A RADIAL BASIS FUNCTION AND SEMANTIC LEARNING SPACE BASED COMPOSITE LEARNING APPROACH TO IMAGE RETRIEVAL

Konstantin Shkurko[1] *and Xiaojun Qi*[2]

[1]kis9@cornell.edu
Mathematics and Physics Departments, Cornell University, Ithaca, NY 14853
[2]xqi@cc.usu.edu
Computer Science Department, Utah State University, Logan, UT 84322-4205

ABSTRACT

This paper introduces a composite learning approach for image retrieval with relevance feedback. The proposed system combines the radial basis function (RBF) based low-level learning and the semantic learning space (SLS) based high-level learning to retrieve the desired images with fewer than 3 feedback steps. User's relevance feedback is utilized for updating both low-level and high-level features of the query image. Specifically, the RBF-based learning captures the non-linear relationship between the low-level features and the semantic meaning of an image. The SLS-based learning stores semantic features of each database image using randomly chosen semantic basis images. The similarity score is computed as the weighted combination of normalized similarity scores yielded from both RBF and SLS learning. Extensive experiments evaluate the performance of the proposed approach and demonstrate our system achieves higher retrieval accuracy than peer systems.

Index Terms— Radial basis function, semantic learning space, content-based image retrieval

1. INTRODUCTION

Current research on content-based image retrieval (CBIR) aims to narrow down the semantic gap between low-level visual features and high-level meaning. Consequently, relevance feedback techniques are extensively studied to close the gap and quickly retrieve the user's desired images.

Relevance feedback techniques refine query concept based on user's feedback upon the returned top ranked images. Three main approaches are query reweighing, query shifting, and query expansion. Both query reweighing and query shifting apply a nearest-neighbor sampling approach to refine query concept. Specifically, query reweighing [1-3] assigns a new weight to each feature of the query, and query shifting [4-6] moves the query to a new point in the feature space. Query expansion [7, 8] uses a multiple-instance sampling approach to select samples from the neighborhood of the positive labeled instances for learning. However, most of these relevance feedback methods require seeding a query with appropriate positive examples and do not effectively use negative-labeled examples. Furthermore, these systems aim at the short-term learning by exclusively refining low-level features using current feedback step. They do not utilize any previous feedback to gather knowledge for further narrowing of the semantic gap. The semantic-space-based system [9] integrates both short-term and long-term learning to gradually improve the retrieval performance. However, it is computationally intensive and intricate to incrementally construct the semantic space. In addition, that system does not integrate any negative examples, which correspond to the failure of the current classifier in learning.

To address the limitations of current CBIR systems, we propose a composite learning approach to retrieving the desired images using as few feedback steps as possible. To this end, we construct a semantic learning space (SLS) by applying the radial basis function (RBF) based relevance feedback technique on a series of queries composed of randomly selected training images. This SLS remembers the user's intent and, therefore, stores a semantic representation of each database image in terms of presence or absence of the semantics of each training image. We incorporate the retrieval results from the SLS and the query-session-based query shifting technique to refine the retrieved images. Furthermore, information from both positive and negative labeled examples is incorporated into the proposed system to improve learning.

The remainder of the paper is organized as follows: Section 2 describes our proposed approach in detail. Section 3 discusses several experimental results. Section 4 concludes the paper and shows the direction for future work.

2. THE PROPOSED COMPOSITE LEARNING

The block diagram of our proposed method is shown in Fig. 1. The system first returns top 30 images based on the low-level visual similarity between each database image and the query image. The user then indicates relevant and non-relevant images from the returned pool. This is designated as short-term feedback and is performed at each iteration step. The accumulated collection of all short-term feedbacks for the current query search is designated session-term feedback. RBF-based low-level query shifting and SLS-based semantic query modification techniques, explained in sections 2.2 – 2.4, use session-term feedback to respectively search the low-level feature database (LLFD) and the SLS. The system then returns top 30 images ranked by a weighted combination of RBF- and SLS-based normalized similarity scores. The user then labels each returned image for the next iteration until he is satisfied with the results. The following subsections will explain each component of our proposed system in detail.

