A FUZZY STATISTICAL CORRELATION-BASED APPROACH TO CONTENT-BASED IMAGE RETRIEVAL

Xiaojun Qi and Ran Chang

Xiaojun.Qi@usu.edu and ranchang@cc.usu.edu
Computer Science Department, Utah State University, Logan, UT 84322-4205

ABSTRACT

This paper presents an effective fuzzy long-term semantic learning method for relevance feedback-based image retrieval. The proposed system uses a statistical correlation-based method to dynamically learn the semantic relations between any relevance feedback image pairs. The learned semantic relations are used to automatically expand the feedback set to balance the number of positive and negative images to improve the fuzzy SVM-based low-level learning. They are also used to more accurately estimate the semantic similarity between the query image and database images. The overall similarity score between query and database images is computed by combining both low-level visual and high-level semantic similarity measures. Our extensive experimental results show the proposed system achieves the best retrieval accuracy when compared with three peer systems.

Index Terms— Content-based image retrieval, long-term-based semantic learning, fuzzy statistical correlation.

1. INTRODUCTION

Content-based image retrieval (CBIR) techniques have become one of viable solutions to quickly find desired images among a large imagery database by supplying sample query image(s). These techniques can be generally classified into three categories [1]: global feature-based, region/object level feature-based, and relevance feedback-based. Among these, relevance feedback-based techniques have been widely used to bridge the semantic gap between low-level visual features and high-level semantic meanings of images. They allow the user to label the returned images as relevant or irrelevant to the query image at each feedback iteration. The labeled images are further used by short-term and/or long-term learning techniques to refine the retrieval results at the next iteration. We will briefly review several techniques related to our proposed approach.

Short-term learning techniques aim to find out top images relevant to the user's query over the course of a single query. They normally apply machine learning techniques to find a decision boundary for classification based on the labeled images. Several typical examples include decision tree learning [2], Bayesian learning [3], support vector machine (SVM) learning [4], fuzzy SVM learning [5], boosting [6], etc. However, they cannot achieve good and reliable classification learning by using a small number of imbalanced relevant and irrelevant images. Furthermore, the semantic knowledge obtained in the feedback process of all query sessions is not remembered.

Long-term learning techniques aim to find out the semantic relations among images over the course of multiple queries. They normally use historical retrieval experiences over many search sessions to estimate the semantic relations among images. For example, Li *et al.* [7] use the triangular matrix to store semantic correlation collected from the statistics of users' feedback information. He *et al.* [8] use the semantic space to store relevancy or irrelevancy information related to query images. Hoi *et al.* [9] use the log base to store the accumulated semantic correlation among images. Han *et al.* [10] apply a memory learning framework to learn hidden semantic relations. However, the sparsity of memorized feedback information may make long-term learning not useful for a large-scale database.

To address the limitations of the current CBIR systems, we propose a fuzzy statistical correlation-based approach to bridge the semantic gap in CBIR. This technique first gathers users' feedback and stores the semantic correlation among images classified by users. It then uses the learned semantic relations to expand the classified positive or negative image set to improve fuzzy SVM-based low-level learning. It also estimates the semantic similarity between query image and each database image using the learned semantic relations and the classified training image set. The fuzzy SVM-based low-level short-term learning and the correlation-based high-level long-term learning are finally combined to improve the retrieval accuracy. The remaining of the paper is organized as follows: Section 2 presents our proposed fuzzy statistical correlation-based CBIR approach. Section 3 demonstrates the effectiveness of our proposed system and shows the experimental results to compare our proposed system with three peer systems. Section 4 draws conclusion and presents the future research direction.

