AN EFFECTIVE NOISE-RESILIENT LONG-TERM SEMANTIC LEARNING APPROACH TO CONTENT-BASED IMAGE RETRIEVAL

Jacob Linenthal[1] and Xiaojun Qi[2]

[1]jlinenthal@ups.edu
Department of Mathematics and Computer Science, University of Puget Sound, Tacoma, WA 98416
[2]Xiaojun.Qi@usu.edu
Department of Computer Science, Utah State University, Logan, UT 84322-4205

ABSTRACT

This paper proposes a noise-resilient long-term semantic learning method for relevance feedback-based image retrieval. Our system accommodates erroneous feedback resulting from the inherent subjectivity of judging relevance, user laziness, or maliciousness. It also addresses three main drawbacks of traditional relevance feedback techniques. Specifically, it uses a statistical memory learning method based on the user's feedback to extract additional high-level semantic information between query and database images. The learned semantic relationship automatically adds potential positive images to the feedback set to improve SVM-based low-level feature learning. These two measures are seamlessly combined to compute the overall similarity between query and database images. Our experimental results on a 6000-image Corel database demonstrate the effectiveness of the proposed system.

Index Terms— Content-based image retrieval, long-term semantic learning, semantic matrix, noise resilient.

1. INTRODUCTION

With the rapidly growing number of digital images, content-based image retrieval (CBIR) has become an important research area. CBIR techniques can be categorized into global feature-based, object/region-level feature-based, and relevance feedback-based. Recently, relevance feedback has been widely used to bridge the semantic gap between low-level features and high-level semantics. This technique allows the user to label the returned images as positive or negative. Such labeled examples are further used to refine results by query updating techniques including query re-weighting [1], query shifting [2], and query expansion [3] or machine learning techniques such as decision tree learning [4], Bayesian learning [5], SVM [6], boosting [7], etc.

Although relevance feedback techniques can improve the retrieval performance, three drawbacks remain. 1) It cannot capture semantics by simply updating the query concept using low-level features. 2) It cannot achieve good and reliable classification learning by using a small number of imbalanced feedback examples. 3) It cannot remember the semantic knowledge obtained in the feedback processes. To overcome these shortcomings, long-term learning techniques [8-11] have been proposed to store the historical retrieval experiences gained by relevance feedback over many queries to guide the new user's queries. These algorithms incorporate users' subjectivities to provide semantic information. However, the sparsity of memorized feedback information collected from the limited interactions may make long-term learning not useful for a large-scale database. Erroneous feedback also leads to store incorrect semantic information and degrade the retrieval performance.

To address the limitations of the current CBIR systems, we propose a noise-resilient long-term semantic learning method to extract additional semantic information. This technique first gathers users' feedback and stores the semantic similarity among images classified by users. It then estimates the hidden semantic relationship between query and other images, which have not been memorized, using semantic transitivity and the classified positive and negative image sets. These hidden semantic relationships are further used to expand the classified positive image set to improve SVM-based low-level learning. Finally, the normal low-level feature-based and the learned semantic-based similarity measures are combined to improve the retrieval accuracy. A series of improvements are also proposed to accommodate erroneous feedback resulting from the inherent subjectivity of determining semantic relevance, user laziness, or maliciousness. Such enhancements are necessary for ensuring the estimated semantic information is as accurate and as resilient to noisy feedback as possible. The rest of the paper is as follows: Section 2 presents the proposed CBIR system. Section 3 shows the experimental results to demonstrate the viability of our improvements. Section 4 concludes the paper.

