RELEVANCE FEEDBACK LEARNING WITH FEATURE SELECTION IN REGION-BASED IMAGE RETRIEVAL

Wei Jiang, Guihua Er, Qionghai Dai

Department of Automation
Tsinghua University
Beijing 100084, China

Lian Zhong, Yao Hou

Hi-tech Industry Department
Development & Reform Bureau
Shenzhen Municipal Government, China

ABSTRACT

Region-based image retrieval and relevance feedback are two important methods to bridge the gap between the low-level visual features and the high-level semantic concepts in content-based image retrieval. In this paper, we address the issue of introducing the relevance feedback mechanism into the region-based image retrieval with online feature selection during each feedback round. Our contribution has two folds. (1) A novel region-based image representation is proposed. Based on a generative model, a fuzzy codebook is extracted from the original region-based features, which represents the images in a uniform real-value feature space. (2) A feature selection criterion is developed and an effective relevance feedback algorithm is implemented in a boosting manner to simultaneously select the optimal features from the fuzzy codebook, and generate a strong ensemble classifier over the selected features. Experimental results show that the proposed scheme can substantially improve the retrieval performance of the region-based image retrieval system.

1. INTRODUCTION

With the effort for about one decade, great improvements have been achieved in *Content-based image retrieval* (*CBIR*). At the same time, researchers have realized that the gap between the low-level visual features and the high-level semantic concepts remains an unsolved difficulty which limits its further development of CBIR systems [9]. *Region-based image retrieval* (*RBIR*) [1, 2, 5] and *Relevance Feeback* (*RF*) [7] are proposed to bridge the gap. The RBIR approaches segment images into several regions and retrieves images with low-level features based on the regions. Since region-based features represent images in the object level, and better fit the human perception than the global features extracted from the entire image, the RBIR methods usually have better performance. The RF mechanism is generally treated as an online supervised learning process, where the user labels some images to be "relevant" or "irrelevant" to

the query concept, and helps the system to successively improve retrieval results.

Although it can be expected that introducing the RF learning into the RBIR will improve the retrieval performance, by far few work has been done to address this issue. This is because in RBIR, different images have different number of regions, and it is hard to represent the images in a uniform feature space. Thus most of the RBIR approaches directly calculate the region-to-region similarity [1] or image-to-image similarity [2] to compare and retrieve images. The latest work in [5] gives a good example to combine the RF mechanism with the RBIR scheme. In [5] vector quantization is used to represent the images with binary index features, and the SVM classifier is adopted as the RF learner where Gaussian kernel is calculated by the region-based Earth Mover's Distance instead of the global-based Euclidean Distance.

To really grasp the query concept from the low-level features of the labeled training samples, it is important for an RF learner to tell which features are more suitable for explaining the current query concept. This refers to the problem of *online feature selection*. In CBIR systems based on global features, many methods are used for feature selection, such as the boosting online learning approach [11]. Again in the RBIR context, since it is difficult to represent images with a uniform feature space, few work has been done to tackle this problem as far as we know.

In this paper, addressing the issue of incorporating RF learning with RBIR, an effective RF algorithm with online feature selection is proposed as follows. (1) Based on the assumption that images in the database are generatively formed by the mixture model, a fuzzy codebook is extracted from the region-based features by grouping method to index the images. Compared with the region-based binary indexing in [5], our fuzzy codebook represents the images with real value fuzzy features, thus more information can be provided to retrieve images. (2) With the fuzzy codebook representation, an effective feature selection criterion is proposed based on calculating the similarity between the positive and

negative training sets. Then a boosting RF algorithm is implemented to select the optimal features one by one from the fuzzy codebook and construct an ensemble classifier over the selected features. The *Fuzzy Feature Contrast Model* (*FFCM*) [8] is adopted to calculate the similarity between images, which is introduced into the *Unified Feature Matching* (*UFM*) [2] measurement to compute the similarity between the "relevant" and "irrelevant" image sets. Experimental results show that our proposal can substantially improve the retrieval performance of RBIR system.

2. REGION-BASED IMAGE REPRESENTATION

In this section, we describe the method to extract region features, followed by the detail of learning fuzzy codebook and representing images in a uniform fuzzy feature space.

