Improving Retrieval Performance by Region Constraints and Relevance Feedback

Tao Wang[1], Yong Rui[2], and Jia-Guang Sun[1]

[1]*Department of Computer Science and Technology, Tsinghua University, Beijing 100084, P.R. China*
[2]*Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA*

E-mail: wangtao@cg.cs.tsinghua.edu.cn; yongrui@microsoft.com; sunjg@mail.tsinghua.edu.cn

Received August 27, 2002; revised March 24, 2003.

Abstract In this paper, region features are used as constraints and relevance feedback to improve the performance of CBIR. Unlike existing region-based approaches where either individual regions are used or only simple spatial layout is modeled, the proposed approach simultaneously models both region properties and their spatial relationships in a probabilistic framework. Furthermore, the retrieval performance is improved by an adaptive filter based relevance feedback. To illustrate the performance of the proposed approach, extensive experiments have been carried out on a large heterogeneous image collection with 17,000 images, which render promising results on a wide variety of queries.

Keywords content-based image retrieval (CBIR), region matching, probabilistic weight estimation, relevance feedback, adaptive filter

1 Introduction

Content-based image retrieval (CBIR) is the process of retrieving images based on visual features that are automatically extracted from images. It has been one of the most active research areas in computer science in the past decades[1,2], and many CBIR systems have been developed, e.g., QBIC[3], PhotoBook[4], VisualSEEK[5], MARS[6], Netra[7], BlobWorld[8], SIMPLIcity[9] etc. Despite years of extensive research, however, assisting users to find their desired images accurately and quickly is still an open problem. The main challenge is the semantic gap between high level query concepts users want and low level features we can extract[33].

In most existing systems, global features such as global color histogram, texture and shape are broadly used[3,5,10]. While global features are easy to implement, they often fail to narrow down the semantic gap because of their limited description power based on objects. Compared with global features, local features have strong correlation with objects, which are prospective to provide a big step towards the semantic-based retrieval.

Existing local-feature-based image retrieval approaches fall into three categories: fixed-layout-based[11], salient-point-based[12,13] and region-based approaches[8,14−18]. Among various local-feature-based approaches, the region-based approach so far has been the winning approach since the images with the same content normally have similar objects and spatial relationship. This approach first segments an image into multiple regions that have high correlation with real-world objects. Then the total similarity between two images is calculated based on the corresponding regions. Because of imperfect image segmentation (e.g., over or under segmentation) and complex spatial relationship, how to construct an accurate and robust similarity model is the main problem. The Netra system compares images based on individual regions[7,19]. Although queries based on multiple regions are allowed, the retrieval is done by merging individual region's query results, which is therefore less robust to imperfect segmentation. The SaFe system, on the other hand, uses a 2D-string approach[20]. While this system is more robust than Netra, it is sensitive to region shifting and rotating. A better approach, the integrated region matching (IRM)[9] is more accurate in similarity calculation and more robust to imperfect image segmentation by allowing a region to be matched to multiple regions from another image. While IRM makes a significant step in region-based matching, its accuracy and robustness suffer from the fact that it does not take into account the spatial relationships between regions. In addition, it uses a greedy algorithm to estimate inter-region weight

*The first author is supported in part by the National Natural Science Foundation of China under Grant No.69902004 and the National Project 2001BA201A07.

w_{ij}, which can easily be trapped into local minimum and all the subsequent weight estimations become unreliable.

In this paper, a novel constraint-based region matching approach to image retrieval (CRM) is proposed. Different from those region-based retrieval approaches, we take into account not only the first-order constraints (e.g., region features) but also the second-order ones (e.g., spatial relationship between them) based on a principled probabilistic framework. Then we furthermore improve retrieval performance by incorporating an adaptive filter based relevance feedback. The combination of better similarity models/features and relevance feedback technique results in a significantly better performance than the existing approaches.

The rest of the paper is organized as follows. In Section 2, we give detailed descriptions of our proposed constraint-based region matching (CRM) approach. In Section 3, we present a framework that integrates relevance feedback learning into CRM. In Section 4, extensive experiments over a large heterogeneous image collection with 17,000 images are reported. Finally, concluding remarks are given in Section 5.

