Biased Subspace Learning for SVM Relevance Feedback in Content-Based Image Retrieval

Lining Zhang
School of EEE
Institute for Media Innovation
Nanyang Technological University
Singapore 639798
zhan0327@ntu.edu.sg

Lipo Wang
School of EEE
Institute for Media Innovation
Nanyang Technological University
Singapore 639798
elpwang@ntu.edu.sg

Weisi Lin
School of Computer Engineering
Institute for Media Innovation
Nanyang Technological University
Singapore 639798
wslin@ntu.edu.sg

Abstract— **Support Vector Machine (SVM) based Relevance Feedback (RF) is one of the most popular techniques for Content-Based Image Retrieval (CBIR). However, it is not appropriate to directly use the SVM as a RF scheme since it treats the positive and negative feedbacks equally. Additionally, it does not take into account unlabelled samples although unlabelled samples are very helpful in constructing a good classifier. To explore solutions to these two problems, we propose a Biased Maximum Margin Analysis (BMMA) and a Semi-Supervised Biased Maximum Margin Analysis (SemiBMMA) combined with SVM RF in this paper. Extensive experiments on a large real world image database demonstrate that the proposed scheme can significantly improve the performance of the traditional SVM-based RF for CBIR.**

Keywords-relevance feedback; content based image retrieval; support vector machine

I. INTRODUCTION

Content-Based Image Retrieval (CBIR) has gained much attention for its potential applications in multimedia management [1]. However, the semantic gap between the low-level visual features and the high-level concepts of images often leads to poor performance of CBIR systems.

To narrow down the so called semantic gap, Relevance Feedback (RF) was introduced as a powerful tool to enhance the performance of CBIR [2]. Huang et al introduced both the query movement and re-weighting techniques [3, 4]. In [5], a one-class Support Vector Machine (SVM) estimated the density of positive feedback samples. Regarding the positive and negative feedbacks as two different groups, classification-based RFs [6, 7] have become popular techniques in the CBIR community.

The two-class SVM, as a small sample learning algorithm, was introduced as a RF technique for CBIR because of its good generalization ability. Guo et al developed a constrained similarity measure for image retrieval [8], which learns a boundary that divides the images into two groups and samples inside the boundary are ranked by their Euclidean distance to the query image. Random sampling techniques were applied to alleviate unstable, biased and overfitting problems in SVM RF [7]. Li et al proposed a multitraining SVM method by adopting a co-training technique and a random sampling method [9].

Nevertheless, most of the SVM RF approaches ignore the basic difference between the two distinct groups of feedbacks, that is, all positive feedbacks share a similar concept while each negative feedback usually varies with different concepts [10]. However, traditional SVM RF techniques treat positive and negative feedbacks equally. Additionally, it is problematic to incorporate the information of unlabelled samples into traditional SVM based RF schemes for CBIR, although unlabelled samples are very helpful in constructing the optimal classifier, alleviating noise and enhancing the performance of the system.

To alleviate the performance degradation when directly using the SVM as a RF scheme for CBIR, we explore solutions based on the argument that different semantic concepts lie in different subspaces and each image can lie in many different concept subspaces [10]. We formally formulate this problem into a general subspace learning problem and propose a Biased Maximum Margin Analysis (BMMA) and a Semi-Supervised Biased Maximum Margin Analysis (SemiBMMA) based on the Graph Embedding framework [11], which is a general platform to formulate new subspace learning algorithms, to integrate the distinct properties of two groups of feedback samples and incorporate the information of unlabelled samples for SVM RF. Experiments on a large Corel Image Dataset have shown significant improvement of the RF performance by the new approaches.

II BMMA AND SEMIBMMA FOR SVM RF IN CBIR

In each round of feedback iteration, there are n samples $X = \{x_1, x_2, \dots, x_n\}$. For simplicity, we assume that the first n^+ samples are positive feedbacks $x_i (1 \le i \le n^+)$, the next n^- samples are negative feedbacks $x_i (n^+ + 1 \le i \le n^+ + n^-)$, and all the others are unlabelled samples $x_i (n^+ + n^- + 1 \le i \le n)$. Let $l(x_i)$ be the class label of sample x_i, we denote $l(x_i) = 1$ for positive feedbacks, $l(x_i) = -1$ for negative feedbacks and $l(x_i) = 0$ for unlabelled samples. To construct the BMMA algorithm, firstly, two different graphs are formed: 1) the intrinsic graph G, which characterizes the local similarity of the feedback samples; 2) the penalty graph G^p , which characterizes the local discriminant structure of the feedback samples.