2.1. Feature extraction and image segmentation

To be segmented, an image is firstly partitioned into small blocks without overlap. Then the feature vectors are extracted based on each block. In our system each block has 16×16 pixels, which is a tradeoff between the computational complexity and the effectiveness of the extracted feature. The color features used are the 9-dimensional color moments in Luv color space (first 3 moments), and the 64-dimensional color histogram in Hsv space; the texture features used are the 10 dimensional coarseness vector, and the 8-dimensional directionality. With the 91-dimensional feature representation, the JSEG algorithm [3] is adopted to segment the image, which can adaptively determine the number of regions. The segmentation performance is fairly good in most situations. After segmentation, an image $\mathbf{x}_i$ is divided into a set of regions $\mathcal{R}(\mathbf{x}_i) = \{r_1(\mathbf{x}_i), \ldots, r_{m_i}(\mathbf{x}_i)\}$, and each region is represented by the mean feature vector of the member blocks in it.

2.2. Image representation by fuzzy codebook

Let $\mathcal{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}$ denote the whole database with N images. Let $\Re = \bigcup_{i=1}^{N} \mathcal{R}(\mathbf{x}_i)$ be the combination of all the segmented regions from all the images in $\mathcal{X}$. Since the low-level feature space here has a large dimensionality and has great redundancy in representing images, the Principle Component Analysis (*PCA*) is exploited to simplify the feature representation. After PCA, the original $\Re$ are projected into a new feature space $\tilde{\mathcal{F}}$ with a much lower dimensionality d, and we denote it by $\tilde{\Re}$. Rename the element in $\tilde{\Re}$ as $\tilde{\Re} = \{r_1, \ldots, r_m\}$, and assume that the region r_j is generatively formed by n codewords $C_1, \ldots, C_n$ in forms of the Gaussian Mixture Model as

$$p(r_j) = \sum_{k=1}^{n} p(C_k)p(r_j|C_k) \qquad (1)$$

where $p(r_j|C_k)$ follows the Gaussian distribution. Then the mixture-based clustering method with model selection
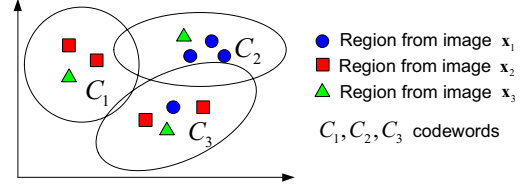

Fig. 1. Generating fuzzy codebook from region-based features.

proposed in [6] is adopted to adaptively determine the following parameters with the Expectation Maximization algorithm and minimum message length model selection criterion: the number of the codewords n, the codeword priors $p(C_k)$, the mean μ_k and co-variance matrix Σ_k of the codewords ($k = 1, \ldots, n$). Fig.1 illustrates the process to generate the fuzzy codebook.

Given an image $\mathbf{x}_i \in \mathcal{X}$, it can be represented by the codewords in $\mathcal{C} = \bigcup_i^n C_i$ as a set of fuzzy features $\mathbf{F}(\mathbf{x}_i) = [f(\mathbf{x}_i, C_1), \ldots, f(\mathbf{x}_i, C_n)]^T$, where fuzzy feature $f(\mathbf{x}_i, C_k)$ denotes the saliency membership of the codeword C_k in representing image $\mathbf{x}_i$, which is given by:

$$f(\mathbf{x}_i, C_k) = \sum_{r \in \mathbf{x}_i} v(r)G(r, C_k) \qquad (2)$$

where r is a region in $\mathbf{x}_i$, $v(r)$ is its region saliency (which describes how well the region represents image $\mathbf{x}_i$), and

$$G(r, C_k) = p(C_k) \cdot p(r|C_k) \qquad (3)$$

The region saliency $v(r)$ is proportional to the size of r, and is inversely proportional to the average distance of each pixel in r to the image center.

3. RELEVANCE FEEDBACK LEARNING WITH BOOSTING FEATURE SELECTION

Assume that during the retrieval process, in the t-th round of relevance feedback, images in $\mathcal{R}^t$ and $\mathcal{IR}^t$ are labeled to be "relevant" and "irrelevant" respectively by the user. Representing images with the fuzzy codebook, boosting method can be adopted to select features one by one from the fuzzy codebook. The original AdaBoost algorithm [4] trains weak classifiers along each feature axis and selects features based on the training error. It usually needs a lot of training samples and candidate features to give good performance. Since in the RBIR problem, the number of the training samples and the fuzzy features are both small, AdaBoost is often overfed. In this paper we propose an effective feature selection criterion. Since the goal of feature selection is to find the optimal feature subspace where $\mathcal{R}^t$ and $\mathcal{IR}^t$ can be best separated, the similarity between $\mathcal{R}^t$ and $\mathcal{IR}^t$ along each feature axis is used as the feature selection criterion in the boosting algorithm. The FFCM measurement is used to calculate the similarity between images, and the UFM method is adopted to calculate the similarity between $\mathcal{R}^t$ and $\mathcal{IR}^t$. The detail is given in the following subsections.

