active SVM-based algorithm [17]; (e) relevance score (RS) [24]; (f) improved nearest neighbor (INN)-based approach [25].

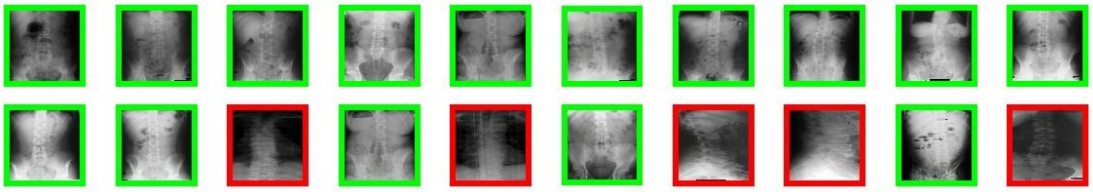
All the algorithms are implemented according to the original literature. In the case of SVM-based approach, we use RBF kernel and tune parameters appropriately. As for the proposed approach, we empirically set $K = 60$ and $\sigma = 0.1$ to computer the KNN similarity matrix. In addition, we set $\mu = 0.6$ for pairwise constraint propagation. Here we only evaluate algorithm 1 and do not use feedback log data. Therefore, we can view our feedback approach as a short-term feedback process.

In this experiment, where the query images are randomly sampled, we conduct 500 queries for each dataset. Scope size is defined as the number of return images for judgment in one query. We consider the scope size of 20 and 40 to evaluate different algorithms. At each round of relevance feedback, the images within this scope are labeled as positive and negative samples by the user. A query example on ImageCLEFmed2009 is shown in **Fig. 3**.

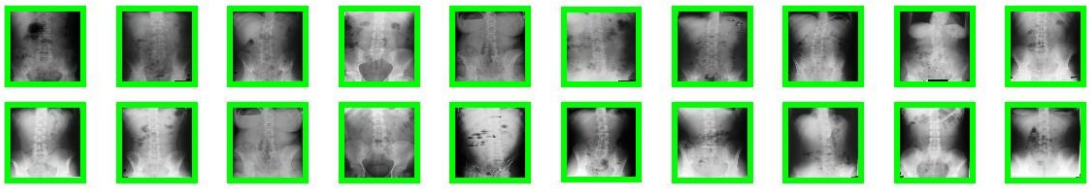
(a) Query image (IRMA code: 1121-115-700-400)



(b) Initial results (25%)



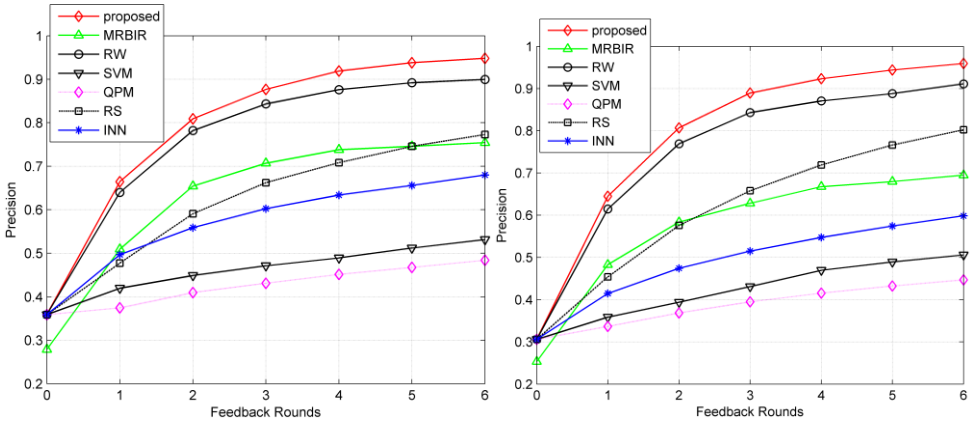
(c) Results after 1st round of relevance feedback (75%)



(d) Results after 2nd round of relevance feedback (100%)