For all the positive feedbacks, we first compute the pair-wise distance between each pair of positive feedbacks. Then for each positive feedback x_i , we find its k_1 nearest

This work is supported by the Institute for Media Innovation (IMI), Nanyang Technological University, Singapore.

neighborhood positive feedbacks, which can be represented as a sample set N_i^s, and put an edge between x_i and its neighborhood positive feedbacks. Then the intrinsic graph is characterized as follows:

$$\begin{aligned}\widetilde{S}_I &= \sum_i \sum_{j:\, j\in \mathbb{N}_i^s or\, i\in \mathbb{N}_j^s} \| \alpha^T x_i - \alpha^T x_j \|^2 * W_{ij} \\ &= 2tr[\alpha^T X(D-W)X^T\alpha]\end{aligned} \quad (1)$$

$$W_{ij} = \begin{cases} 1/|\mathbb{N}^s|, \ if\ l(i)=1 and\ l(j)=1,\ i\in \mathbb{N}_j^s or\ j\in \mathbb{N}_i^s \\ 0,\ else \end{cases} \quad (2)$$

where D is a diagonal matrix whose diagonal elements are calculated by $D_{ii}=\sum_j W_{ij}$; $|\mathbb{N}^s|$ denotes the total number of k_1 nearest neighborhood positive sample pairs for each positive feedback. Basically, the intrinsic graph measures the total average distance of the $|\mathbb{N}^s|$ nearest neighborhood sample pairs, and is used to characterize the local within-class compactness for all the positive feedbacks.

The penalty graph G^p is constructed to represent the local separability between the positive class and the negative class. More strictly speaking, we expect that the total average margin between the sample pairs with different labels should be as large as possible.

For each feedback sample, we find its k_2 neighbor feedbacks with different labels and put edges between corresponding pairs of feedback samples with weights W_{ij}^p.

Then, the penalty graph can be formed as follows:

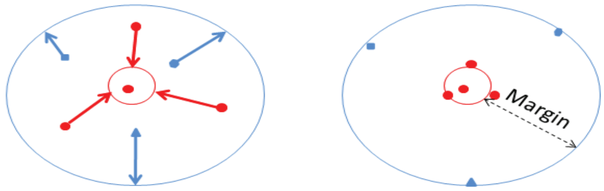
$$\begin{aligned}\widetilde{S}_p &= \sum_i \sum_{j:\, j\in \mathbb{N}_i^p or\, i\in \mathbb{N}_j^p} \| \alpha^T x_i - \alpha^T x_j \|^2 * W_{ij}^p \\ &= 2tr[\alpha^T X(D^p-W^p)X^T\alpha]\end{aligned} \quad (3)$$

$$W_{ij}^p = \begin{cases} 1/|\mathbb{N}^p|, \ if\ l(i)=1\ and\ l(j)=-1,\ i\in \mathbb{N}_j^p or\ j\in \mathbb{N}_i^p \\ 0,\ else \end{cases} \quad (4)$$

where D^p is a diagonal matrix whose diagonal elements are calculated by $D_{ii}^p=\sum_j W_{ij}^p$; $|\mathbb{N}^p|$ denotes the total number of k_2 neighborhood sample pairs with different labels. Similarly, the penalty graph measures the total average distance of the $|\mathbb{N}^p|$ nearest neighbor sample pairs in different class, and is used to characterize the local between-class separability.