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 a specified number of images, which are classified by the user as either relevant or irrelevant to q. This process is continued for a specified number of feedback iterations, or until the user is satisfied with the retrieval results. For each iteration step, the system returns top images ranked by combining low-level feature-based and high-level semantic-based similarity scores. That is, the similarity between query q and an arbitrary image D_i in the database, denoted as $S(q, D_i)$, is defined as:

$$S(q, D_i) = w_L \cdot SimScore_L(q, D_i) + w_H \cdot SimScore_H(q, D_i) \quad (1)$$

where $SimScore_L(q, D_i)$ and $SimScore_H(q, D_i)$ respectively measure the low- and high-level similarity scores between q and D_i; w_L and w_H respectively are the contributing weights assigned to the low- and high-level similarity measures, and $w_L + w_H = 1$. In our system, a higher similarity score means smaller distance in terms of both low- and high-level features and corresponds to stronger similarity.

Here, we define a query session as the overall iterative process to retrieve desired images for a query image. We also define the positive feedback set $\boldsymbol{P}$ as all images classified as relevant by the user during a query session. The negative feedback set $\boldsymbol{N}$ is defined similarly.

2.1. Low-Level Feature-Based Retrieval

The initial retrieval is essential for facilitating quick learning. We use the expanded MPEG-7 edge histogram descriptor (EHD) and the 64-bin HSV-based scaled color descriptor (SCD) to extract low-level features. The inverted weighted and normalized Euclidean distances between EHDs and SCDs of q and D_i are respectively computed to measure $SimScore_L(q, D_i)$. In the following iterations, the positive and negative feedback sets (i.e., $\boldsymbol{P}$ and $\boldsymbol{N}$) are used to train a radial basis function (RBF) kernel SVM classifier. The normalized distance from D_i to the trained separating hyperplane is computed to measure $SimScore_L(q, D_i)$.

2.2. High-Level Semantic-Based Retrieval

Our system transforms users' relevance feedback into the semantic similarity among images. Three observations guide this transformation. 1) If two images returned in the same query session are both marked positive, they belong to the same semantic category as the query image. 2) If one returned image is marked positive while the other is marked 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. Thus, the semantic similarity of two images is defined as:

$$S_H^0(i, j) = P(i, j) / C(i, j) \quad (2)$$

where $P(i, j)$ denotes the number of query sessions where both images i and j are marked as positive and $C(i, j)$ denotes the number of query sessions where both images i and j are returned and at least one is marked as positive.

A semantic matrix of size $M \times M$, where M is the total number of images in the database, stores the similarity measure for each pair of images returned in each query session. Half of the cells are filled due to the symmetric property (for every i and j, $S_H^0(i, j) = S_H^0(j, i)$). Furthermore, this matrix is sparse since an image normally belongs to a few semantic categories relative to the total number of categories present in the image database. An adjacency list stores this sparse matrix to reduce memory requirements.

In order to learn the potential semantic relationship between query and other images that have not been memorized, we utilize semantic transitivity as well as the distance from each un-memorized image to the positive feedback set $\boldsymbol{P}$ in the estimation. Semantic transitivity considers images i and k as semantically related if images i and j are related, and images j and k are also related. Thus, our system uses all the images in $\boldsymbol{P}$ as the intermediate links to estimate the semantic relationship between un-memorized images and query q. The semantic similarity between query q and a database image $D_i \notin P$ is calculated:

$$\begin{aligned} S_H^1(q, \mathrm{D}_i) &= \max\left\{S_H^0(q, \mathrm{D}_i), SD_H^1(D_i, P)\right\} \\ &= \max\left\{S_H^0(q, \mathrm{D}_i), \max\left(S_H^0(\mathrm{D}_i, P_1), \ldots, S_H^0(\mathrm{D}_i, P_{|P|})\right)\right\} \end{aligned} \quad (3)$$

where $SD_H^1(D_i, P)$ denotes the distance from D_i to $\boldsymbol{P}$, P_i is the i^{th} image in $\boldsymbol{P}$, and $|P|$ is the total number of images in $\boldsymbol{P}$. That is, if an image is sufficiently related to one image in $\boldsymbol{P}$, it is considered related to $\boldsymbol{P}$ and therefore related to query q.

