LETTER

QDFA: Query-Dependent Feature Aggregation for Medical Image Retrieval

Yonggang HUANG[†a)], ***Student Member***, **Dianfu MA**[†], **Jun ZHANG**[††], ***and*** **Yongwang ZHAO**[†], ***Nonmembers***

SUMMARY We propose a novel query-dependent feature aggregation (QDFA) method for medical image retrieval. The QDFA method can learn an optimal feature aggregation function for a multi-example query, which takes into account multiple features and multiple examples with different importance. The experiments demonstrate that the QDFA method outperforms three other feature aggregation methods.

key words: *CBIR, feature aggregation, query-dependent, fuzzy SVM*

1. Introduction

Large numbers of medical images are now available in online repositories, in which the images are generally stored and accessed in common formats (such as JPEG, GIF, etc) other than DICOM format for convenience and anonymization purposes [1]. Since there is no textual description attached, medical images cannot be effectively indexed and retrieved by traditional text-based retrieval techniques. This has lead to the use of the content-based image retrieval (CBIR) techniques which can search for medical images based on the modality, anatomic region and different acquisition views [1].

In CBIR, the images are retrieved according to their visual similarities on extracted low level features, such as color, texture and spatial location. Due to low retrieval accuracy of using a single feature, current CBIR systems usually take the approach of feature fusion to enhance the retrieval performance. There are two main ways to perform feature fusion for image retrieval [2]. One is called as early fusion, in which, the descriptor values of multiple visual features are stacked as a single, large vector, and the images are ranked by calculating the distances between the large vectors [3]. The early fusion usually suffers from the problem of dimensionality arising [2]. The other is called as late fusion, also known as feature aggregation [2], which obtains image similarities through combining multiple feature similarities. By comparison with early fusion, feature aggregation alleviates the dimensionality arising, and can adopt special designed distance functions for different visual features [2]. Because of these merits, several feature aggregation methods have been proposed. We focus on developing a new feature aggregation method in this letter.

Deselaers et al. [4] applied CombSumScore, CombMaxScore, CombSumRank, CombMaxRank functions to aggregate multiple similarities for multi-feature and multi-example queries. Deselaers' methods are query-independent, which apply the same feature aggregation function for different queries, without considering that a special feature is not equally important for different queries.

Kushki et al. [3] introduced a hierarchical decision fusion framework formulated with fuzzy logic. The feature aggregation functions for different image queries are presented using fuzzy logic based expressions, which require users to tune aggregation parameters.

A few query-dependent feature aggregation methods have also been reported for multi-example queries to date. Zhang et al. [2] proposed a local feature aggregation function based on support vector machine (LSVMC). In LSVMC, the query-dependent feature aggregation problem is formulated as a binary classification problem and solved by support vector machine (SVM). The authors [5] proposed a query-dependent feature fusion method based on one-class support vector machine (OSVM-QDFF). In OSVM-QDFF, the query-dependent feature aggregation problem is formulated as a one-class classification problem and solved by one-class support vector machine (OSVM). LSVMC and OSVM-QDFF take into account the different importance of a single visual feature for different queries. However, LSVMC and OSVM-QDFF treat multiple example images equally, without considering that different example images may play different roles to express the user's query.

In this letter, we propose a new query-dependent feature aggregation method for medical image retrieval. Different from existing solutions, for a multi-example query, the proposed feature aggregation method can learn an optimal feature aggregation function, which can take into account multiple features and multiple examples with different importance to express the user's query.

The remaining of the letter is organized as follows. In Sect. 2, a new query-dependent feature aggregation method is presented. The experiments and results are reported in Sect. 3. Finally, Sect. 4 concludes this letter.

2. Query-Dependent Feature Aggregation (QDFA)

2.1 Problem Definition

Let us consider a medical image collection $\mathbb{I}$ which contains

Manuscript received July 27, 2011.

[†]The authors are with the School of Computer Science and Engineering, BeiHang University, Beijing, China.