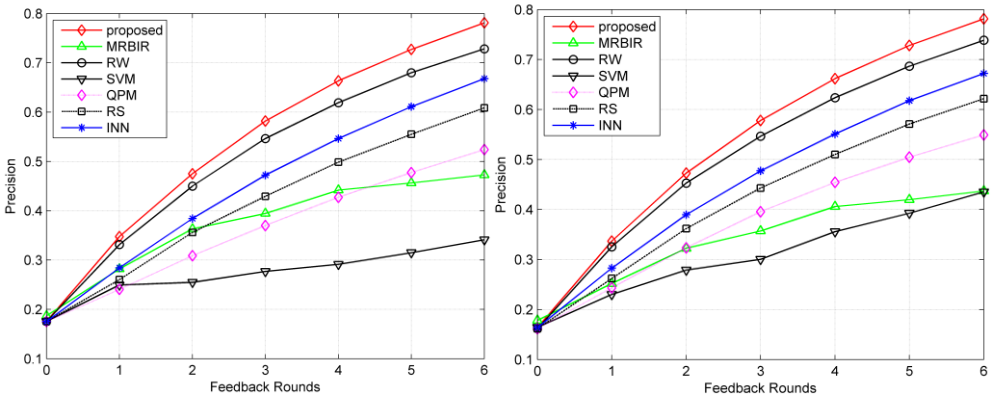
Fig. 3. An query example of the proposed method. Green framed images denote relevant images and red framed images denote irrelevant images. (a) The query image. (b) Initial results of image ranking. (c-d) Results after different rounds of relevance feedback.

For comparison, the image ranking approach in section 3.2 is applied to calculate the initial retrieval results. However, we do not use this approach for MRBIR, which obtains the results based on its own ranking scheme. Then, the top images are automatically selected and re-ranked at each round of the relevance feedback by different algorithms. The average precision is used to measure the performance of each algorithm within different scope sizes. These results are shown in **Figs. 4-5**.



(a) Scope size 20 (b) Scope size 40

Fig. 4. Average precision with different scope size measured on ImageCLEFmed2009 dataset.



(a) Scope size 20 (b) Scope size 40

Fig. 5. Average precision with different scope size measured on Mammographic patches dataset.

As these figures show, it is evident that the proposed approach explicitly outperforms the other traditional relevance feedback algorithms with different datasets and scope sizes. For example, on the ImageCLEFmed2009 dataset, the average precision of the results with scope size 20 and 40 achieves a 5.4% and 5.2% improvement over the RW algorithm after 6 iteration feedbacks. We also notice that QPM shows the worst results since it is influenced by the dependence between image features and may fail if the values of the features are continuous. The performance of the active SVM-based algorithm cannot achieve satisfactory performance due to the limited images labeled in the relevance feedback. It is difficult, when the number of feedback sample is small, to train an appropriate learning model. In comparison with RW and MRBIR, the proposed approach is not only concerned about the manifold structure of the dataset but also integrates the pairwise constraint information from user feedback to refine the image query. Therefore, the results demonstrate the proposed approach is better than the two manifold ranking-based image retrieval algorithms.

We obtained similar conclusions on the Mammographic patches dataset. However, the overall performances of the retrieval algorithms degraded even though this dataset was smaller than the first one. The reason was that the images in the second dataset had high intra-class variability and inter-class similarity. Despite of this, the proposed method still achieved the best result.

4.3 Comparison of Image Ranking Schemes

According to section 3.2, we introduce a biased image ranking approach that treats positive and negative feedback differently. Since the relevant images are more important than the irrelevant images in image retrieval, this approach focuses the ranking score on the positive feedback rather than on the negative feedback. In this section, we evaluate the performance of the proposed approach with two ranking schemes: (1) biased ranking introduced in our paper (referred to as BR); (2) unbiased ranking by setting initial score of positive and negative feedback sample 1 and -1, respectively (referred to as UR). We also report the results of MRBIR as another manifold ranking-based algorithm combining these two schemes. Mean average precision of the top 100 images is used to measure the retrieval performance. These results are shown in **Table 1 and 2**.

Table 1. Mean average precision of top 100 images measured on ImageCLEFmed2009 dataset.