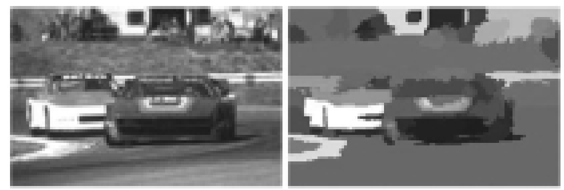
2 Constraint-Based Region Matching

Object and relationship modeling has been extensively studied for the past three decades in spatial layout planning[21,22]. Pfefferkorn[21] defines three important components. (1) *Objects.* Examples are tables and floor. (2) *Specified areas.* They encode domain knowledge of the position of objects, e.g., floor is at the bottom. (3) *Spatial relationship constraints.* For example, a table is on top of a floor.

Because the first two constraints concern only the individual objects, we call them *first-order* constraints. Similarly, we call the spatial relationship constraints *second-order* constraints, as they involve two objects. Even though the above concepts were first developed in spatial layout planning, the ideas apply well to characterize the image's content. Namely, if two images have similar objects and spatial relationships, their content should be highly relevant. In the rest of this section, we present detailed description of the constraint-based region matching approach (CRM) which integrates both 1st and 2nd order constraints to estimate matching weight in a probabilistic frame.

2.1 Image Segmentation

Image segmentation divides an image into regions that have strong correlation with real-world objects. We use a region growing approach to segment images based on HSV color model[23]. An example segmentation result is shown in Fig.1.



(a) (b)

Fig.1. A car image (a) and its segmentation (b).

More sophisticated segmentation techniques can be used, e.g., Pavlidis and Liow[24], and Felzenszwalb *et al.*[25]. But that is beyond the scope of this paper.

2.2 Region Features

After image segmentation, the color, size, shape, position and other features of each region are extracted to represent the content of an image.

1) *Color feature.* The color feature of a region is its mean color in HSV color space.

$$\boldsymbol{C} = (\bar{s}\sin(\bar{h}), \bar{s}\cos(\bar{h}), \bar{v}). \tag{1}$$

2) *Size feature.* The size feature is represented by the area percentage of the region to the whole image.

$$\rho = \frac{\text{Area of the region}}{\text{Area of the image}}. \tag{2}$$

3) *Shape feature.* The shape feature is region's eccentricity e, which is the ratio of the minor axis length $I_{\min}$ and the major axis length $I_{\max}$ of the best-fit ellipse of the region[26].

$$\begin{aligned} e &= \frac{I_{\min}}{I_{\max}} \\ &= \frac{u_{20} + u_{02} - \sqrt{(u_{20} - u_{02})^2 + 4u_{11}^2}}{(u_{20} - u_{02})^2 + 4u_{11}^2} \in [0, 1] \end{aligned} \tag{3}$$

where $u_{pq} = \sum_{(x,y)\in region}(x - \bar{x})^p(y - \bar{y})^q$, and $(\bar{x}, \bar{y})$ is the center of the region. Given that segmented regions may not be accurate in boundary, litter benefit will be gained by using more sophisticated shape features, such as region's contour and its Fourier descriptor[27].

4) *Position feature.* The position of the region is the normalized region center.

$$\boldsymbol{O} = \Big(\frac{\bar{x}}{W}, \frac{\bar{y}}{H}\Big) \tag{4}$$

where W and H are image width and height.

2.3 Regional and Spatial Constraints

Let r_i, r_k be regions in image 1 and r_j, r_l be matched regions to them respectively in image 2. The 1st and 2nd order constraints and their associated similarities are defined as follows:

1) *Region property constraint.*

$$S_c(r_i, r'_j) = \exp(-\|\boldsymbol{C}_i - \boldsymbol{C}_j\|^2/2\sigma_c^2) \tag{5}$$

$$S_e(r_i, r'_j) = \exp(-\|e_i - e_j\|^2/2\sigma_e^2) \tag{6}$$

$$S_p(r_i, r'_j) = \exp(-\|\boldsymbol{O}_i - \boldsymbol{O}'_j\|^2/2\sigma_p^2). \tag{7}$$