In the following, we describe how to utilize the graph embedding framework to develop algorithms based on the designed intrinsic and penalty graphs. Different from the original formulation of the graph embedding framework in [11], the BMMA algorithm optimizes the objective function in a trace difference form instead, i.e.,

$$\begin{aligned}\alpha^* &= \arg\max_\alpha 2tr[\alpha^T X(D^p-W^p)X^T\alpha] - 2tr[\alpha^T X(D-W)X^T\alpha] \\ &= \arg\max_\alpha tr(\alpha^T XBX^T\alpha) - tr(\alpha^T XLX^T\alpha) \\ &= \arg\max_\alpha tr[\alpha^T X(B-L)X^T\alpha]\end{aligned} \quad (5)$$



(a) **(b)**

Figure.1 (a) red dots are positive samples and blue dots are negative samples in the original space (b) the positive samples and negative samples in the maximum margin subspace

In order to remove an arbitrary scaling factor in the projection, we additionally require that α is constituted by the unit vectors, i.e., $\alpha_k^T\alpha_k=1, k=1,2\ldots,l$. This means that we need to solve the following constraint optimization.

$$\begin{aligned}&\max_\alpha tr(\alpha^T X(B-L)X^T\alpha) \\ &= \sum_{k=1}^{l} \alpha_k^T X(B-L)X^T\alpha_k \\ &s.t.\ \ \alpha_k^T\alpha_k - 1 = 0,\ k=1,2,\ldots,l\end{aligned} \quad (6)$$

The motivation for using the constraint $\alpha_k^T\alpha_k=1,\ k=1,2,\ldots,l$ is to avoid calculating the inverse of XBX^T, which leads to the potential "*Small Sample Size*" problem. It is easy to see that the problem (i.e., Eq.6) is a Standard Eigenvalue Decomposition problem.

The BMMA can be illustrated in Fig.1.

In the previous subsection, we have formulated the BMMA algorithm and shown that the optimal projection matrix α can be obtained by Standard Eigenvalue Decomposition on a matrix. Then the problem is how to determine an optimal dimensionality for RF, i.e., the projected subspace. To achieve such a goal, we give the detail of determining the optimal dimensionality.

In general,

$$\max_\alpha tr(\alpha^T X(B-L)X^T\alpha) = \sum_{i=1}^{l}\lambda_i \quad (7)$$

where λ_i's are the associated eigenvalues and we have

$$\lambda_1 \ge \lambda_2 \ge \ldots \ge \lambda_{d-1} \ge 0 \ge \lambda_d \ge \ldots \ge \lambda_l \quad (8)$$

To maximize the margin between the positive samples and negative samples, we should preserve all the eigenvectors associated with the positive eigenvalues. Therefore, the optimal dimensionality of the projected subspace just corresponds to the number of nonnegative eigenvalues of the matrix. Therefore, compared to the original formulation of the graph embedding framework in [11], the new formulation (6) can easily avoid the intrinsic "Small Sample Size" problem and also provide us with a simple way to determine the optimal dimensionality for this subspace learning problem.

Then to incorporate the information of unlabelled samples, we design a regularization term based on intrinsic graph for the unlabelled samples in the image database.

For each unlabelled sample $x_i (n^+ + n^- + 1 \le i \le n)$, we expect that the nearby unlabelled samples are likely to have the similar low-dimensional representations. Specifically, for each unlabelled sample, we find its k_1 nearest neighborhood unlabelled samples,which can be represented as a sample set N_i^u , and put an edge between the unlabelled x_i and its neighborhood unlabelled samples. Then the intrinsic graph for the unlabelled samples is characterized as follows:

$$\begin{aligned} \widetilde{S}_U &= \frac{1}{2} * \sum_i \sum_{j: j \in \mathbb{N}_i^u or i \in \mathbb{N}_j^u} \| \alpha^T x_i - \alpha^T x_j \|^2 * W_{ij}^u \\ &= tr[\alpha^T X (D^u - W^u) X^T \alpha] \\ &= tr[\alpha^T X U X^T \alpha] \end{aligned} \tag{9}$$

$$W_{ij}^u = \begin{cases} \frac{1}{|\mathbb{N}^u|} \exp(-\| x_i - x_j \|^2 / \delta^2), \ if \ l(i) = l(j) = 0, i \in \mathbb{N}_j^u or \ j \in \mathbb{N}_i^u \\ 0, \ else \end{cases} \tag{10}$$

which reflects the affinity of the sample pairs; D^u is a diagonal matrix whose diagonal elements are calculated by $D_{ii}^u = \sum_j W_{ij}^u$; $| D^u |$ denotes the total number of k_1 nearest neighborhood unlabelled sample pairs for each unlabelled sample. $L^u = D^u - W^u$ can be known as a Laplacian matrix. Hence, we call this term as a Laplacian regularizer. Actually, this scheme can preserve weak (probably correct) similarities between all unlabeled sample pairs and thus effectively integrate the similarity information of unlabeled samples into the BMMA. By integrating the Laplacian regularizer into the supervised BMMA, we can easily obtain the SemiBMMA for the SVM RF, i.e.,