3.1. Criterion for feature selection

– Similarity between two images

The FCM measurement is firstly proposed as a psychological similarity measurement between two objects [10], and in [8] it is generalized to the FFCM measurement and is applied to the computer vision field to represent the similarity between human faces. Let a, b denote two stimuli, which are represented by two fuzzy feature vectors $\tilde{A} = [A_1, \ldots, A_d]$, $\tilde{B} = [B_1, \ldots, B_d]$ respectively. FFCM defines the similarity between a, b as:

$$S(\tilde{A}, \tilde{B}) = \sum\nolimits_{i=1}^{d} (\min\{A_i, B_i\} - \alpha \max\{A_i - B_i, 0\} \\ - \beta \max\{B_i - A_i, 0\}) \qquad (4)$$

If $\alpha \neq \beta$, we emphasize the features in different stimuli unequally. The asymmetry direction is determined by the relative "salience" of the stimuli: if a is more salient, $\alpha \geq \beta$. Represented by the fuzzy codebook as Eqn.(2), the similarity between two images can be calculated with Eqn.(4).

– Similarity between image sets

Given two fuzzy set $\mathcal{A} = \{\tilde{A}_1, \ldots, \tilde{A}_p\}$, $\mathcal{B} = \{\tilde{B}_1, \ldots, \tilde{B}_q\}$, the UFM method [2] defines the similarity between them as:

$$S(\mathcal{A}, \mathcal{B}) = \vec{W}(\mathcal{A}, \mathcal{B})^T \vec{L}(\mathcal{A}, \mathcal{B}) \qquad (5)$$

where $\vec{W}(\mathcal{A}, \mathcal{B}) = [w(\tilde{A}_1), \ldots, w(\tilde{A}_p), w(\tilde{B}_1), \ldots, w(\tilde{B}_q)]^T$, $w(\tilde{A}_i)$ and $w(\tilde{B}_j)$ are the saliency membership of the member $\tilde{A}_i$ and $\tilde{B}_j$ in feature set $\mathcal{A}$ and $\mathcal{B}$ respectively. $\vec{L}(\mathcal{A}, \mathcal{B})$ is given by the following formula:

$$\vec{L}(\mathcal{A}, \mathcal{B}) = \\ [l(\tilde{A}_1, \mathcal{B}), \ldots, l(\tilde{A}_p, \mathcal{B}), l(\tilde{B}_1, \mathcal{A}), \ldots, l(\tilde{B}_q, \mathcal{A})]^T \quad (6)$$

$$l(\tilde{A}_i, \mathcal{B}) = \max_{j=1,\ldots,q} S(\tilde{A}_i, \tilde{B}_j)$$

$$l(\tilde{B}_j, \mathcal{A}) = \max_{i=1,\ldots,p} S(\tilde{B}_j, \tilde{A}_i)$$

By Eqn.(5) the similarity between $\mathcal{R}^t = \{\mathbf{x}_1^+, \ldots, \mathbf{x}_p^+\}$ and $\mathcal{IR}^t = \{\mathbf{x}_1^-, \ldots, \mathbf{x}_q^-\}$ along the feature axis C_k can be calculate, where $S(\mathbf{x}_i^+, \mathbf{x}_j^-)$ is given by the FFCM measurement by Eqn.(4). Since in image retrieval, the "relevant" images often share some common features, while the "irrelevant" images are different from the "relevant" ones in different ways. The common features in $\mathcal{R}^t$ is usually more important than those in $\mathcal{IR}^t$. We set $\alpha > \beta$ to meet this asymmetry requirement. In practical $\alpha = 0.5$, $\beta = 0.3$.

3.2. Boosting feature selection

Given a set of optimal feature axis selected by the UFM similarity, the Real AdaBoost method [4] is adopted to select a new added feature axis by re-weighting the training samples, and combine the classifiers constructed over the incrementally learned features into an ensemble classifier with decreased training error.