[††]The author is with the School of Information Technology, Deakin University, Victoria, Australia.

a) E-mail: yonggang.h@gmail.com

DOI: 10.1587/transinf.E95.D.275

n images, $\mathbb{I} = \{I_i\}_{i=1}^{n}$. Suppose m visual features are available, $\mathbb{F} = \left\{F_j\right\}_{j=1}^{m}$, and the special designed distance functions for these visual features are $\left\{D_j(\cdot,\cdot)\right\}_{j=1}^{m}$. Assuming the user provides some example images as a query, $\mathbb{Q} = \{Q_k\}_{k=1}^{q}$.

The feature representation of the image I_i can be denoted as a set of feature vectors $\left\{F_{ij}\right\}_{j=1}^{m}$. The distance between the example image Q_k and the image $I_i \in \mathbb{I}$ on the visual feature F_j can be calculated as $d_{ij}^k = D_j(F_{kj}, F_{ij})$. The distance can be normalized as:

$$\bar{d}_{ij}^k = \frac{d_{ij}^k - d_{min,j}^k}{d_{max,j}^k - d_{min,j}^k} \tag{1}$$

where $d_{max,j}^k$ and $d_{min,j}^k$ denote the maximum and minimum distances between Q_k and images in $\mathbb{I}$ on the visual feature F_j. The normalized distances can be converted to feature similarities as $s_{ij}^k = 1 - \bar{d}_{ij}^k$.

Given the features set $\mathbb{F}$, the feature similarities between Q_k and I_i can be represented as a feature similarity vector:

$$\mathbb{S}_i^k = (s_{i1}^k, \cdots, s_{ij}^k, \cdots, s_{im}^k) \tag{2}$$

By considering a linear feature similarities aggregation solution, the example similarity between Q_k and I_i can be obtained as:

$$s_i^k = \mathbb{S}_i^k \cdot \mathbf{w}^k \tag{3}$$

where $\mathbf{w}^k = (w_1^k, \cdots, w_j^k, \cdots, w_m^k)^{\mathrm{T}}$ is the feature weight vector, and w_j^k is the feature weight assigned for F_j, which reflects the importance of F_j for the example image Q_k.

The example similarities between all example images in $\mathbb{Q}$ and the image I_i can be denoted as an example similarity vector:

$$\mathbb{S}_i = (s_i^1, \cdots, s_i^k, \cdots, s_i^q) \tag{4}$$

By considering a linear example similarities aggregation solution, the final relevance of I_i to $\mathbb{Q}$ can be calculated as:

$$R_i = \mathbb{S}_i \cdot \mathbf{v} \tag{5}$$

where $\mathbf{v} = (v_1, \cdots, v_k, \cdots, v_q)^{\mathrm{T}}$ is the example weight vector, and v_k denotes the example weight assigned for Q_k, which reflects the importance of Q_k for the query $\mathbb{Q}$.

Consequently, the goals of the proposed query-dependent feature aggregation method are: (**I**). to find the optimal feature weight vectors $\mathbf{w}^k, k \in [1, q]$ for formula (3) for each example image Q_k; (**II**). to find the optimal example weight vector $\mathbf{v}$ for formula (5).

2.2 Importance Degree Estimation for Example Images

Considering different example images play different roles to express the user's query, we take a naive approach to estimate the importance degrees of example images in $\mathbb{Q}$.

For an example image Q_k in the query $\mathbb{Q}$, the lower of the sum of distances between Q_k and other images in $\mathbb{Q}$, the higher is its importance degree. Otherwise, the higher of the sum of distances, the lower is its importance degree. Based on this argument, we firstly calculate the sum of distances for Q_k with the features set $\mathbb{F}$ as:

$$\theta(Q_k) = \sum_{I_i \in \{\mathbb{Q} - Q_k\}} \sum_{F_j \in \mathbb{F}} \bar{d}_{ij}^k \tag{6}$$

Then the exponential function is used to normalize the sum of distances to importance degree of Q_k as:

$$\rho(Q_k) = e^{-a \cdot \theta(Q_k)} \tag{7}$$

where a is the slack factor.