2. PROPOSED SYSTEM

The retrieval process of our proposed system is as follows: The user first supplies a query image q. The system then returns top n images, which are classified by the user as either relevant or irrelevant to q. This process continues for a few feedback iterations, or until the user is satisfied with the retrieval results. For each iteration step, the system returns top images ranked by the similarity score $S(q, D_i)$ between query q and an arbitrary image D_i in the database:

$$S(q, D_i) = w_{short} \cdot S_{short}(q, D_i) + w_{long} \cdot S_{long}(q, D_i) \quad (1)$$

where $S_{short}(q, D_i)$ and $S_{long}(q, D_i)$ respectively measure the short-term-based low-level and long-term-based high-level similarity scores between q and D_i; w_{short} and w_{long} respectively are the contributing weights assigned to the short-term and long-term-based similarity measures, and are set to 0.5 in our system. Here, a higher similarity score means a smaller distance in terms of both low and high-level features and therefore corresponds to higher similarity.

A few definitions are listed here for the convenience of the following discussions. A query session is the overall iterative process to retrieve desired images. The current positive feedback set ***CurPos*** contains the images marked as relevant to q in the current feedback iteration. The current negative feedback set ***CurNeg*** contains the images marked as irrelevant to q in the current feedback iteration. The positive feedback set ***AllPos*** contains the accumulative distinct positive images (i.e., the images marked as relevant to q) in all feedback iterations of the current query session. The negative feedback set ***AllNeg*** contains the accumulative distinct negative images (i.e., the images marked as irrelevant to q) in all feedback iterations of the current query session. The number of images in a set x is denoted by $|x|$.

2.1. Correlation-Based Long-Term Semantic Relation

Our system automatically transforms users' relevance feedback into the semantic similarity among images. The following three observations guide this transformation. 1) If two images returned in a feedback iteration are both marked as positive, they belong to the same semantic category as the query image. 2) If one returned image is marked as positive while the other is marked as negative, they are not semantically related. 3) If two returned images are marked as negative, they could be semantically related, just not relevant to the query image, or they could be in different semantic categories. Therefore, we define the semantic similarity of two images i and j as:

$$S_H(i, j) = P(i, j) / CP(i, j) \quad (2)$$

where $P(i, j)$ is the number of feedback iterations where both images i and j are marked as positive and $CP(i, j)$ is the number of feedback iterations where both images i and j are returned and at least one is marked as positive. This ratio measures the similarity correlation strength between images i and j and ranges from 0 to 1. Similarly, we define the semantic dissimilarity of two images as:

$$O_H(i, j) = N(i, j) / CN(i, j) \quad (3)$$

where $N(i, j)$ is the number of feedback iterations where images i and j are marked with opposite labels and $CN(i, j)$ is the number of query sessions where images i and j are returned and at least one is marked as negative. This ratio measures the dissimilarity correlation strength between images i and j and ranges from 0 to 1.

These two semantic correlations are updated after the user provides the feedback information on returned images for each iteration step. That is, they store accumulative retrieval experiences gathered from all search sessions to date. Fig. 1 summarizes the algorithmic view of this updating process. Due to the symmetric property, i.e., $S_H(i, j) = S_H(j, i)$ and $O_H(i, j) = O_H(j, i)$, the update operation is performed on the upper triangular pairs (i, j)'s where $i < j$.

1. Initialize $S_H(i, j)$, $O_H(i, j)$, $P(i, j)$, $CP(i, j)$, $N(i, j)$, and $CN(i, j)$ as 0's.
2. For each feedback session, split the labeled images into two sets: ***CurPos*** and ***CurNeg***.
3. For every image pair (i, j) in ***CurPos***, $P(i, j) = P(i, j)+1$ and $CP(i, j) = CP(i, j)+1$.
4. For every image pair (i, j) where $i \in$ ***CurPos*** and $j \in$ ***CurNeg***, $CP(i, j) = CP(i, j)+1$, $N(i, j) = N(i, j)+1$, and $CN(i, j) = CN(i, j)+1$
5. For every image pair (i, j) in ***CurNeg***, $CN(i, j) = CN(i, j)+1$.
6. Update $S_H(i, j)$ and $O_H(i, j)$ using Equations (2) and (3) for all feedback image pairs (i, j)'s, respectively.
7. Repeat steps 2 through 6 when a new relevance feedback session starts.

Fig. 1: Algorithmic view of updating the correlation-based long-term semantic relation.