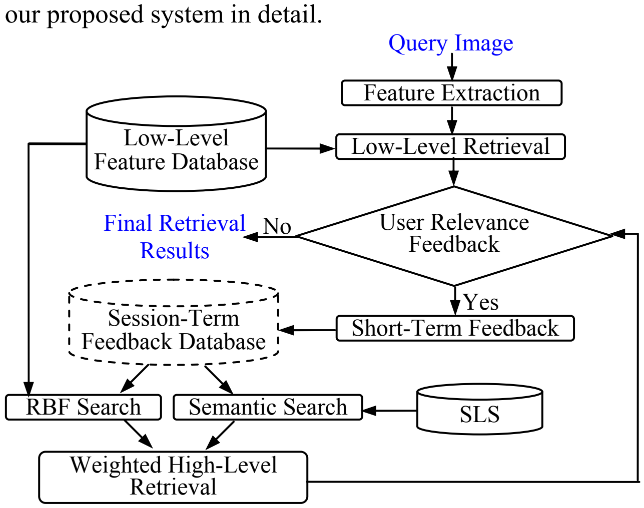


Fig. 1: The block diagram of our CBIR system

2.1. Initial Retrieval

The initial retrieval is essential to later search iterations. We use 192-bin (18×3×3) HSV color histogram, the first three RGB color moments, and 80-bin MPEG-7 edge histogram to extract low-level features for each image in the database. The LLFD stores these low-level features.

To measure the similarity between the query and each image in the LLFD, the inverted Euclidian distance and the histogram intersection is computed for color moments and two histograms, respectively. As a result, higher similarity score corresponds to stronger relevance.

2.2. RBF-Based Low-Level Learning and Search

We employ the RBF to designate images as relevant and non-relevant due to its effectiveness in learning and quick convergence for one-class-relevance classification using a small volume of training sets [6]. Specifically, we use a network of RBFs, with one RBF classifying one low-level feature element, to progressively model query concept for effective searching. The chosen RBF is the Laplacian normal distribution with a controlling parameter σ. This learning and search process consists of the following:

1. Perform initial retrieval (section 2.1) to return 30 images most similar to query image $q(t)$ and empty the session-term feedback database (STFD).
2. Let the user select relevant images, which are most similar to the user's query concept, while regarding the rest of the returned images as non-relevant.
3. Add the exclusively new relevant and non-relevant images to STFD.
4. Modify the query using the low-level features of relevant and non-relevant images in STFD:
$$q(t+1) = \bar{x}^R - \alpha_N\left(\bar{x}^N - q(t)\right) \tag{1}$$
where $\bar{x}^R$ and $\bar{x}^N$ are the average features of relevant and non-relevant images, respectively. The parameter α_N controls the influence of non-relevant images on the update and is empirically set to 0.4 in our system.
5. Use the RBF network to evaluate the similarity in a new search by employing the 2-norm of the Laplacian distance measure:
$$\left\| f_l\left(x, q(t+1)\right)\right\|_2 = \left\| \frac{1}{2\sigma(t+1)} \mathrm{e}^{-\frac{|x-q(t+1)|}{\sigma(t+1)}} \right\|_2 \tag{2}$$
where x is the feature of an image in LLFD, $q(t+1)$ is the modified query feature, and $\sigma(t+1)$ is computed as:
$$\sigma(t+1) = e^{\beta \cdot Std(t+1)} \tag{3}$$
with $Std(t+1)$ as the standard deviation of the relevant images with respect to the average feature of all images in STFD. The parameter β determines the sensitivity amplification of the deviation from the query concept and is set to 2.6 in our system.
6. Return top images based on the similarity scores computed by (2).
7. Repeat steps 2 through 6 until the user is satisfied with the retrieval results.

2.3. SLS Construction

The SLS stores semantic relationships between database images and semantic basis images (SBIs), which are composed of unique, randomly selected training images in each category. These SBIs correspond to the columns of the SLS and the database images correspond to the rows of the SLS. We refer to the number of SBIs as the size of the SLS.

Initially empty, the SLS fills up with relevance feedback information gathered from a number of iterations of searching for each SBI using RBF-based low-level method. To complement the potential limitations of the Laplacian normal distribution, we use the Cauchy distribution as the RBF function in the SLS construction. In addition, all the images returned in each feedback step are exclusively new for the current search to increase the learning diversity and speed. The relevance between a returned image x_i and a SBI b_j updates the SLS via:

$$S_{ij} = \begin{cases} 1 & x_i \text{ and } b_j \text{ are relevant} \\ -1 & x_i \text{ and } b_j \text{ are non-relevant} \\ 0 & \text{All non-returned images} \end{cases} \tag{4}$$

where S_{ij} indicates the value in the i^{th} row and j^{th} column of the SLS. Specifically, the i^{th} row of the SLS is the semantic feature vector (SFV) of database image x_i.