2.3 Aggregation of Feature Similarities Based on Fuzzy Support Vector Machine

According to formula (2), given the visual features set $\mathbb{F}$, the feature similarities between Q_k and I_j can be represented as a feature similarity vector $\mathbb{S}_i^k$. Consequently, the feature similarities between Q_k and all the images in $\mathbb{I}$ can be represented as a feature similarity space $\mathbb{P}_k$ with the size of $n \times m$:

$$\begin{bmatrix} s_{11}^k & \cdots & s_{1j}^k & \cdots & s_{1m}^k \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{i1}^k & \cdots & s_{ij}^k & \cdots & s_{im}^k \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ s_{n1}^k & \cdots & s_{nj}^k & \cdots & s_{nm}^k \end{bmatrix} \tag{8}$$

We find the optimal feature weight vector $\mathbf{w}^k$ for formula (3) through solving a binary classification problem defined in the feature similarity space $\mathbb{P}_k$. It is finding an optimal $\mathbf{w}^k$ such that all relevant images are closer to Q_k and all irrelevant images stay away from it.

The example images in $\mathbb{Q}$ are regarded as positive examples. Some images randomly selected from the image collection $\mathbb{I}$, with the same number of example images in $\mathbb{Q}$, are regarded as negative examples. Since the example images in $\mathbb{Q}$ have different importance, in this letter, we solve the specific binary classification problem using the fuzzy support vector machine (FSVM) algorithm [6].

Considering a linear binary classification problem in the feature similarity space $\mathbb{P}_k$ with the training data set as:

$$\{(\mathbb{S}_t^k, L_t, \mu_t)\}_{t=1}^{2q} \tag{9}$$

where $\mathbb{S}_t^k$ is a m-dimensional feature similarity vector of positive or negative example image I_t, $L_t \in \{-1, 1\}$ is the class label of the image I_t, and μ_t is the fuzzy membership value of the image I_t with the class label L_t. For all positive examples, the fuzzy membership values are set according to formula (7) as $\mu_t = \rho(I_t), I_t \in \mathbb{Q}$. For all negative examples, the fuzzy membership values are set equally to 1. The goal of training FSVM is to find the optimal separating hyperplane that maximizing the margin of separation and minimizing the classification errors in $\mathbb{P}_k$, which can be represented as:

$$\begin{cases} min & \frac{1}{2}\left\|\mathbf{w}^k\right\|^2 + C\sum_{t=1}^{2q}\mu_t\xi_t \\ s.t & L_t(\mathbb{S}_t^k \cdot \mathbf{w}^k + b) \geq 1 - \xi_t \\ & \xi_t \geq 0,\ t = 1, 2, \cdots, 2q \end{cases} \tag{10}$$

where $\mathbf{w}^k$ is the adaptive feature weight vector of the hyperplane, and b is the bias.

By comparison with the regular SVM, the error term ξ_t is scaled by the membership value μ_t in the cost function of FSVM. The membership values reflect the relative fidelities of the training examples. The example images with larger importance degrees are more important for the FSVM training than the example images with smaller degrees.

The solution can be found through a dual problem with the undetermined Lagrange multiplier α_t. With a properly chosen non-linear mapping function $\varphi(\mathbb{S}^k)$, the feature similarity vector can be mapped into a high-dimensional feature space to get a potentially better representation. The kernel function:

$$K(\mathbb{S}_t^k, \mathbb{S}_r^k) = \varphi(\mathbb{S}_t^k) \cdot \varphi(\mathbb{S}_r^k) \tag{11}$$

is used to compute the inner product of two feature similarity vectors in the input feature similarity space. Then we get the kernel vision of the dual problem. For a given kernel function $K(\mathbb{S}_t^k, \mathbb{S}_r^k)$, the output hyperplane decision function of FSVM is:

$$f(\mathbb{S}^k) = \sum_{t=1}^{2q} \alpha_t L_t K(\mathbb{S}_t^k, \mathbb{S}^k) + b \tag{12}$$

According to formula (3), the similarity between Q_k and I_i can be obtained using the output decision function as:

$$s_i^k = \mathbb{S}_i^k \cdot \mathbf{w}^k = f(\mathbb{S}_i^k) - b \tag{13}$$

2.4 Aggregation of Example Similarities Using Weighted Bayes Sum Rule

Corresponding to formula (5), we combine the output similarities of multiple FSVMs to obtain the final relevance of I_i to query $\mathbb{Q}$.