Method	**Feedback rounds**					
	1	2	3	4	5	6
MRBIR-UR	0.353	0.507	0.577	0.604	0.620	0.635
MRBIR-BR	0.474	0.556	0.592	0.614	0.634	0.641
Proposed-UR	0.553	0.625	0.668	0.698	0.723	0.750
Proposed-BR	**0.646**	**0.767**	**0.826**	**0.847**	**0.860**	**0.869**

Table 2. Mean average precision of top 100 images measured on Mammographic patches dataset

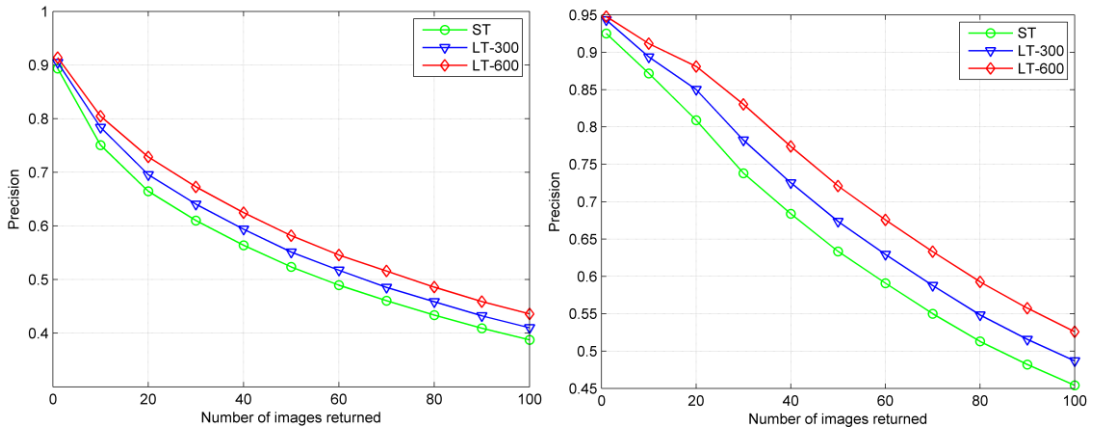
Method	Feedback rounds					
	1	2	3	4	5	6
MRBIR-UR	0.341	0.404	0.440	0.477	0.489	0.503
MRBIR-BR	0.336	0.423	0.479	0.505	0.522	0.531
Proposed-UR	0.366	0.453	0.517	0.571	0.616	0.653
Proposed-BR	**0.416**	**0.536**	**0.616**	**0.669**	**0.707**	**0.737**

We observed that the biased ranking scheme performed better than the unbiased ranking scheme on both datasets. Furthermore, we obtained the same conclusion from the results of MRBIR, although its overall mean average precision was lower. This is because of asymmetrical training where the number of relevant images is much smaller than the irrelevant images in image retrieval. The biased ranking scheme treats relevant images as more important than irrelevant ones and produces a higher score from positive feedback than from negative feedback. Thus, it can outperform unbiased ranking as expected.

4.4 Performance Evaluation of Long-term Feedback Strategy

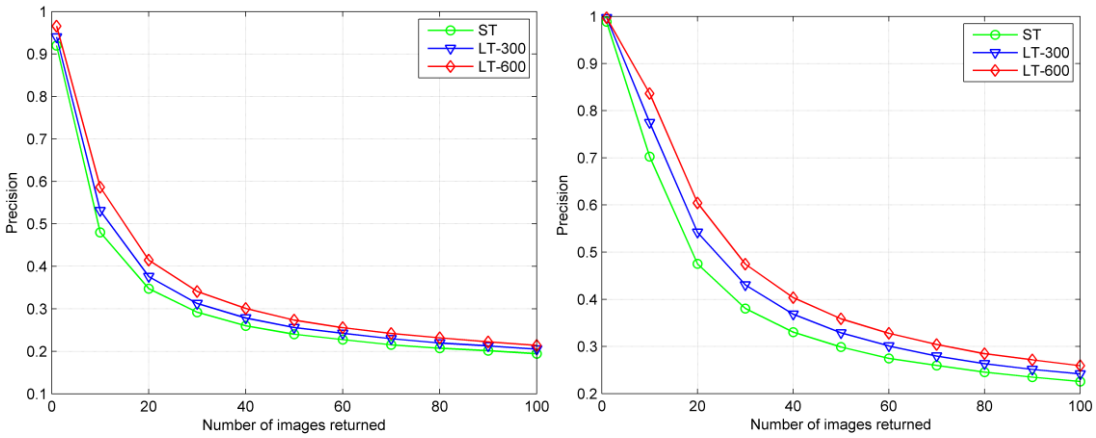
In order to verify the effectiveness of exploiting the past user feedback in medical image retrieval, we evaluate the long-term feedback strategy of the proposed approach. In this experiment, we collect the log data for recording historical feedback as mentioned in section 3.3. Specifically, we first randomly sample images from the dataset for query. In each query session, we label the top 20 returned images using ground truth to simulate user's feedback. Then, the feedback is converted to log data and recorded. For long-term feedback evaluation, we apply log data of 300 query sessions (LT 300) and 600 query sessions (LT 600) as compared to the short-term feedback strategy (ST). The average precision of the top 100 returned images after the 1st and 2nd rounds of feedback on two datasets are used to measure the different feedback strategies (**Fig. 6-7**).