$$\alpha^* = \arg\max_\alpha tr[\alpha^T X (B - L - \beta * U) X^T \alpha] \tag{11}$$

where β is used to trade off the contributions of the labeled samples and unlabelled samples. Similarly, the solution of (11) is obtained by conducting the Standard Eigenvalue Decomposition and α is calculated as a set of eigenvectors.

The procedure of SemiBMMA SVM RF for CBIR is illustrated in TABLE I.

TABLE I. THE PROCEDURE OF SEMIBMMA SVM RF FOR CBIR

1) Construct the supervised intrinsic graph G , according to the formulation (1) and calculate the matrix value XLX^T .
2) Construct the supervised penalty graph G^p , according to the formulation (3) and calculate the matrix value XBX^T .
3) Construct the Laplacian regularizer according to the formulation (9) and calculate the matrix value XUX^T .
4) Calculate the projection matrix α^* according to Standard Eigenvalue Decomposition on the matrix $X(B - L - \beta * U)X^T$.
5) Calculate the new representations: project all positive, negative and remaining samples in the database onto the reduced subspace respectively, i.e., $Y^+ = \alpha^{*T} X^+$, $Y^- = \alpha^{*T} X^-$.
6) Train a standard SVM classifier on $[Y^+, Y^-]$

III. IMAGE RETRIEVAL SYSTEM

In experiments, we use a subset of the Corel Photo Gallery, which comprises totally 10,763 real-world images with 80 concepts, as the test data to evaluate the performance of the proposed scheme. The images are represented by low level features, such as color [12], texture [13] and shape [14], each of which can capture the content of an image to some extent. In this paper, we select the color histogram, the Weber Local Descriptors [15] and the edge directional histogram to represent images. All of the features will result in vectors with 510 values to represent images.

IV. EXPERIMENTAL RESULTS

The CBIR system was implemented on a small scale image database and a large scale image database, respectively. We used the average precision, which is the ratio of the number of relevant images retrieved to the top N retrieved images, to evaluate the retrieval performance. We empirically select the parameters $k_1, k_2 = 4$. To implement the semi-supervised method, we randomly select 300 unlabeled samples in each round of feedback iteration. For the trade off parameter between labeled samples and unlabeled samples, we simply set $\beta = 1$ .We choose the Gaussian kernel for all the SVM-based algorithms:

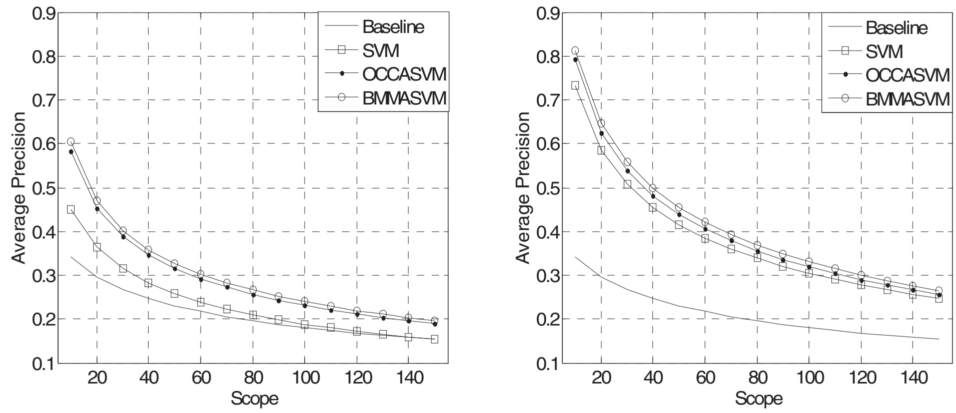
$$K(x, y) = e^{-\rho |x-y|^2}, \rho = 0.001 \tag{12}$$