2) *Spatial relationship constraint*:

$$S_o(r_i, r_k, r'_j, r'_l) = \Big(\frac{(\boldsymbol{O}_i - \boldsymbol{O}_k)\cdot(\boldsymbol{O}'_j - \boldsymbol{O}'_l)}{\|\boldsymbol{O}_i - \boldsymbol{O}_k\| \cdot \|\boldsymbol{O}'_j - \boldsymbol{O}'_l\|} + 1\Big)\Big/2 \tag{8}$$

$$S_i(r_i, r_k, r'_j, r'_l) = \overline{r_i \text{ in/out } r_k)\ (\text{XOR}(r'_j \text{ in/out } r'_l)} \tag{9}$$

$$S_s(r_i, r_k, r'_j, r'_l) = \Big(\frac{(\rho_i, \rho_k)\cdot(\rho'_j, \rho'_l)}{\|(\rho_i, \rho_k)\| \cdot \|(\rho'_j, \rho'_l)\|} + 1\Big)\Big/2 \tag{10}$$

where σ_c, σ_e and σ_p are features' standard variants to control the penalty of different variations respectively. The total similarity based on the *first-order* constraints, e.g., (5)–(7), for two regions r_i and r_j, is defined as:

$$S_1(r, r'_j) = w_c S_c(r_i, r'_j) + w_e S_e(r_i, r'_j) + w_p S_p(r_i, r'_j) \tag{11}$$

$$\text{s.t.,}\ \ w_c + w_e + w_p = 1$$

Similarly, the total similarity based on the *second-order* constraints, e.g., (8)–(10), is

$$\begin{aligned} S_2(r_i, r_k, r'_j, r'_l) =& w_o S_o(r_i, r_m, r'_j, r'_l) \\ &+ w_i S_i(r_i, r_k, r'_j, r'_l) \\ &+ w_s S_s(r_i, r_k, r'_j, r'_l) \end{aligned} \tag{12}$$

$$\text{s.t.,}\ \ w_o + w_i + w_s = 1$$

where w_c, w_e, w_p, w_o, w_i, and w_s are proper weights for corresponding constraints. In our current implementation, we experimentally find that $w_c = 0.5$, $w_e = 0.3$, $w_p = 0.2$, $w_o = 0.4$, $w_i = 0.3$, and $w_s = 0.3$ give good results on a 17,000 image database.

It is worth mentioning that the proposed CRM is a general approach. The 1st order constraints can be texture and semantic features besides color, shape features etc. Similarly, the 2nd order constraints can be adjacent, contained, semantic features etc. The invariance of translation, rotation and scaling depends on the used constraints.

2.4 Probabilistic Weight Estimation

After image segmentation, let images 1 and 2 be represented by region sets $R_1 = \{r_1, r_2, \ldots, r_M\}$ and $R_2 = \{r'_1, r'_2, \ldots, r'_N\}$ respectively, shown in Fig.2, where M and N are the number of regions in the two images. To increase region matching robustness against inaccurate segmentation, a region in one image can be matched to multiple regions from another image[9]. The total similarity between two images can therefore be defined as the similarity between the two region sets $S(R_1, R_2)$:

$$S(R_1, R_2) = \sum_{i=1}^{M}\sum_{j=1}^{N} w_{ij} S_1(r_i, r'_j)$$

$$\text{s.t.,}\ \sum_{i=1}^{M}\sum_{j=1}^{N} w_{ij} = 1 \tag{13}$$

where weight w_{ij} indicates the importance of region pair (r_i, r'_j) with respect to the overall similarity. It is clear that the weight w_{ij} plays an important role in determining the overall similarity value.

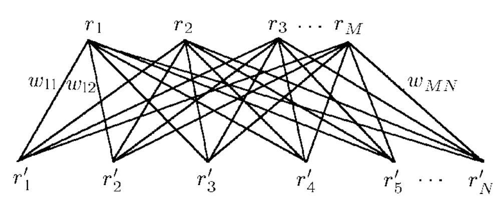


Fig.2. Region matching between region sets R_1, R_2.

Note that all the similarities defined in (5)–(12) are in the range of $[0, 1]$ and the similarity between two entities can be interpreted as the probability of the two entities being similar. Let $x \sim y$ and $x \approx y$ denote that x matches y based on the *first-order* and *second-order* constraints respectively. We therefore have:

$$P(r \sim r'_j 0 \doteq S_1(r_1, r'_j) \tag{14}$$

$$P(r_1 \approx r'_j, r_k \approx r'_l) | r_i \sim r'_j, r_k \sim r'_l) \doteq S_2(r_i, r_k, r'_j, r'_l) \tag{15}$$

where $\doteq$ stands for "can be estimated by".