Assume that during round t, k features $C_1^*, \ldots, C_k^* \in \mathcal{C}$ are already selected during the previous k iterations in the boosting method, and k optimal weak classifiers $H_{g(i)}$, $i = 1, \ldots, k$ are constructed, one for each selected feature in each iteration. For a training sample $\mathbf{x}$, the current weight is $w^t(\mathbf{x})$. Assume that $\hat{y}$ is the class label of $\mathbf{x}$ predicted by the system, and y is true label. The class probability estimated by $H_{g(i)}$ is $q_{g(i)}(\mathbf{x}) = p(\hat{y} = 1 | H_{g(i)})$. To select a new feature which is optimal in discriminating sets $\mathcal{R}^t$ and $\mathcal{IR}^t$, we project the training samples to each feature axis C_i, and calculate the similarity $S_i(\mathcal{R}^t, \mathcal{IR}^t)$ with the current sample weights along C_i by Eqn.(5). Then $C_{g(k+1)}$ is selected as the optimal feature, along which the two image sets have the smallest similarity. $C_{k+1}^* = C_{g(k+1)}$. A weak classifier $H_{g(k+1)}$ is constructed along the optimal C_{k+1}^* and a set of class probabilities $q_i(\mathbf{x})$ are given. Then set

$$h_{g(i)}(\mathbf{x}) = 0.5 \log q_{g(i)}(\mathbf{x}) / (1 - q_{g(i)}(\mathbf{x})) \qquad (7)$$

and the class probability are combined as:

$$h_{en}^{k+1}(\mathbf{x}) = \sum\nolimits_{j=1}^{k+1} h_{g(j)}(\mathbf{x}) \qquad (8)$$

The weight for sample $\mathbf{x}$ is then updated as:

$$w^{k+1}(\mathbf{x}) = \frac{1}{Z} w^k(\mathbf{x}) exp\{-y h_{g(k+1)}(\mathbf{x})\} \qquad (9)$$

where Z is a normalization factor to make $\sum_{\mathbf{x}} w^{k+1}(\mathbf{x}) = 1$.

The *Fuzzy K-Nearest Neighbor* (*FKNN*) classifier is used as the weak learner, whose class probability is given by

$$q(\mathbf{x}) = \frac{\sum_{j=1}^{k} w'(\mathbf{x}_j^*) \left(1/||\mathbf{x} - \mathbf{x}_j^*||^2\right)}{\sum_{j=1}^{k} \left(1/||\mathbf{x} - \mathbf{x}_j^*||^2\right)} \qquad (10)$$

where $\mathbf{x}_j^*, j = 1, \ldots, k$ are k nearest neighbors of $\mathbf{x}$ in the training set, $w'(\mathbf{x}_j^*) = w(\mathbf{x}_j^*)$ or $-w(\mathbf{x}_j^*)$ if $\mathbf{x}_j^* \in \mathcal{R}$ or $\mathcal{IR}$.

The entire framework of the RBIR system is summarized in Fig.2. In practical, we set $K = 100$.

- The User selects a query image $\mathbf{x}_q$ to start query, and in the first round of retrieval the rank of an image $\mathbf{x}$ is determined by the UFM similarity $S(\mathbf{x}, \mathbf{x}_q)$
- For subsequent round $t > 1$
1. Initialize $w^0(\mathbf{x}) = \frac{1}{2|\mathcal{R}^t|}$ or $\frac{1}{2|\mathcal{IR}^t|}$ for $\mathbf{x} \in \mathcal{R}$ or $\mathcal{IR}$
2. Begin iteration: for k=1,...,K
    - Get $S_i(\mathcal{R}^t, \mathcal{IR}^t)$ by Eqn.(5) along each C_i;
    - Select $C_k^* = C_{g(k)}$ with smallest $S_{g(k)}(\mathcal{R}^t, \mathcal{IR}^t)$;
    - Construct an FKNN classifier along C_k^*;
    - Update weight by Eqn.(9);
    - Get ensemble class probability $h_{en}^k(\mathbf{x})$ by Eqn.(8);
3. Rank images by $h_{en}^k(\mathbf{x})$ in descending order;

Fig. 2. The framework of the RBIR system.

II - 511

4. EXPERIMENTAL RESULTS

The image database has 10,000 real world images from Corel CDs, which come from 100 semantic categories, 100 images for each category. In all the following experiments, the performance measurement used is the top-k precision P_k (the percentage of the "relevant" images in the returned k images), and the average P_k is calculated by the retrieval results of 1000 retrieval sessions. The 1000 query images are randomly selected from the database. During each retrieval round, 10 images are labeled by user to feedback.