The sigmoid function is firstly employed to covert the example similarities to the class-conditional probabilities as:

$$P_i^k = \frac{1}{1 + e^{-s_i^k}} \tag{14}$$

Then, we use the weighted bayes sum rule to obtain the final relevance of the image I_i to the query $\mathbb{Q}$ as:

$$R_i = \mathbb{S}_i \cdot \mathbf{v} = \sum_{k=1}^{p} P_i^k \cdot v_k = (P_i^1, \cdots, P_i^p) \cdot \mathbf{v} \tag{15}$$

where $\mathbf{v} = (v_1, \cdots, v_k, \cdots, v_q)^{\mathrm{T}}$ is the example weight vector. v_k is the example weight assigned for Q_k, which can be computed according to formula (7) as $v_k = \rho(Q_k)$.

Finally, the images in $\mathbb{I}$ are ranked according to the their final relevance to the query $\mathbb{Q}$.

3. Experiments and Results

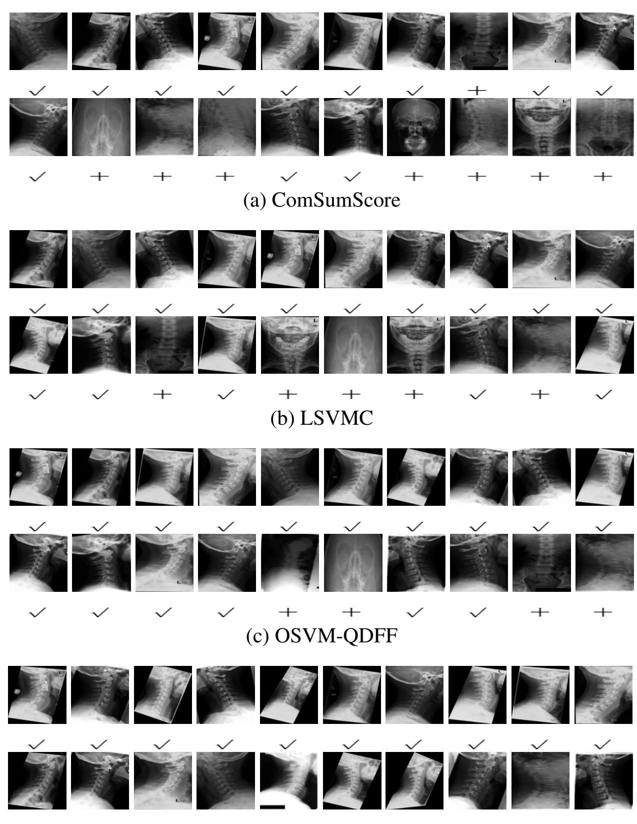
A number of experiments were carried out on the IRMA medical image collection [7] which contains 9000 medical images and are subdivided into 57 classes. The images are classified manually by reference coding with respect to a mono-hierarchical coding scheme. The scheme describes the imaging modality, the body orientation, the body region examined and the biological system examined. To evaluate the content based medical image retrieval, the query example images were randomly selected from each class and the remained images in the class were regarded as the corresponding ground truth set.

In the experiments, three different low-level features were used to represent the content of images, which are described as follows:

Spatial Layout Feature: The Color Layout Descriptor (CLD), with 64 luma component Y, was extracted to represent the spatial layout. The distance between two CLD vectors was calculated as $D_{cld}(Q, I) = \sqrt{\sum_i (Y_{Q_i} - Y_{I_i})^2}$.

(a) (b) (c) (d) (e)

Fig. 1 An example query with five example images.