It is intuitive that a region pair (r_i, r_j) should receive higher weight w_{ij} in the similarity model if it is a better matching pair. In (15), the 2nd constraint involves two pairs of matched regions, e.g. (r_i, r_j) and (r_k, r'_l). Considering all of possible related regions r_k, r'_l to (r_i, r'_j), $P(r_i \sim r'_j, r_i \approx r'_j) = \sum_{k=1}^{M} \sum_{l=1}^{N} P(r_i \sim r'_j,\ r_k \sim r'_l, r_i \approx r'_j,\ r_k \approx r'_l)$. $P(r_i \sim r'_j, r_i \approx r'_j)$ is the probability with which region r_i matches r'_j based on both 1st order (region features) and 2nd order (spatial relationship) constraints. It is therefore a good estimation for w_{ij}. According to (14) and (15), we have:

$$\begin{aligned} W_{ij} &\doteq P(r_i \sim r'_j, r_i \approx r'_j) \\ &= \sum_{k=1}^{M} \sum_{l=1}^{N} P(r_i \sim r'_j, r_k \sim r'_l, r_i \approx r'_j, r_k \approx r'_l) \\ &= \sum_{k=1}^{M} \sum_{l=1}^{N} P(r_i \approx r'_j, r_k \approx r'_l | r_i \sim r'_j, r_k \sim r'_l) \\ &\quad \cdot P(r_i \sim r'_j, r_k \sim r'_l) \\ &\doteq \sum_{k=1}^{M} \sum_{l=1}^{N} S_2(r_l, r_k, r'_j, r'_l) P(r_i \sim r'_j) P(r_k \sim r'_l) \\ &\doteq \sum_{k=1}^{M} \sum_{l=1}^{N} S_2(r_i, r_k, r'_j, r'_l) S_1(r_i, r'_j) S_1(r_k, r'_l) \end{aligned} \tag{16}$$

In the above derivation, we use Bayesian rule in Step 3 and the independence assumption between region pairs (r_i, r'_j), and (r_k, r'_l) in Step 4.

Different regions may have different interested importances in an image. To normalize the summary of W_{ij} to 1 and set the user-interested weight q_i to region r_i, we further define the normalized weight w_{ij} as:

$$w_{ij} = q_i \times \frac{W_{ij}}{\sum_{j=1}^{N} W_{ij}} \qquad \text{s.t.,} \ \sum_{i=1}^{M} q_i = 1 \tag{17}$$

In our current implementation, we initialize user-interested weight as $q_i = 1/M$ (experimentally assuming that every region is equally important in the query image). Or they can be dynamically learned by using relevance feedback (see Section 3).

Examining this new weight estimation technique, CRM uses a principled way to estimate the weights based on probabilities, which is more robust to inaccurate image segmentation. Furthermore, it integrates both first-order and second-order constraints and avoids the shortcoming of the greedy algorithm used in IRM.

3 Relevance Feedback

Relevance feedback can greatly improve the retrieval performance whose basic idea is to use a learning mechanism that adapts image features and similarity measures to best reflect high-level query concepts[10,28,29]. In Section 2, we have presented in detail how CRM supports a better similarity model and image features. In this section, we will further improve the retrieval performance by incorporating an adaptive filter based relevance feedback, i.e., a relevance feedback enhanced CRM.

3.1 Interested Region Weights q_i

Let R_1 represent the region set of the query image and R_2 represent the region set of an image in the database. Referring to (13), it is assumed that all the regions in R_1 are of equal importance to a user. This is a reasonable assumption to start with, but in reality a user may have different interests in different regions of an image. For example, he or she may be more interested in some regions (e.g., a car running on a road) of the image than the other regions (e.g., buildings and trees near the car). By on-line relevance feedback to adjust the interested region weights, the system dynamically learns user's query concepts (e.g., a car on a road) to perform better.