However, this learning does not work well when the user incorrectly labels images during a feedback session. For instance, suppose that the user erroneously marks image k as relevant to query q and image k is semantically related to a set of images $\{i_1, i_2, \ldots, i_n\}$. Using (3), all the images i_1, i_2, …, i_n would be considered as semantically identical to q. However, since the semantic link between q and k was erroneous, so too is the connection between q and images i_1, i_2, …, i_n. The inaccurate semantic links may be propagated if more erroneous relevance feedback is provided in the query session. Thus, the effectiveness of the retrieval system will be degraded. To counteract this degradation, we use the average semantic distance from all images in $\boldsymbol{P}$ to D_i ($D_i \notin P$) to compute $SD_H^2(D_i, P)$, the distance from D_i to $\boldsymbol{P}$. Thus, the basic noise-resilient semantic similarity between query q and D_i is computed as:

$$\begin{aligned} S_H^2(q, D_i) &= \max\left\{S_H^0(q, D_i),\ SD_H^2(D_i, P)\right\} \\ &= \max\left\{S_H^0(q, D_i),\ \sum_{j=1}^{|P|} S_H^0(D_i, P_j) \cdot |P|^{-1}\right\} \end{aligned} \quad (4)$$

Here, each image in $\boldsymbol{P}$ is equally weighted. This is not optimal since some semantic similarities are more accurate than others. For example, the semantic information between images i and j is more accurate if $C(i, j)$ is larger (i.e., images i and j have been classified more often). Thus, a weighted scheme is employed to compute the distance from D_i to $\boldsymbol{P}$ (i.e., $SD_H^3(D_i, P)$). The weighted noise-resilient semantic similarity between query q and D_i is computed as:

$$S_H^3(q,D_i)=\max\left\{S_H^0(q,D_i),SD_H^3(D_i,P)\right\}$$
$$=\max\left\{S_H^0(q,D_i),\frac{\sum_{j=1}^{|P|}\max\left\{C(D_i,P_j),\rho\right\}\cdot S_H^0(D_i,P_j)}{\sum_{j=1}^{|P|}\max\left\{C(D_i,P_j),\rho\right\}}\right\} \quad (5)$$

where ρ is the weight given to images which have not been directly given feedback by the user. We experimentally set ρ to be 0.8.

Since some images might be more representative of the query semantic concept than others, we further estimate the representative strength of each image P_i in ***P***. This strength is computed as the average similarity for any image to the remaining images in ***P***. That is, for every image $P_i \in P$, its representative strength is computed as:

$$T(P_i)=\sum_{j=1,j\neq i}^{|P|}S_H^0(P_i,P_j)\cdot(|P|-1)^{-1} \quad (6)$$

The representative strength-based weighted noise-resilient semantic similarity between query q and D_i is computed as:

$$S_H^4(q,D_i)=\max\left\{S_H^0(q,D_i),SD_H^4(D_i,P)\right\}$$
$$=\max\left\{S_H^0(q,D_i),\sum_{j=1}^{|P|}\left[T(P_j)+SD_H^3(P_j,D_i)\right]\right\} \quad (7)$$

where $SD_H^4(D_i,P)$ computes the distance from D_i to ***P*** using the representative strength-based weighted scheme.

2.3. Incorporating the Negative Feedback Set

Negative feedback is incorporated into our system since it contains information about the irrelevant features. Similar techniques are employed to calculate the distance from any database image D_i ($D_i \notin N$) to ***N***. That is, dual operations of equations (3)-(7) are used to compute a series of $SD_H^1(D_i,N), SD_H^2(D_i,N), SD_H^3(D_i,N)$ and $SD_H^4(D_i,N)$. The final semantic similarity between q and D_i is updated as:

$$FS_H^k(q,D_i)=\max\left\{S_H^0(q,D_i),SD_H^k(i,P)-SD_H^k(i,N)\right\},\ k=1,2,3,4 \quad (8)$$

This computation ensures an image, which is semantically similar to ***P*** and not semantically similar to ***N***, is considered as more semantically similar to q. The high-level semantic-based similarity score $SimScore_H(q, D_i)$ in (1) is computed:

$$SimScore_H(q,D_i)=FS_H^k(q,D_i),\ k=1,2,3,or\ 4 \quad (9)$$

where k=4 achieves the best retrieval performance and k=1 achieves the worst retrieval performance.