A. Experiments on a small size image database

The first evaluation experiment is executed on a small size database, which includes 3899 images with 30 different categories. We use all 3899 the images in 30 categories as queries. In practice, the first 5 query relevant images and first 5 irrelevant images in the top 20 retrieved images in the previous iterations were automatically selected as positive and negative feedbacks respectively. Fig.2 shows the average precision-scope curves of the algorithms for the 1st and 2nd iterations. We can notice that both the BMMA SVM and the Orthogonal Complement Component Analysis (OCCA) [19] SVM can perform much better the traditional SVM on the entire scope, especially the 1st round of feedback. The BMMA SVM algorithm and the OCCA SVM algorithm can significantly improve the performance of SVM by treating the positive and negative feedbacks unequally. Moreover, the performance of BMMA SVM and OCCA SVM will be degraded by overfitting after a few iterations.

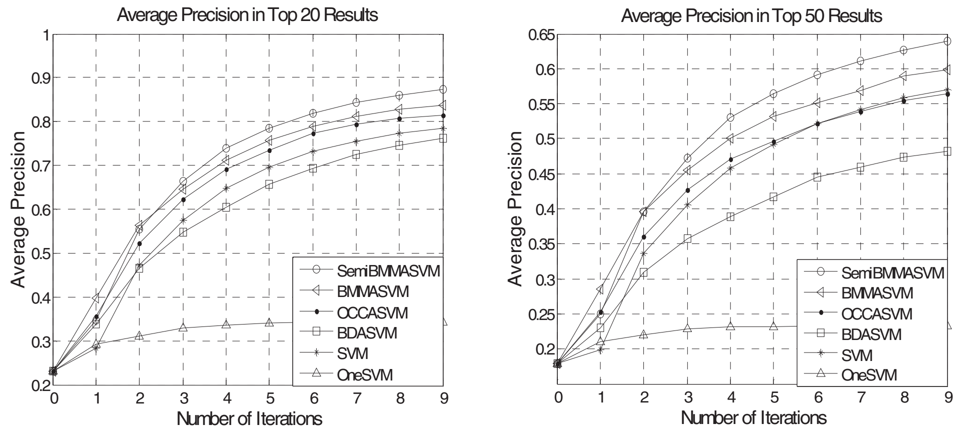
B. Experiments on a large scale image database

In this subsection, we divide the whole image database into five subsets of equal size. Thus, there are 20 percent images per category in each subset. At each run of cross validation, one subset is selected as the query set, and the other four subsets are used as the database for retrieval. Then 400 query samples are randomly selected from the query subset and the relevance feedback is automatically implemented by the system. For each query image, the system retrieves and ranks the images in the database and 9 RF iterations are automatically executed. The first 3 relevant images are labeled



(a) (b)

Figure.2 The precision-scope curves after the 1st feedback and 2nd feedback for SVM, OCCA SVM and BMMA SVM



(a) (b)

Figure.3 The average precisions in top 10- top 60 results of the six approaches from the fivefold cross validation.

as positive feedbacks and all other irrelevant images in top 20 results are automatically marked as negative feedbacks. The first 3 relevant images are labelled as positive feedbacks and all other irrelevant images in top 20 results are automatically marked as negative feedbacks. We show the average precisions in top 20 and top 50 retrieved images in Fig.3 (a) and (b), respectively. To demonstrate the effectiveness of the proposed scheme, we compare them with traditional SVM, the OCCA SVM, the Biased Discriminant Analysis (BDA) SVM, and one-class SVM (OneSVM).

As shown in Fig.3, SemiBMMA SVM outperforms all the other algorithms on the entire scope. Both BMMA and OOCA can improve the performance of the SVM RF as shown in Fig.3. However, both BMMA and OCCA SVM will encounter the overfitting problem, i.e., OCCA combined with SVM will degrade the performance of SVM after a few rounds of feedbacks although they can improve the performance of SVM in the first a few rounds of feedback. The SemiBMMA combined with SVM can significantly improve the performance of the traditional SVM since it can effectively utilize the basic property of different groups of the feedback samples and integrate the information of the unlabelled samples into the construction of the classifier.