2.4. SLS-Based Semantic Learning and Search

The SLS-based semantic learning and search starts with finding the semantic rows corresponding to the relevant and non-relevant images labeled in the initial retrieval. The query's semantic feature vector (QSFV) is initialized as:

$$q_k^S(t) = (s_k^{R,1} \vee \ldots \vee s_k^{R,Nr}) \wedge \overline{(s_k^{N,1} \vee \ldots \vee s_k^{N,Nn})} \qquad (5)$$

where $q_k^S(t)$ is the k^{th} element of the QSFV, $s_k^{R,i}$ and $s_k^{N,i}$ are the k^{th} element of the SFV of the i^{th} relevant and non-relevant images, respectively. The values of Nr and Nn correspond to the number of relevant and non-relevant images. Here, we treat all negative values as 0's.

For the following feedback iterations, relevant images reinforce the semantically relevant features of the QSFV and non-relevant images suppress the non-relevant features of the QSFV. This process is summarized by:

$$q_i^S(t+1) = \begin{cases} \alpha q_i^S(t) & (s_i^R = 1 \ or \ s_i^N = \text{-}1),\ q_i^S(t) \neq 0 \\ 1 & (s_i^R = 1 \ or \ s_i^N = \text{-}1),\ q_i^S(t) = 0 \\ q_i^S(t) & s_i^R = 0 \ or \ s_i^N = 0 \\ q_i^S(t)/\alpha & s_i^R = -1 \ or \ s_i^N = 1 \end{cases} \qquad (6)$$

where $q_i^S(t+1)$ is the i^{th} element of the updated QSFV, s_i^R and s_i^N correspond to the i^{th} element of the SFVs of the relevant and non-relevant images, respectively. The parameter α is the adjustment rate and is set to 1.1.

The dot product computes the semantic similarity scores between query $q^S(t)$ and database image x_i.

$$S_{x_i}^{Sem} = x_i \cdot q^S(t) = \sum_k x_{i,k} q_k^S(t) \qquad (7)$$

The higher the similarity score, the more semantically relevant the images are.

2.5. Composite Learning

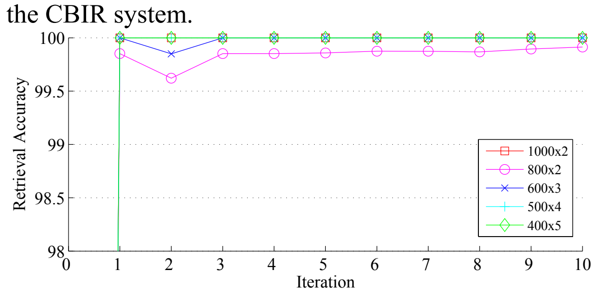
For each feedback, the similarity scores from the RBF-based low-level search and the SLS-based semantic search are combined to obtain top 30 returned images. The min-max normalization defined in (8) scales low-level or semantic similarity score S_i to the range of [0, 1] before combination.

$$S'_i = \frac{S_i - \min(S_i)}{\max(S_i) - \min(S_i)} \qquad (8)$$

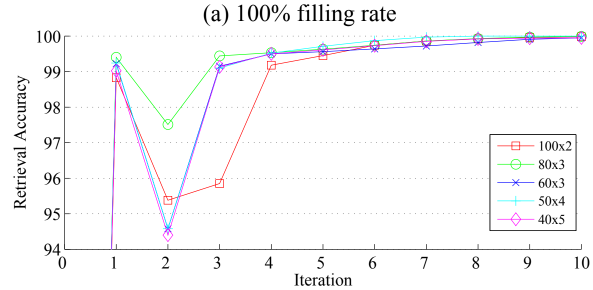
3. EXPERIMENTAL RESULTS

To date, we have tested our CBIR system on 6000 images from COREL. These images have 60 distinct semantic categories with 100 images in each. A retrieved image is considered to be relevant if it belongs to the same category as the query image. To facilitate the evaluation process, we designed an automatic feedback scheme to model the RBF- and SLS-based query session, consisting of 10 iterations. The retrieval accuracy is computed as the ratio of the relevant images to the total returned images. Five experiments have been specifically designed to evaluate the proposed system. The first 4 experiments have been tested on 2000 images in the 20-category COREL database. The initial retrieval accuracy for this subset is 43.9% and is omitted from the figures 2, 3, and 5 to ensure readability of subsequent iterations.