We observed that the long-term feedback strategy achieved higher retrieval performance than the short-term strategy. Furthermore, as the number of historical query sessions increased, the relevance feedback showed an improvement in performance. For instance, LT-600 achieved a 5.42% and a 7.18% improvement in mean average accuracy over the LT-300 on ImageCLEFmed2009 and Mammographic patches datasets after two rounds of relevance feedback. This proves that historical feedback information can effectively enhance the performance of the proposed medical image retrieval approach by using a long-term feedback strategy.



(a) Precision after 1st round (b) Precision after 2nd round

Fig. 6. Average precision on ImageCLEFmed2009 dataset.



(a) Precision after 1st round (b) Precision after 2nd round

Fig. 7. Average precision on Mammographic patches dataset.

5. Conclusion and Discussion

In this paper, we proposed a novel interactive medical image retrieval approach based on pairwise constraint propagation. The proposed approach converts user feedback to pairwise constraints and refines the similarity matrix by constraint propagation. Based on the reconstructed similarity matrix, we introduced a biased image ranking scheme to tackle the asymmetrical training in relevance feedback. Furthermore, we proposed a long-term feedback strategy for the approach by exploiting historical feedback log data. Exhaustive experiments on two medical image datasets demonstrated that the proposed approach outperforms other relevance feedback algorithms. The results also showed the effectiveness of the proposed approach with the biased image ranking scheme and a long-term feedback strategy. Although this paper focuses on medical image retrieval, the proposed algorithm can

be viewed as a general retrieval framework and could apply to ordinary image retrieval task. In the future we will extend the retrieval approach presented in this paper, to incorporate heterogeneous medical information.

Acknowledgement

The authors would like to thank TM Deserno, Dept. of Medical Informatics, RWTH Aachen, Germany, for providing the medical image databases used in the experiments. This work is supported by the National Natural Science Foundation of China under Grant No. 61273251 and Qing Lan Project.

References

[1] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications-clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, pp. 1-23, February, 2004. Article (CrossRef Link)

[2] L. Yang, R. Jin, L. Mummert, R. Sukthankar, A. Goode, B. Zheng S. Hoi and M. Satya-narayanan, "A boosting framework for visuality-preserving distance metric learning and its application to medical image retrieval," *IEEE Pattern Analysis and Machine Intelligence*, vol. 32, no. 1, pp. 30-44, January, 2010. Article (CrossRef Link)

[3] U. Avni, H. Greenspan, E. Konen, M. Sharon and J. Goldberger, "X-ray categorization and retrieval on the organ and pathology level, using patch-based visual words," *IEEE Medical Imaging*, vol. 30, no. 3, pp. 733-746, March, 2011. Article (CrossRef Link)

[4] G. Quellec, M. Lamard, G. Cazuguel, "Case retrieval in medical databases by fusing heterogeneous information," *IEEE Medical Imaging*, vol. 30, no. 1, pp. 108-118, January, 2011. Article (CrossRef Link)

[5] M.M. Rahman, S.K. Antani, and G.R. Thoma, "A learning-based similarity fusion and filtering approach for biomedical Image retrieval using SVM classification and relevance feedback," *IEEE Information Technology in Biomedicine,* vol. 15, no. 4, pp. 640-646, July, 2011. Article (CrossRef Link)

[6] H.W. Chia, Y Li, C. Li. "Effective extraction of Gabor features for adaptive mammogram retrieval," in *Proc. of IEEE Conf. on Multimedia & Expo 2007*, pp. 1053-1056, July 2-5, 2007. Article (CrossRef Link)

[7] Y. Rui , T.S. Huang, M. Orgega and S. Mehrotra, "Relevance feedback: a power tool for interactive content-based image retrieval," *IEEE Circuits and Systems for Video Technology*, vol. 8, no. 5, pp. 644-655, September, 1998. Article (CrossRef Link)