V. CONCLUSIONS

Support Vector Machine (SVM) based Relevance Feedback (RF) is one of the most important techniques to improve the performance of a Content-Based Image Retrieval (CBIR) system. To integrate the distinct properties of feedback samples and the information of the unlabelled samples into the SVM RF for CBIR, we have designed a Biased Maximum Margin Analysis (BMMA) and a Semi-Supervised Biased Maximum Margin Analysis (SemiBMMA) for the traditional SVM RF. The novel approaches can distinguish the positive feedbacks and negative feedbacks by maximizing the local margin and integrity the information of unlabeled sample by introducing a Laplacian regularizer. Extensive experiments on a large real world Corel image database have shown that the proposed scheme combined with the traditional SVM RF can significantly improve the performance of CBIR systems.

ACKNOWLEDGMENT

This work is supported by the Institute for Media Innovation (IMI), Nanyang Technological University, Singapore.

References

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, Content-Based Image Retrieval at the End of the Early Years, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 22, No. 1, pp. 1349–1380, Jan.2000.

[2] Y. Rui, T. Huang, M. Ortega, and S. Mehrotra, Relevance Feedback: A Power Tool in Interactive Content-Based Image Retrieval, *IEEE Trans. Circuits Syst. Video Technol.*, Vol. 8, No. 5, pp. 644-655, Sep. 1998.

[3] Y. Rui, T. S. Huang, and S. Mehrotra, Content-Based Image Retrieval with Relevance Feedback in MARS, in *Proc. IEEE Int. Conf. Image Process.*, 1997, Vol. 2, pp. 815–818.

[4] T. S. Huang and X. S. Zhou. Image Retrieval by Relevance Feedback: from Heuristic Weight Adjustment to Optimal Learning Methods. In *Proc. IEEE International Conference on Image Processing (ICIP'01)*, Thessaloniki,Greece, Oct. 2001.

[5] Y. Chen, X.-S. Zhou, and T.-S. Huang, One-class SVM for Learning in Image Retrieval, *In Proceedings of IEEE Int. Conf. Image Processing (ICIP'01)*, pp. 815–818, 2001.

[6] P. Hong, Q. Tian, and T. S. Huang. Incorporate support vector machines to content-based image retrieval with relevant feedback. In *Proc. IEEE International Conference on Image Processing (ICIP'00)*, Vancouver,BC, Canada, 2000.

[7] D. Tao, X. Tang, X. Li, and X. Wu, Asymmetric Bagging and Random Subspace for Support Vector Machines-Based Relevance Feedback in Image Retrieval, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 28, No. 7, pp. 1088–1099, Jul. 2007.

[8] G. Guo, A.K. Jain, W. Ma, and H. Zhang, Learning Similarity Measure for Natural Image Retrieval with Relevance Feedback, *IEEE Trans. Neural Networks*, Vol. 12, No. 4, pp. 811-820, 2002.

[9] J. Li, N. Allinson, D. Tao and X. Li, Multitraining Support Vector Machine for Image Retrieval. *IEEE Trans. On Image Processing*. Vol.15 No.11, Nov.2006

[10] X. Zhou and T. Huang, Small Sample Learning during Multimedia Retrieval Using Biasmap, *In Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition,*Vol.1, pp.11-17,2001.

[11] S. Yan, D. Xu, B. Zhang, H. Zhang, Q. Yang, and S. Lin, Graph embedding and extensions: A general framework for dimensionality reduction, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 29, No. 1, pp.40–51, Jan. 2007.

[12] M. J. Swain and D. H. Ballard, Color Indexing, *International. Journal of Compututer. Vision.,* Vol.7, No. 1, pp. 11–32, 1991.

[13] H. Tamura, S. Mori, and T.Yamawaki, Texture Features Corresponding to Visual Perception, *IEEE Trans. Syst., Man, Cybern.*, Vol. 8, No.6, pp. 460–473, Jun. 1978.

[14] A. Jain and A.Vailaya, Image Retrieval Using Color and Shape, *Pattern Recognit.*, Vol. 29, No. 8, pp. 1233–1244, Aug. 1996.

[15] J. Chen, S. Shan, G. Zhao, X. Chen, W. Gao, Matti Pietikainen, WLD: A Robust Descriptor based on Weber's Law, *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 32, No.9, pp. 1705–1720, Sep. 2010.

[16] D. Tao, X. Tang and X. Li, Which Components Are Important for Interactive Image Searching? *IEEE Trans. Circuits System Video Technology*, Vol. 18, No.1, Jan. 2008.