Experiment 1: Optimal filling of the SLS with 200 SBIs. Fig. 2 summarizes the average retrieval accuracy on approximate 100% and 10% fillings with varying number of returned images and iterations in building the SLS. It clearly shows 10% filling rate (i.e., 10% rows are filled in with nonzero values) obtains decent retrieval accuracy with substantially smaller computational cost. The combination of 3 iterations with 80 returned images per iteration performs the best on the 2nd and 3rd iterations. This combination will be chosen to build the SLS, since higher retrieval accuracy at a few feedback steps is most desirable in the CBIR system.



(a) 100% filling rate



(b) 10% filling rate

Fig. 2: Retrieval accuracy of the SLS with 200 SBIs

Experiment 2: Optimal size of the SLS. Fig. 3 shows the average retrieval accuracy for different sizes of the SLS (i.e., 1% to 18% of the database images with the step size of 2%) built by 3 iterations with 80 returned images per iteration. The accuracy is above 90% for all tested sizes. In general, 10% of the database images yield the best retrieval accuracy with rather low computational cost and storage requirement, and therefore, are used to build the SLS.

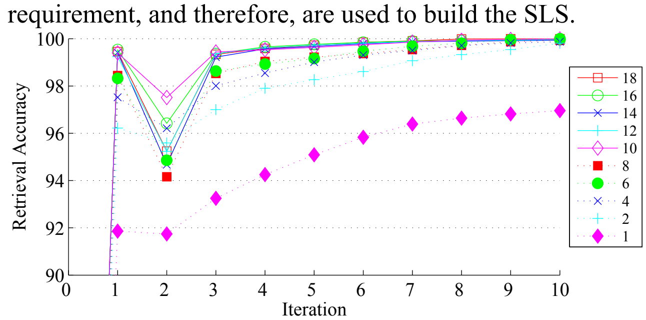


Fig. 3: Retrieval accuracy using different sizes of the SLS

Experiment 3: Selection of RBFs. Fig. 4 shows the average retrieval accuracy of RBF-based low-level search using different normal distributions: Gaussian with exponential powers of 2, 3, and 4, Laplacian, Logistic, and Cauchy. It clearly shows the Laplacian RBF performs the best and Cauchy is the second best. To balance the effects of distributions' shortcomings, Laplacian distribution is used in the RBF network for the low-level search, while the Cauchy distribution is used in the SLS construction.

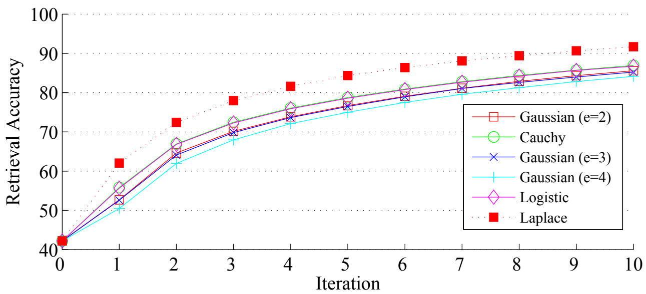


Fig. 4: Retrieval accuracy using RBFs with different distributions

Experiment 4: Combination of RBF- and SLS-based searches. Fig. 5 shows the average retrieval accuracy of combining the normalized similarity scores obtained from both RBF- and SLS-based searches with different ratios. The ratio in the legend is the contribution rate of RBF vs. SLS. It clearly shows the ratio of 4:6 (i.e., 40% RBF and 60% SLS) achieves the best accuracy at the 2^{nd} to the 4^{th} feedback step. Consequently, this combination ratio is used in our CBIR system.