2.4. Automatically Expanding the Positive Feedback Set

One of the drawbacks of traditional relevance feedback techniques is the imbalance of feedback sets. That is, there are typically more negative than positive feedback examples. This issue makes SVM-based classification learning less accurate and reliable. To address this problem, we use the semantic similarity values obtained from the expanded semantic links to automatically supplement the positive feedback set ***P***. That is, if an image is sufficiently related to ***P***, it should be related to query q and should be automatically added to ***P***. Any database image D_i satisfying the following condition will be added to ***P***:

$$FS_H^k(q,D_i)\geq\mu,\qquad k=1,2,3,4 \quad (10)$$

where μ is an empirically determined value (i.e., 0.8). These additional positive examples help SVM learning to improve classification accuracy and reliability, and provide long-term learning more chances to discover new images for memorizing.

3. EXPERIMENTAL RESULTS

We have tested our CBIR system on a 6000-image Corel database, with 100 images in each of 60 distinct 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. Four experiments have been designed to evaluate the retrieval performance. In each experiment, we randomly chose 10% of the database as queries and performed a query session for each chosen query to construct the semantic matrix. In each query session, we performed 4 iterations of relevance feedback with top 25 images returned in each iteration using (1) with w_L=w_H, where $SimScore_H(q,D_i)$ is computed by either $S_H^k(q,D_i)$ (i.e., without incorporating ***N***) or (9) (i.e., incorporating ***N***). The positive feedback set ***P*** may or may not be expanded for comparison purposes. For each query session, we introduced 5% random noise by having the simulated "user" classify some relevant images as irrelevant and some irrelevant images as relevant. The system is then tested using the remaining 5400 images in the database as queries. No semantic knowledge is stored during the testing.

Experiment 1: Basic memory learning versus low-level learning. Fig. 1 compares the average retrieval accuracy using low-level SVM-based learning (w_L=1 and w_H=0 in (1)) and basic memory learning (i.e., no statistical learning is applied to obtain additional information), where $SimScore_H(q, D_i)$ in (1) is computed by $S_H^0(q,D_i)$ and the system does not incorporate ***N*** and expand ***P***. It shows that basic memory learning achieves better retrieval accuracy than traditional SVM-based learning at each iteration step.

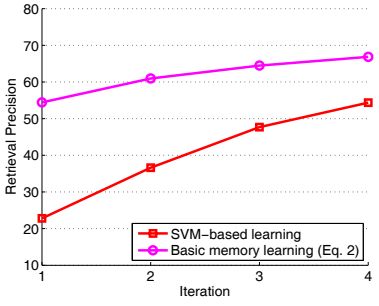

Fig. 1: Accuracy of SVM-based and basic memory learning

Experiment 2: The viability of basic noise-resilient semantic learning (NRSL). Basic semantic learning (SL) [11] and our basic NRSL respectively use $S_H^1(q, D_i)$ (Eq. 3) and $S_H^2(q, D_i)$ (Eq. 4) to compute $SimScore_H(q, D_i)$. Fig. 2 compares the retrieval accuracy of these two systems and the basic memory learning system. Due to erroneous semantic transitivity, the basic SL system degrades its retrieval accuracy. However, the retrieval accuracy of our system increases after each iteration step. It clearly shows the effectiveness of its resilience to noisy feedback. It also improves the basic memory learning in iterations 2 to 4.

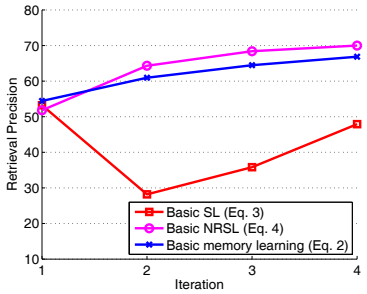

Fig. 2: Comparisons of basic memory learning and two SL systems without incorporating ***N*** and expanding ***P***

Experiment 3: Improvements of basic NRSL by using weight and representative strength. Fig. 3 compares basic NRSL with weighted (Eq. 5) and representative strength-based weighted (Eq. 7) NRSL. It shows the system derived by (7) performs the best and the system derived by (5) performs second best.