2.2. Fuzzy SVM-Based Short-Term Learning

We use a compact feature vector to represent each database image for our retrieval task. This feature vector consists of 9 color, 18 edge, and 9 texture components. Color features are computed by mean, variance, and skewness in HSV color space. Edge features are computed by 18-bin edge direction histogram obtained by applying the Sobel edge detector on the converted grayscale image. Texture features are computed by the entropy of each of 9 detail subbands in the wavelet domain of the converted grayscale image.

For initial retrieval, our system returns top n images based on $S_{short}(q, D_i)$, which is computed by the sum of the inverted Euclidean distance between color and texture features and the histogram intersection between edge features. In the following iterations, the user feedback information is first used to update the correlation-based long-term semantic relation as described in section 2.1. The updated semantic relation is then used to find additional images to ensure balanced positive and negative images in the training set. The feature vectors of this expanded, balanced training image set are further input to the radial basis function (RBF) kernel-based fuzzy SVM classifier to more accurately locate the decision boundary. The updated semantic relation is also used to estimate $S_{long}(q, D_i)$. The detailed steps of fuzzy SVM-based short-term learning are summarized in Fig. 2.

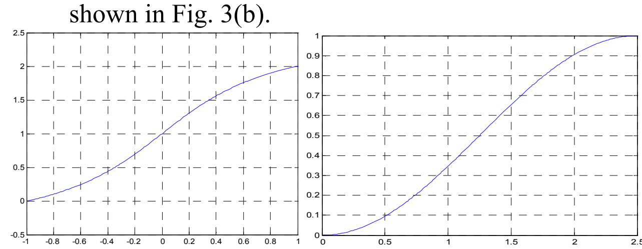
1. The user labels each returned image as either relevant or irrelevant to the query image q.
2. Update $S_H(i, j)$ and $O_H(i, j)$ using the semantic relation updating algorithm as summarized in Fig. 1.
3. For each database image D_i, estimate its overall similarity to query q by:
$$OS(q, D_i) = S_H(q, D_i) - O_H(q, D_i) \tag{4}$$
4. Sort the OS values in the descending order.
5. If **|*AllPos*|>|*AllNeg*|,** sequentially choose new images from the tail of the sorted OS list to ensure that **|*AllPos*| = |*AllNeg*|**.
6. If **|*AllPos*|<|*AllNeg*|,** sequentially choose new images from the head of the sorted OS list to ensure that **|*AllPos*| = |*AllNeg*|**.
7. The feature vectors of the expanded training images are input to the fuzzy SVMs to find the decision boundary. The memberships of the user's positively and negatively user labeled images are 1's and -1's, respectively. The membership of each expanded image is its OS value computed by Equation (4).
8. The normalized distance from D_i to the trained separating hyperplane is computed as $S_{short}(q, D_i)$.
9. The system returns top n images based on $S(q, D_i)$ in Equation (1), where the computation of $S_{long}(q, D_i)$ will be explained in section 2.3.
10. Repeat steps 1 through 9 for a few feedback iterations, or until the user is satisfied with the retrieval results.

Fig. 2: Algorithmic view of fuzzy SVM-based short-term learning.

2.3. Fuzzy-Correlation-Based Long-Term Learning

We compute the long-term-based high-level similarity scores between q and D_i using the updated long-term semantic relations. The detailed steps are as follows:

1. Roughly estimate the semantic similarity between query q and D_i as:
$$Similarity(q, D_i) = \frac{1}{|AllPos|}\sum_{j=1}^{|AllPos|} S_H(D_i, Pos_j) \tag{5}$$
where Pos_j is the j^{th} image in ***AllPos***, and |***AllPos***| is the total number of images in ***AllPos***.
2. Roughly estimate the semantic dissimilarity between query q and D_i as:
$$Dissimilarity(q, D_i) = \frac{1}{|AllNeg|}\sum_{j=1}^{|AllNeg|} O_H(D_i, Neg_j) \tag{6}$$
where Neg_j is the j^{th} image in ***AllNeg***, and |***AllNeg***| is the total number of images in ***AllNeg***.
3. Compute the fuzzy membership f of the semantic similarity between q and D_i by:
$$f = Similarity(q, D_i) - Dissimilarity(q, D_i) \tag{7}$$
Here, $f \in [-1, 1]$. The closer the f is to 1, the higher possibility for D_i being semantically related to q.
4. Compute weight w to evaluate the confidence of f:
$$w = Similarity(q, D_i) + Dissimilarity(q, D_i) \tag{8}$$
Here, $w \in [0, 2]$. The closer the w is to 0, the lower confidence of f. If w = 0 for a (q, D_i) pair, its *Similarity*= *Dissimilarity*=0. It indicates that there is no semantic relation learned between this pair. Therefore, a smaller w means fewer semantic relations learned between this pair and a larger w means more semantic relations learned between this pair.
5. More accurately evaluate the semantic similarity between q and D_i by:
$$RS_H(q, D_i) = \left[\tan^{-1}\left(\frac{\pi}{2} f\right) + 1\right] \cdot w \tag{9}$$
We use the arc tangent operation to convert f to a range close to [0, 2], as shown in Fig. 3(a). This mapping brings f to the same range as w to ensure an appropriate weight is given to each f.
6. Compute the long-term-based high-level similarity scores between q and D_i by:
$$S_{long}(q, D_i) = 0.5 \cdot \sin[0.4\pi(RS_H(q, D_i) - 1.25)] + 0.5 \tag{10}$$
It converts the RS_H value to the range of [0, 1], as shown in Fig. 3(b).



(a) (b)

Fig. 3: Illustrations of the two mapping functions.

3. EXPERIMENTAL RESULTS

We have tested our CBIR system on two subsets of the Corel database. The 1st subset contains 2000 images with 100 images in each of 20 semantic categories such as beach, vehicle, etc. The 2nd subset contains 6000 images with 100 images in each of 60 semantic categories. To facilitate the evaluation process, the CBIR system automatically selects query images and performs the relevance feedback process. Specifically, a retrieved image is automatically classified as relevant if it is in the same semantic category as the query.

We compared our proposed system (the fuzzy SVM- and semantic correlation-based system) with three peer systems, namely, the SVM-based CBIR system, the fuzzy SVM-based CBIR system (using S_{short} in Equation (1)), and the SVM- and semantic correlation-based CBIR system on both two subsets. The four systems are compared in terms of the retrieval precision, which is defined as the ratio between the number of relevant images returned and the total number of image returned. In each experiment, we randomly chose 10% of the database as queries and performed a query session for each chosen query to generate prior semantic knowledge by updating the correlation-based long-term semantic relation. In each query session, we performed 6 iterations with top 20 images returned in each iteration using Equation (1). The four systems were then tested using the remaining images in the 2000-image and 6000-image databases as queries, respectively. No semantic relation is stored during the testing process.

Fig. 4 compares the average retrieval accuracy of the four systems on the 2000-image database. The SVM-based CBIR system does not use the prior semantic knowledge. It clearly shows that our proposed system consistently achieves the best retrieval accuracy in all iterations, while the SVM-based system achieves the worst retrieval accuracy in all iterations. The SVM- and correlation-based CBIR system outperforms the fuzzy SVM-based system.

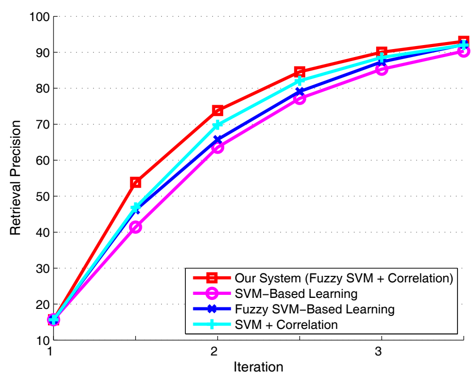


Fig. 4: Comparison of four systems on 2000 images.