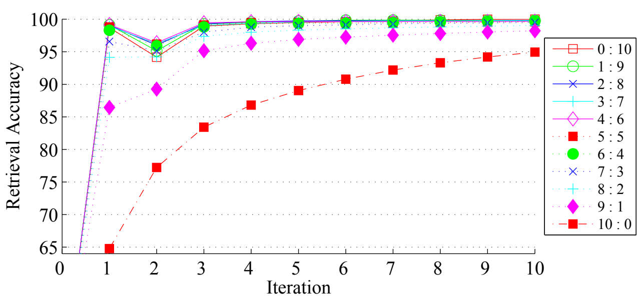


Fig. 5: Retrieval accuracy of combined RBF and SLS searches

Experiment 5: Comparisons with peer CBIR systems. From the first four experiments, we build our retrieval system by combining Laplacian RBF and SLS with a 4:6 ratio, where the SLS is constructed with 10% size and 10% filling rate using Cauchy RBF. We compare our composite approach with Laplacian RBF search (an improved system over [6]), SLS search, and MARS-1 [1] on 6000 COREL images. It shows our approach performs the best in the first 3 iterations. Overall, the two subsystems of our approach also achieve better accuracy than MARS-1.

5. CONCLUSIONS AND FUTURE DIRECTION

A novel CBIR system with relevance feedback is proposed. Major contributions consist of: 1) Constructing the SLS to learn the user's intention with appropriate number of SBIs; 2) Learning the semantic meaning of each database image using SBIs; 3) Combining SLS and RBF searches to achieve better retrieval accuracy with fewer than 3 feedback iterations. Experimental results show the proposed system outperforms peer systems, and, furthermore, the system achieves remarkably high retrieval accuracy for a large database.

The principal component analysis will be considered to update the SLS. An adaptive weight combination will be studied to further improve the retrieval results as well. The performance in a multi-class database, where an image may belong to multiple semantic classes, will be evaluated.

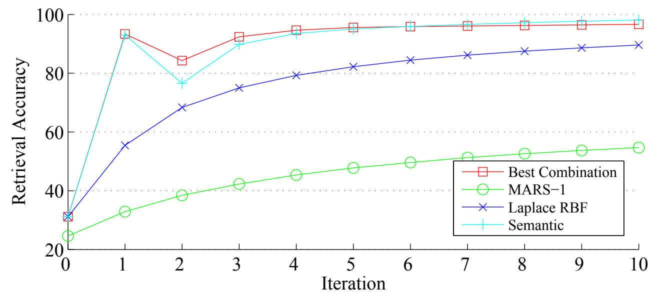


Fig. 6: Comparisons with other CBIR systems

6. REFERENCES

[1] Y. Rui, T. S. Huang, M. Ortega, and S. Mahrota, "Relevance feedback: A Powerful Tool for Interactive Content-based Image Retrieval," *IEEE Trans. on Circuits and Video Technology*, Vol. 8, No. 5, pp. 644-655, 1998.

[2] G. Aggarwal, T. V. Ashwin, and G. Sugata, "An Image Retrieval System with Automatic Query Modification," *IEEE Trans. on Multimedia*, Vol. 4, No. 2, pp. 201-214, 2002.

[3] A. Kushki, P. Androutsos, K. N. Plataniotis, and A. N. Venetsanopoulos, "Query Feedback for Interactive Image Retrieval," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 14, No. 5, pp. 644-655, 2004.

[4] Y. Rui, T. S. Huang, and S. Mahrota, "Content-based Image Retrieval with Relevance Feedback in MARS," *Proc. of Int. Conf. on Image Processing*, pp. 815-818, 1997.

[5] G. Giacinto, F. Roli, and G. Fumera, "Comparison and Combination of Adaptive Query Shifting and Feature Relevance Learning for Content-based Image Retrieval," *Proc. of 11th Int. Conf. on Image Analysis and Processing*, pp. 422-427, 2001.

[6] 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.

[7] K. Porkaew, S. Mehrota, and M. Ortega, "Query Reformulation for Content-Based Multimedia Retrieval in MARS," *Proc. of Int. Conf. on Multimedia Computing and Systems*, Vol. 2, pp. 747-751, 1999.

[8] 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.

[9] X. He, O. King, W. Y. Ma, M. J. Li, and H. J. Zhang, "Learning a Semantic Space From User's Relevance Feedback for Image Retrieval," *IEEE Trans. on Circuits and Systems for Video Technology*, Vol. 13, No. 1, pp. 39-48, 2003.