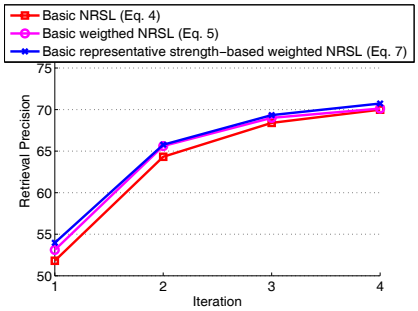

Fig. 3: Comparisons of three NRSL systems without incorporating ***N*** and expanding ***P***

Experiment 4: Improvements of NRSL by incorporating *N* and expanding *P*. Fig. 4 compares three NRSL systems constructed by respectively using (9) with k=2, 3, and 4 as $SimScore_H(q, D_i)$. Additional positive images are automatically added to ***P*** using (10) to improve SVM learning. It shows a similar improvement pattern as Fig. 3. However, a significant retrieval accuracy improvement occurs at the last two iterations when comparing with the three peer systems in Experiment 3.

4. CONCLUSIONS AND FUTURE WORK

This paper introduces a noise-resilient long-term semantic learning method for image retrieval. The proposed system uses a statistical memory learning method to learn additional semantic relationship between query and database images. The learned relationship automatically adds potential positive images to the positive feedback set to improve SVM-based low-level learning. Both high-level semantic and low-level feature similarity measures are combined to compute the overall similarity score between query and database images. Our experimental results demonstrate the effectiveness of the proposed system.

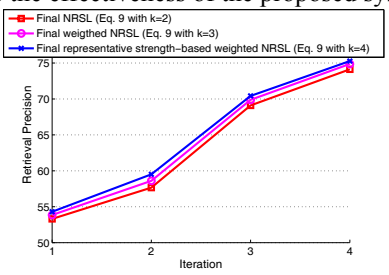

Fig. 4: Comparisons of three NRSL systems incorporating ***N*** and expanding ***P***

One advantage of our system is that the statistical method is effective even when images belong to multiple semantic categories. Clustering techniques will be considered to group images into appropriate semantic categories to facilitate the learning process.

5. REFERENCES

[1] A. Kushki, P. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Query Feedback for Interactive Image Retrieval," *IEEE Trans. on CSVT*, Vol. 14, pp. 644-655, 2004.

[2] P. Muneesawang and L. Guan, "An Interactive Approach for CBIR Using a Network of Radial Basis Functions," *IEEE Trans. on Multimedia*, Vol. 6, No. 5, pp. 703-716, 2004.

[3] D. H. Widyantoro and J. Yen, "Relevant Data Expansion for Learning Concept Drift from Sparsely Labeled Data," *IEEE Trans. on Knowledge and Data Eng.*, Vol. 17, No. 3, pp. 401-412, 2005.

[4] 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.

[5] Z. Su, H. Zhang, S. Li, and S. Ma, "Relevance Feedback in CBIR: Bayesian Framework, Feature Subspaces, and Progressive Learning," *IEEE Trans. on Image Processing*, pp. 924-936, 2003.

[6] S. Tong and E. Chang, "Support Vector Machine Active Learning for Image Retrieval," *Proc. of ACM Int. Conf. on Multimedia*, pp. 107-118, 2001.

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

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

[9] 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. on CSVT*, Vol. 13, No. 1, pp. 39-48, 2003.

[10] S.C.H. Hoi, M. R. Lyu, and R. Jin, "A Unified Log-Based Relevance Feedback Scheme for Image Retrieval," *IEEE Trans. on KDE*, Vol. 18, No. 4, pp. 509-524, 2006.

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