Given the interested weight q_i which models a user's degree of interest in region i of the query image, a more generic similarity between the query image and an image in the database is defined:

$$\begin{aligned} S(R_1, R_2) &= \sum_{i=1}^{M} q_i s_1(r_i, R_2) \\ &= \sum_{i=1}^{M} q_i \sum_{j=1}^{N} w'_{ij} S_1(r_i, r'_j) \end{aligned}$$
$$\text{s.t.} \ \sum_{i=1}^{M} q_i = 1, \ \sum_{j=1}^{N} w'_{ij} = 1, \quad i = 1, 2, \ldots M \tag{18}$$

with

$$w'_{ij} = W_{ij} \Big/ \sum_{j=1}^{N} W_{ij}$$

$$W_{ij} = \sum_{k=1}^{M} \sum_{l=1}^{N} q_l q_k S_2(r_i, r_k, r_j, r'_l) S_1(r_i, r'_j) S_1(r_k, r'_l)$$

where, $s_1(r_i, R_2)$ is the aggregated similarity of region r_i. Particularly, the relationship weight W_{ij} is also affected by adjusting the interested region weight q_i of query image. Compared with (13), the improved similarity model (18) is more flexible and accurate to model user's query content due to decreasing the effect of not interested objects and relationship between them.

3.2 Clustering Based Query Examples

In CBIR, images with similar region features are normally identical in content. On the contrary, images with the same content, however, may be different in region features. For example, according to the query content of a *car*, an image of a red car on a road is the same as another image of a blue truck near a building although they are very different in region features and spatial relationship. Motivated by this, instead of using only one averaged image as the query, we represent user's query concepts as a few clustered patterns C_k $(k = 1, 2..K)$. If an image is similar to any of them, it is similar to the query example. By using these clustering patterns C_k as query examples, we can find more relative images to any of them.

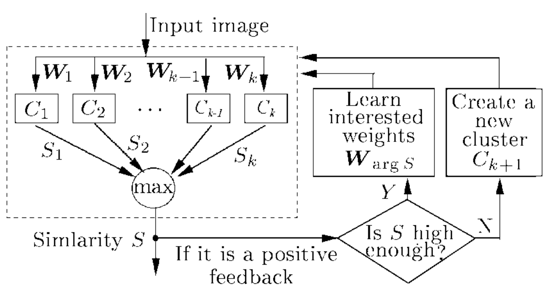


Fig.3. The framework of clustering based relevance.

The framework of clustering-based relevance feedback is described in Fig.3, where $\boldsymbol{W}_k = [q_1, q_2, \ldots, q_M]_k^T$ is the interested weight vector, and the finial output is the maximum similarity to all clustering patterns by (18). The more positive relevance feedbacks are learned, the more cluster patterns are created. When increasing clustering patterns improves retrieval performance, it also decreases the retrieval speed due to more calculations for comparison. To decrease the calculation, we set the maximum number of clustering patterns as K (e.g., experimentally set $K = 8$). When a new clustering pattern is created, we remove one of the worst patterns from $K+1$ patterns and keep the total number of clustering patterns still to be K. The problem of clustering pattern selection is just like the feature selection. Given a set of candidate patterns, select a subset that performs the best under some classification standard. In [30], the sequential forward floating selection (SFFS) method has been found to be extremely powerful for feature selection[31]. Using this SFFS algorithm, K best informative query patterns can be selected.

3.3 Learn Weights q_i by Adaptive Filter

The least-mean-square (LMS) filter relevance feedback algorithm is elegant in theory, easy to implement and requires very little computation[32]. It has the additional benefit of supporting on-line learning, which can learn recursively when a new example arrives.

In Fig.4, $\boldsymbol{X}(n)$ is the input features of both the unkonwn visual system and our simulator – adaptive filter. The output from the unknown system is $d(n)$, and the output from the adaptive filter is $y(n) = \boldsymbol{W}(n)^T\boldsymbol{X}(n)$. By comparing $y(n)$ with $d(n)$, we can obtain an error signal $e(n)$, which can then be used to drive the filter to automatically adapt to the minimum mean square error (MMSE) solution of the estimated linear similarity model's parameters efficiently without any prior knowledge on data distribution. The LMS filter algorithm can be summarized as follows:

(I) Initialization:

Choose step size $0 < \mu < 2$, and set the filter coefficients to

$$\boldsymbol{W}(0) = [1/M, 1/M, \ldots, 1/M]^T \tag{19}$$

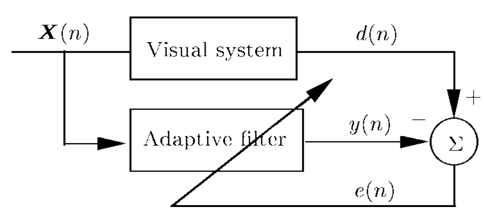


Fig.4. An adaptive filter based feedback model.