Considering both the computational complexity and the effectiveness to represent the images, the size of the fuzzy codebook is selected through experiment. The average P_{30}, P_{50} and P_{100} for the 5-th feedback round are calculated with different codebook sizes in Fig.3 (a), and the corresponding average retrieval time cost is given in Fig.3 (b). The figures show that for all P_{20}, P_{50} and P_{100}, the larger the codebook size, the better the retrieval performance, and the longer the retrieval time. When the number of codewords exceeds 1000, the improvement of the precision becomes not very significant, but the time cost increases greatly. Thus in practical, the system use 1000 codewords as a tradeoff.
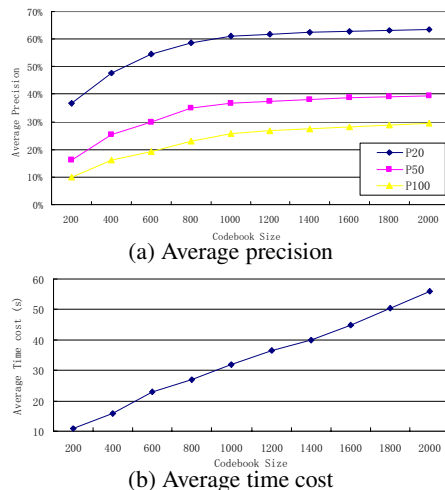

(a) Average precision

(b) Average time cost

Fig. 3. The precision and time cost for different codebook sizes.

To evaluate the effectiveness of our proposal in improving the retrieval performance, we compare our algorithm with the UFM method [2]. For fair comparison, all the methods use the same query images. Fig.4 gives the average P_{20} of our method and UFM. The figure shows that our method consistently outperforms the UFM. For example the performance improvement in round 5 is 31%. The experimental result indicates that the proposed image representation method and the relevance feedback learning method with boosting can improve the retrieval accuracy significantly.

5. CONCLUSIONS

In this paper we have proposed an effective relevance feedback method with online feature selection in the RBIR con-
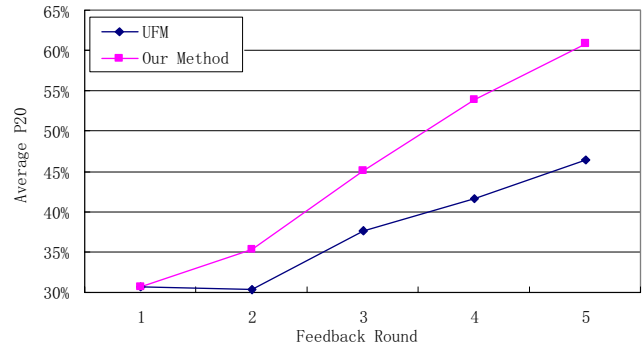

Fig. 4. Comparison between our proposal and the UFM method.

text. Based on the generative model, a fuzzy codebook is extracted to represent the images in a uniform fuzzy feature space. And based on the feature selection criterion to measure the similarity between the positive and negative training sets, an RF algorithm is implemented in a boosting manner to select the optimal features and generate an ensemble classifier during each feedback round. Experimental results show the effectiveness of our proposal.

7. REFERENCES

[1] C. Carson, et al., "Blobworld: A system for region-based image indexing and retrieval," *Proc. of Int. Conf. Visual Information System*, pp.509-516, 1999.

[2] Y.X. Chen, J.Z. Wang, "A region-based fuzzy feature matching approach for content-based image retrieval," *IEEE Trans. on PAMI*, 24(9): 1252-1267, 2002

[3] Y. Deng, et al., "Color image segmentation," *Proc. of IEEE Int. Conf. CVPR*, vol.2, pp.446-451, 1999.

[4] J. Friedman, et al., "Additive logistic regression: a statistical view of boosting ," *The Annals of Statistic*, 38(2): 337-374, 2000.

[5] F. Jing, et al., "An effective and efficient region-based image retrieval framework," *IEEE Trans. on Image Processing*, 13(5): 699-709, 2004.

[6] M.H.C. Law, et al., "Simultaneous feature selection and clustering using mixture models," *IEEE Trans. on PAMI*, 26(9):1154-1166, 2004.

[7] Y. Rui. et al, "Relevance feedback: A powerful tool in interactive content-based image retrieval," *IEEE Trans. on CSVT*, 8(5): 644-655, 1998.

[8] S. Santini, R. Jain, "Similarity measures," *IEEE Trans. on PAMI*, 21(9), pp. 871-883, 1999.

[9] A.W.M. Smeulders, et al., "Content-based image retrieval at the end of the ealy years", *IEEE Trans. PAMI*, 22(12): 1349-1380, 2002.

[10] A. Tvesky, "Feature of similarity," *Psychological review*, 84(4), pp. 327-352, 1977.

[11] K. Tieu, P. Viola, "Boosting image retrieval," *Proc. of IEEE Int. Conf. CVPR*, pp.228–235, 2000.