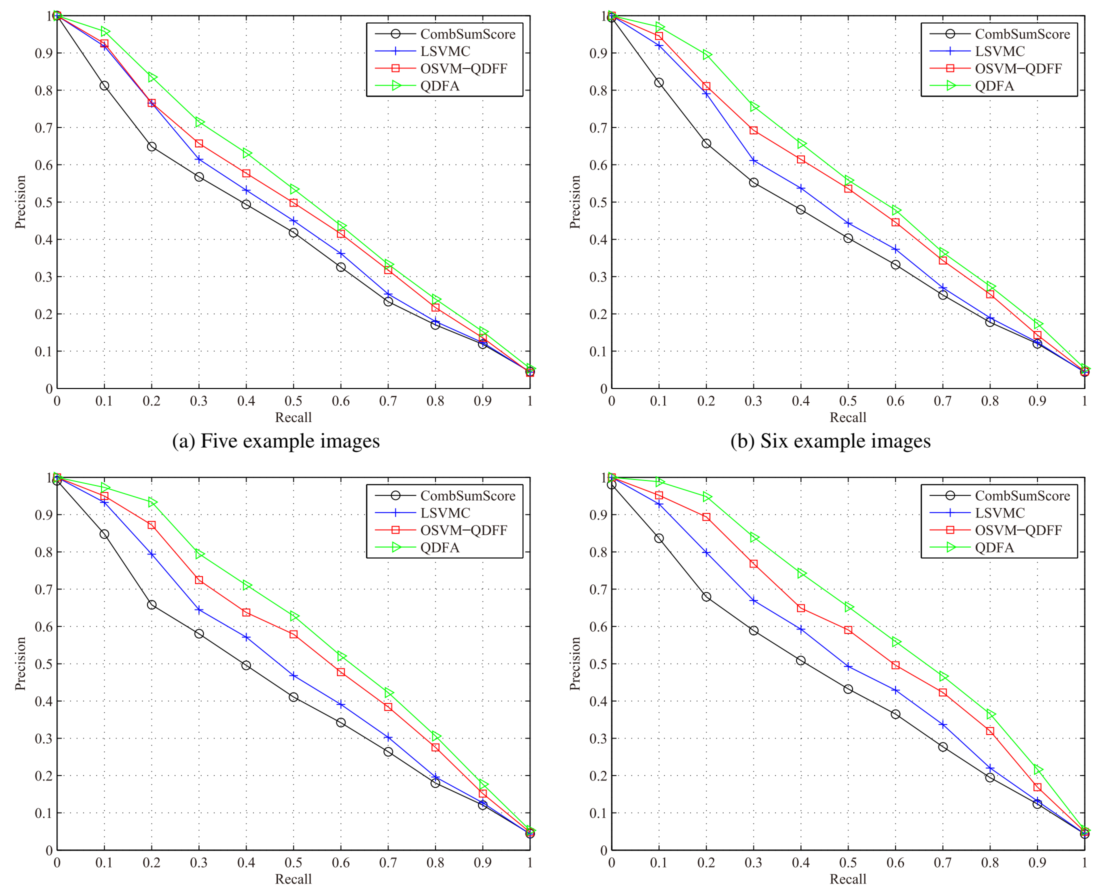
(a) ComSumScore

(b) LSVMC

(c) OSVM-QDFF

(d) QDFA

Fig. 2 Retrieval results for an example query.



(a) Five example images

(b) Six example images

(c) Seven example images

(d) Eight example images

Fig. 3 Retrieval performance.

Table 1 The importance degrees assigned in QDFA.

Example	a	b	c	d	e
Importance degree	0.7101	0.7019	0.7026	0.6772	0.6812

Texture Feature: The Tamura feature, with 1 coarseness, 1 contrast and 16 directionality from 16 directions, was extracted to represent the textual feature. The distance between two Tamura vectors was calculated as $D_{tamura}(Q, I) = \sqrt{\sum_i (T_{Q_i} - T_{I_i})^2}$.

Edge Feature: The Edge Histogram Descriptor (EHD) was used to represent the global edge feature. The distance between two EHD vectors was calculated as $D_{ehd} = \sum_i \left|H_{Q_i} - H_{I_i}\right|$.

To evaluate the retrieval performance of QDFA, CombSumScore [4], which has the best performance in Deselaers' feature aggregation functions, LSVMC [2] and OSVM-QDFF [5] were implemented as references. We used the LIBSVM [8] to solve SVMs (for LSVMC), OSVMs (for OSVM-QDFF) and FSVMs (for QDFA). Polynomial kernels were applied in our experiments, and we applied grid search for optimal parameter set that produces the best retrieval performance.