(II) For each $n = 1, 2, \ldots, N$ feedback samples,

i. Compute the distance $y(n)$ based on the current weights

$$y(n) = \boldsymbol{W}^T(n)\boldsymbol{X}(n) \tag{20}$$

ii. Compute the error signal

$$e(n) = d(n) - y(n) \tag{21}$$

iii. Compute the updated weights

$$\boldsymbol{W}(n) = \boldsymbol{W}(n-1) + \frac{\mu}{a + \boldsymbol{X}^T(n)\boldsymbol{X}(n)}\boldsymbol{X}(n)e(n) \tag{22}$$

where a is a small positive constant to avoid denominator to be 0. In our specific context, we have

$$\boldsymbol{W} = [q_1, q_2, \ldots, q_M]^T,$$
$$\boldsymbol{X} = [s_1, s_2, \ldots, s_M]^T,$$
$$s_i = S_1(r_i, R_2) = \sum_{j=1}^{N} w'_{ij} S_1(r_i, r'_j).$$

For the pattern C_k, the total similarity $S(R_1, R_2)$ is the linear function of the aggregated similarity $s_1(r_i, R_2)$ weighted by interested region weight q_i. Given user's relevance feedback value $d(n)$ and calculated $s_1(r_i, R_2)$, $i = 1, .., M$, the interested weight q_i of the query pattern is learned by the LMS adaptive filter.

In the clustering-based relevance feedback frame (see Fig.3), each clustering pattern C_k stores its own interested region weights q_i. If a positive feedback image is not similar enough to any pattern by (18), a new clustering pattern is created by it. Otherwise, assuming C_k is the best similar pattern to it, parameters q_i of C_k are learned by the feedback image. Although visual similarity model may be a complex non-linear model, each clustering pattern C_k spans so small feature subspace that is well approximated as a linear model on which the optimal adaptive filter can be implemented better. Detailed experiments on the performance of this technique are reported in Subsection 4.5.

4 Experiments

To validate the presented approaches, extensive experiments have been carried out on a large heterogeneous image collection. First, we will examine and compare the retrieval performance of CRM against some existing approaches. Then, we compare the relevance-feedback enhanced CRM against the plain CRM.

4.1 Data Set

For all the experiments reported in this section, the Corel image collection is used as the test data set. We choose this data set with the following considerations:

• It is a large-scale and heterogeneous data set. The data set includes 17,000 images which covers a wide variety of content from animals and plants to natural and cultural images.

• It is professionally-annotated by Corel professionals. All the images have been classified into 170 categories and there are 100 images in each category.

The Corel data set has also been used in other systems and relatively high retrieval performance has been reported[6,9,13]. However, those systems only use pre-selected categories (e.g., cars vs. mountains) with distinctive visual characteristics. In our experiments, no pre-selection is made for 17,000 images. Since average users want to retrieve images based on high-level concepts, not low-level visual features[33], the ground truth we use is based on high-level categories such as car, flower, people etc. In experiments, in order to obtain an objective evaluation of the different retrieval techniques, we use the categories to evaluate the retrieval performance. But in practice, the system is to enable user to guide the system to the images that are meaningful, while not being subjected to categorization.

4.2 Queries

To fairly evaluate the retrieval performance of different methods and similarity models, we randomly generated 400 queries in all the 17,000 images for each retrieval condition. For all the experimental results reported in this section, they are the average of all the 400 random query results.

4.3 Performance Measures

The most widely used performance measures for information retrieval are precision (Pr) and recall (Re)[34]. In general, Pr will decrease when Re increases. The performance of an "ideal" system should be that *the precision is higher at the same recall value*. Because of this, the $Pr(Re)$ curve is also used to better characterize the performance of a retrieval system.