[8] S.C.H Hoi, M.R. Lyu and R Jin, "A unified log-based relevance feedback scheme for image retrieval," *IEEE Knowledge and Data Engineering*, vol. 18, no. 4, pp. 509-524, April, 2006. Article (CrossRef Link)

[9] P. Yin, B. Bhanu, K. Chang and A. Dong, "Long-term cross-session relevance feedback using

virtual features," *IEEE Knowledge and Data Engineering*, vol. 20, no. 3, pp. 352-368, March, 2008. Article (CrossRef Link)

[10] Z. Li, J. Liu and X. Tang, "Pairwise constraint propagation by semidefinite programming for semi-supervised classification," in *Proc. of 25th Int. Conf. on Machine learning*, pp. 576-583, July 5-9, 2008. Article (CrossRef Link)

[11] Z. Lu and Y. Peng, "Exhaustive and Efficient Constraint Propagation: A Semi-Supervised Learning Perspective and Its Applications," *International Journal of Computer Vision*, vol. 103, no. 3, pp. 306-325, July, 2013. Article (CrossRef Link)

[12] Z. Fu, H. Lp, H. Lu and Z. Lu, "Multi-modal constraint propagation for heterogeneous image clustering," in *Proc. of 19th ACM Int. Conf. on Multimedia*, pp. 143-152, November 28 - December 1, 2011. Article (CrossRef Link)

[13] B. Thomee and M.S. Lew, "Interactive search in image retrieval: a survey," *International Journal of Multimedia Information Retrieval*, vol. 1, no. 2, pp. 71-86, July, 2012. Article (CrossRef Link)

[14] H. Muller, W. Muller, S. Marchand-Maillet, T. Pun and D.M. Squire, "Strategies for positive and negative relevance feedback in image retrieval," in *Proc. of 15th Int. Conf. on Pattern Recognition*, pp. 1043-1046, September 3-7, 2000. Article (CrossRef Link)

[15] G. Das, S. Ray and C. Wilson, "Feature re-weighting in content-based image retrieval," in *Proc. of 5th Int. Conf. on Image and Video Retrieval*, pp. 193-200, July 13-15, 2006. Article (CrossRef Link)

[16] J. Laaksonen, M. Koskela and E. Oja, "PicSOM: self-organizing image retrieval with MPEG-7 content descriptors," *IEEE Neural Networks*, vol. 13, no. 4, pp. 841-853, July, 2002. Article (CrossRef Link)

[17] S. Tong and E. Chang, "Support vector machine active learning for image retrieval," in *Proc. of 9th ACM Int. Conf. on Multimedia*, pp. 107-118, September 30 - October 5, 2001. Article (CrossRef Link)

[18] C.H. Hoi, C.H. Chan, K.Z. Huang, M.R. Lyu and I. King. "Biased support vector machine for relevance feedback in image retrieval," in *Proc. of IEEE Int. Conf. on Neural Networks*, pp. 3189-3194, July 25-29, 2004. Article (CrossRef Link)

[19] D. Tao, X. Tang, X. Li and X. Wu, "Asymmetric bagging and random subspace for support vector machines-based relevance feedback in image retrieval," *IEEE Pattern Analysis and Machine Intelligence*, vol. 28, no. 7, pp. 1088-1099, July, 2006. Article (CrossRef Link)

[20] J. Ye, R. Janardan and Q. Li, "Two-dimensional linear discriminant analysis," in *Proc. of the Advances in Neural Information Processing Systems*, pp. 1569-1576, December 13-18, 2004. Article (CrossRef Link)

[21] D. Xu, S. Yan, D. Tao, S. Lin and H.J. Zhang, "Marginal Fisher analysis and its variants for human gait recognition and content- based image retrieval," *IEEE Image Processing*, vol. 16, no. 11, pp. 2811-2821, November, 2007. Article (CrossRef Link)

[22] X. He, D. Cai and J. Han. "Learning a maximum margin subspace for image retrieval," *IEEE Knowledge and Data Engineering*, vol. 20, no. 2, pp. 189-201, February, 2008. Article (CrossRef

Link)

[23] E.D. Vesa, J. Domingob, G. Ayalac and P. Zuccarello, "A novel Bayesian framework for relevance feedback in image content-based retrieval systems," *Pattern Recognition*, vol. 39, no. 9, pp. 1622-1632, September, 2006. Article (CrossRef Link)