Figure 1 shows an example query with five example images. Figure 2 shows 20 top ranked images for the example query using the four different methods. The results demonstrate the better performance of QDFA than CombSumScore, LSVMC and OSVM-QDFF. Table 1 reports the importance degrees automatically estimated by QDFA. QDFA assigns the higher importance degrees to the example images that can express the user's query better.

Figure 3 reports the retrieval performance in terms of precision and recall. The average of precision and recall are calculated using 228 queries, which are formed by 4 randomly generated queries for each of the 57 classes. The number of example images in the queries varies from 5 to 8. The experimental results show that the retrieval performance of QDFA are always better than ComSumScore, LSVMC and OSVM-QDFF. In the case of five example images, the precision of QDFA is higher than that of CombSumScore, LSVMC and OSVM-QDFF about 15 percent, 8 percent and 5 percent respectively, when recall is less than 0.4. In the case of eight example images, the precision of QDFA is higher than that of CombSumScore, LSVMC and OSVM-QDFF about 25 percent, 15 percent and 8 percent respectively, when recall is less than 0.4.

The proposed method and other three competing methods adopt different retrieval strategies of using multiple features and multiple examples. For ComSumScore, different features and example images are treated equally for all queries. LSVMC and OSVM-QDFF can take into account the different importance of a single visual feature for different queries. In QDFA, different visual features and different example images are treated query-adaptive.

4. Conclusions

This letter proposed a new query-dependent feature aggregation method for medical image retrieval. For a multi-example query, the proposed feature aggregation method can learn an optimal feature aggregation function, which takes into account multiple features and multiple examples with different importance. A number of experiments were carried out on a real-world medical image dataset, and the results showed the proposed QDFA method outperforms CombSumScore, LSVMC and OSVM-QDFF.

Acknowledgements

The authors thank the anonymous reviewer for valuable feedback to this work. The authors also thank courtesy of TM Deserno, Dep. of Medical Informatics, RWTH Aachen, Germany, for providing IRMA dataset. This work was supported by Natural Science Foundation (61003017) and Core-High-Base Project (2010ZX01042-002-001) of China.

References

[1] M.M. Rahman, B.C. Desai, and P. Bhattacharya, "Medical image retrieval with probabilistic multi-class support vector machine classifiers and adaptive similarity fusion," Computerized Medical Imaging and Graphics, vol.32, no.2, pp.95–108, March 2008.

[2] J. Zhang and L. Ye, "Local aggregation function learning based on support vector machines," Signal Processing, vol.89, no.11, pp.2291–2295, Nov. 2009.

[3] A. Kushki, P. Androutsos, K.N. Plataniotis, and A.N. Venetsanopoulos, "Retrieval of images from artistic repositories using a decision fusion framework," IEEE Trans. Image Process., vol.13, no.3, p.277, March 2004.

[4] K. Donald and A. Smeaton, "A comparison of score, rank and probability-based fusion methods for video shot retrieval," The 4th International Conference on Image and Video Retrieval, pp.61–70, Singapore, July 2005.

[5] Y. Huang, D. Ma, J. Zhang, Y. Zhao, and S. Yi, "A new query dependent feature fusion approach for medical image retrieval based on one-class svm," J. Computational Information Systems, vol.7, no.3, pp.654–665, March 2011.

[6] C.F. Lin and S.D. Wang, "Fuzzy support vector machines," IEEE Trans. Neural Netw., vol.13, no.2, pp.464–471, March 2002.

[7] T. Lehmann, H. Schubert, B. Ott, and M. Leisten, "Image retrieval in medical applications," RWTH Aachen University, http://ganymed.imib.rwth-aachen.de/irma/datasets, accessed July 25, 2011.

[8] C. Chang and C. Lin, "LIBSVM: A library for support vector machines," National Taiwan University, http://www.csie.ntu.edu.tw/~cjlin/libsvm/, accessed July 25, 2011.