4.4 CRM vs. Existing Approaches

Tested on a PC with PIII 1GHz CPU and 256MB memory, the average time for one query is 1.66 seconds for IRM and 3.61 seconds for CRM over 17,000 images. Table 1 compares the image retrieval results using CRM and IRM[9], color histogram[3] and wavelet texture[6]. In IRM, we use the same color, shape and position features as CRM. For the wavelet-based texture, the original image is fed into a wavelet filter bank, and decomposed to the third level, resulting in 10 decorrelated sub-bands. For each sub-band, we extract the standard deviation of the wavelet coefficients and therefore have a texture feature vector of length 10. The wavelet-based texture feature has been proven to be quite effective in modeling image features[6,32].

Table 1. Retrieval Performance Comparison Between CRM and IRM, Color Histogram, Wavelet texture

Precision (%)	Return top 20	Return top 100	Return top 180
CRM	17.30	9.83	6.91
IRM	15.96	8.53	6.49
Color	14.68	7.88	6.96
Texture	12.96	5.83	5.27

To better compare the two approaches, we also plot their $Pr(Re)$ curves in Fig.5. Fig.6 shows example comparisons between CRM and IRM, where the top-left image is the query image. Based on the above table and figures, the following observations can be made:

• The CRM approach performs better than the color histogram and wavelet texture global features. This demonstrates the effectiveness of the region-based features because they have strong correlation with real-world objects. Care must be taken, however, when computing the weights. For example, IRM's weight estimation is not very accurate. While it performs better than the wavelet texture feature, it is about the same as the color histogram approach.

• CRM is more accurate in similarity model and more robust to imperfect image segmentation than IRM. Due to using both 1st and 2nd order constraints for the probabilistic weight estimation, the computational complexity of CRM is greater than IRM. However the gain is that the CRM performs better than IRM in terms of retrieval precision.

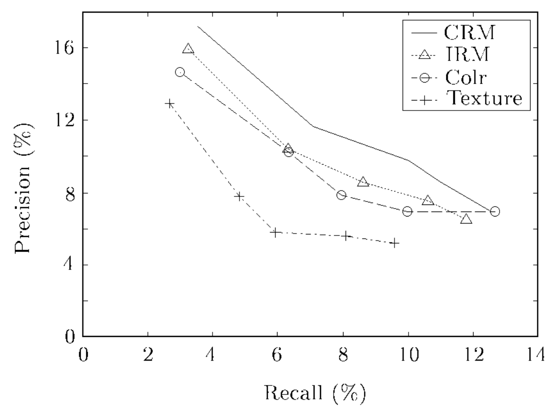


Fig.5. Retrieval performance $Pr(Re)$ comparison between CRM and IRM, color histogram, wavelet texture.

Fig.6. Comparisons between CRM and IRM. (1) Retrieving sunset images, where (a) uses CRM and (b) uses IRM. (2) Retrieving car images, where (c) users CRM and (d) uses IRM.

4.5 Relevance Feedback Enhanced CRM

Table 2 shows the comparison between relevance feedback enhanced CRM and plain CRM. When the top 20, 100, and 180, most similar images are returned, Fig.7 shows their $Pr(Re)$ curves at 0, 1, 2 feedback iterations. The examples of relevance feedback are shown in Fig.8. Based on the table and figures, the following observations can be made:

Table 2. Retrieval Performance Comparison Between Relevance Feedback Enhanced CRM and Plain CRM

Precision (%)	Return top 20	Return top 100	Return top 180
0 feedback	17.30	9.83	6.91
1 feedback	22.47	14.60	10.80
2 feedbacks	26.92	15.62	12.89

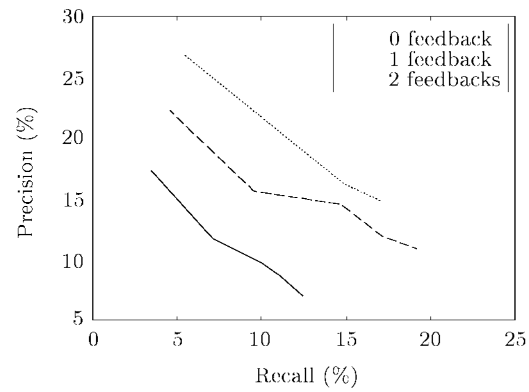


Fig.7. Retrieval performance $Pr(Re)$ comparison between relevance feedback enhanced CRM and plain CRM.