Fig. 5 compares the average retrieval accuracy of the four systems on the 6000-image database. It shows the same performance trend except that each system decreases the retrieval accuracy at each iteration due to the large-scale image database.

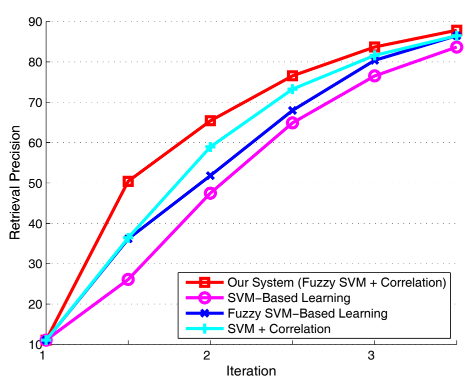


Fig. 5: Comparison of four systems on 6000 images.

These two sets of experiments demonstrate that the correlation-based long-term semantic relation does help in learning the semantic similarity between query image and each database image since both of the two correlation-based systems achieve better retrieval accuracy. The expanded, balanced positive and negative images also make the fuzzy SVM-based system outperform the SVM-based system.

4. CONCLUSIONS AND FUTURE WORK

This paper introduces a long-term semantic learning method for CBIR. Our system uses a statistical correlation-based method to learn the semantic relations between any relevance feedback image pairs. The learned semantic relations are used to expand the feedback set to improve the fuzzy SVM-based low-level learning and compute the high-level semantic similarity between query and each database image. High-level semantic and low-level visual similarity measures are combined to compute the overall similarity score between query and database images. Our extensive experimental results show the effectiveness and scalability of our proposed system.

The semantic clustering technique will be studied to group images into semantic categories to facilitate the learning process and reduce the storage. More heuristic strategies will also be investigated to determine the semantic relations of images.

5. REFERENCES

[1] A. Smeulders, M. Worring, S. Santini, A. Gupta, and J. Ramesh, "Content-Based Image Retrieval at the End of the Early Years," *IEEE Trans. PAMI*, vol. 22, no. 12, 1349-1380, 2000.

[2] S. D. MacArthur, C. E. Brodley, and C. R. Shyu, "Relevance Feedback Decision Trees in CBIR," *Proc. of Workshop on Content Based Access of Image and Video Libraries*, pp. 68-72, 2000.

[3] Z. Su, H. Zhang, S. Li, and S. Ma, "Relevance Feedback in CBIR: Bayesian Framework, Feature Subspaces, and Progressive Learning," *IEEE Trans. IP*, vol. 12, no. 8, pp. 924-936, 2003.

[4] S. Tong and E. Chang, "SVM Active Learning for Image Retrieval," *Proc. of ACM Multimedia*, pp. 107-118, 2001.

[5] K. Wu and K.-H. Yap, "Fuzzy SVM for Content-Based Image Retrieval," *IEEE Computational Intelligence Magazine*, vol. 1, no. 2, pp. 10–16, May 2006.

[6] K. Tieu and P. Viola, "Boosting Image Retrieval," *Proc. of IEEE Int. Conf. on CVPR*, pp. 228-235, 2000.

[7] M. Li, Z. Chen, and H. Zhang, "Statistical Correlation Analysis in Image Retrieval," *Patt. Recog.*, vol. 35, pp. 2687-2693, 2002.

[8] X. He, O. King, W. Y. Ma, M. Li, and H. Zhang, "Learning a Semantic Space from User's Relevance Feedback for Image Retrieval," *IEEE Trans. CSVT*, vol. 13, no. 1, pp. 39-48, 2003.

[9] S. Hoi, M. Lyu, and R. Jin, "A Unified Log-Based Relevance Feedback Scheme for Image Retrieval," *IEEE Trans. Knowledge and Data Engineering*, vol. 18, no. 4, pp. 509-524, 2006.

[10] J. Han, K. N. Ngan, M. Li, and H. J. Zhang, "A Memory Learning Framework for Effective Image Retrieval," *IEEE Trans Image Processing*, vol. 14, no. 4, pp. 511-524, 2005.