[24] G. Giacinto and F. Roli. "Instance-based relevance feedback in image retrieval using dissimilarity spaces," *Case-Based Reasoning on Images and Signals*, vol. 73, no. 1, pp. 419-436, January, 2008. Article (CrossRef Link)

[25] M.A. Herráez and F.J. Ferri, "An improved distance-based relevance feedback strategy for image retrieval," *Image and Vision Computing*, vol. 31, no. 10, pp. 704-713, October, 2013. Article (CrossRef Link)

[26] T. Amin, M. Zeytinoglu and L. Guan, "Application of Laplacian mixture model to image and video retrieval," *IEEE Multimedia*, vol. 9, no. 7, pp. 1416-1429, November, 2007. Article (CrossRef Link)

[27] M. Arevalillo-Herráeza, M. Zacarésb, X. Benaventc and E.D. Vesa, "A relevance feedback CBIR algorithm based on fuzzy sets," *Signal Processing: Image Communication*, vol. 23, no. 7, pp. 490-504, August, 2008. Article (CrossRef Link)

[28] H. Chang, D.Y. Yeung, "Kernel-based distance metric learning for content-based image retrieval," *Image and Vision Computing*, vol. 25, no. 5, pp. 695-703, May, 2007. Article (CrossRef Link)

[29] Z.H. Zhou, K.J Chen and H.B. Dai. "Enhancing relevance feedback in image retrieval using unlabeled data," *ACM Information Systems*, vol. 24, no. 2, pp. 219-244, April, 2006. Article (CrossRef Link)

[30] L. Zhang, L. Wang and W. Lin, "Semi-supervised biased maximum margin analysis for interactive image retrieval," *IEEE Image Processing*, vol. 21, no. 4, pp. 2294-2308, April, 2012. Article (CrossRef Link)

[31] J. He, M. Li, H. Zhang and H. Tong, "Generalized manifold-ranking-based image retrieval," *IEEE Image Processing*, vol. 15, no. 10, pp. 3170-3177, October, 2006. Article (CrossRef Link)

[32] S.R. Bulò, M. Rabbi and M. Pelillo, "Content-based image retrieval with relevance feedback using random walks," *Pattern Recognition*, vol. 44, no. 9, pp. 2109-2122, September, 2011. Article (CrossRef Link)

[33] D. Zhou, O. Bousquet, T.N. Lal, J. Weston and B. Schölkopf. "Learning with local and global consistency," in *Proc. of the Advances in Neural Information Processing Systems*, pp. 321-328, December 8-13, 2003. Article (CrossRef Link)

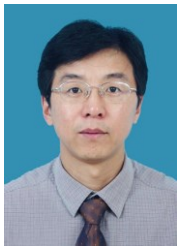
[34] X. Zhu, Z. Ghahramani and J. Lafferty. "Semi-supervised learning using gaussian fields and harmonic functions," in *Proc. of 20th Int. Conf. on Machine Learning*, pp. 912-919, August 21-24, 2003. Article (CrossRef Link)

[35] T. Tommasi, B. Caputo, P. Welter, M.O. Güld and T.M. Deserno, "Overview of the CLEF 2009 medical image annotation track," in *Proc. of 10th Int. Conf. on Cross-language evaluation forum*, pp. 85-93, September 30- October 2, 2009. Article (CrossRef Link)

[36] J. Oliveira, A. Machado, G. Chavez, A. Lopes, T.M. Deserno and A. Araújo, "MammoSys: a content-based image retrieval system using breast density patterns," *Computer Methods and Programs in Biomedicine*, vol. 99, no. 3, pp. 289-297, September, 2010. Article (CrossRef Link)

Menglin Wu received the B.S. and M.S. degree in the School of Computer Science and Technology, Nanjing University of Science and Technology, China in 2004 and 2006, respectively. Now he is a PhD candidate in the same affiliation. His research interests mainly include biomedicine image analysis, pattern recognition and machine learning.

Qiang Chen is an Assistant Professor in the School of Computer Science and Technology, Nanjing University of Science and Technology. His current interests include medical image analysis, image processing and pattern recognition.



QunSen Sun is a Professor in the School of Computer Science and Technology, Nanjing University of Science and Technology. He received his Ph.D. degree in pattern recognition and intelligence system from the same affiliation. His current interests include pattern recognition, image processing, computer vision and data fusion.