Fig.8. Relevance feedback examples with CRM. (a) With no feedback. (b) With one iteration of feedback. (c) With two iterations of feedback. Retrieving images with both mountain and sea. (d) With no feedback. (e) With one iteration of feedback. (f) With two iterations of feedback. Retrieving images with Santa Claus.

• With more feedback iterations, the retrieval performance improves because relevance feedback leverages user's knowledge to adapt similarity model better.

• The performance increase in the first iteration is the biggest. This is important because users prefer to have iterations as few as possible. It is clear that combining relevance feedback with CRM is a good direction.

5 Conclusions

In this paper, we proposed a novel constraint-based region matching approach to image retrieval. This approach simultaneously models both *first-order* constraints and *second-order* constraints in a principled probabilistic framework. Furthermore, the retrieval performance is improved by an adaptive filter based relevance feedback. Experimental results show that the CRM performs better than the state-of-the-art techniques.

Acknowledgement The first author wants to thank Prof. Wenping Wang of Dept. CSIS, Hong Kong University, for his helpful comments and discussions about the paper.

References

[1] Ru, Y, Huang T S, Chang S F. Image retrieval: Current techniques, promising directions and open issues. *Journal of Visual Communication and Image Representation*, 1999, 10: 39–62.

[2] Smeulders A W M, Worring M, Santini S. Content-based image retrieval at the end of the early years. *IEEE Trans. PAMI,* 2000, 22(12): 1349–1380.

[3] Niblack W *et al.* The QBIC project: Querying images by content using color, texture and shape. In *SPIE Proc, Storage and Retrieval for Image and Video Databases,* San Jose, 1993, 1908: 1723–187.

[4] Pentland A, Picard R, Sclaroff S. Photobook: Content-based manipulation of image databases. *Int. J. Comput. Vis.,* 1996, 18(3): 233–254, 1996.

[5] Smith J R, Chang S F. VisualSEEK: A fully automated content-based image query system. ACM Multimedia, Bostan MA, 1996.

[6] Rui Y, Huang T, Mehrotra S. Content-based image retrieval with relevance feedback in MARS. In *Proc. IEEE Int. Conf. on Image Processing,* 1997.

[7] Ma W Y, Manjunath B S. Netra: A toolbox for navigation large image database. In *Proc. of IEEE ICIP,* 1997, pp.568–571.

[8] Carson C, Belongie S, Greenspan, H, Malik J. Region-based image querying. In *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries,* 1997, pp.42–49.

[9] Li J, Wang J Z, Wiederhold G. IRM: Integrated region matching for image retrieval. In *Proceedings ACM Multimedia 2000,* Los Angeles, CA, 2000, pp.147–156, 2000.

[10] Rui Y, Huang T S. Optimizing learning in image retrieval. In *Proc. IEEE Int. Conf. on Computer Vision and Pattern Recognition* (*CVPR2000*), 2000, 1: 236–243.

[11] Tian Q, Wu Y, Huang T S. Combine user defined region-of-interest and spatial layout for image retrieval. In *IEEE International Conference on Image Processing,* Vancouver, BC, 2000, 3: 746–749.

[12] Gouet V, Boujemaa N. Object-based queries using color points of interest. In *Proc. of IEEE Workshop on Content-Based Access of Images and Videos* (*CBAIVL*)*,* 2001, Kauai, Hawaii.

[13] Tian Q, Sebe N, Lew M S, Loupias E, Huang T S. Image retrieval using wavelet-based salient points. In *Journal of Electronic Imaging,* Special Issue on Storage and Retrieval of Digital Media, 2001, 10(4): 835–849.

[14] Ma W Y, Manjunath B S. Netra: A toolbox for navigation large image database. In *Proc. of IEEE ICIP,* 1997, pp.568–571.

[15] Ko B C, Lee H S, Byun H. Region-based image retrieval system using efficient feature description. In *Proc. of 15th International Conference on Pattern Recognition,* 2000, 4: 283–286.

[16] Moghaddam B, Biermann H, Margaritis D. Image retrieval with local and spatial queries. In *Proc. IEEE ICIP 2000,* 2000, 2: 542–545.

[17] Wang W, Song Y, Zhang A. Semantics retrieval by content and context of image regions. In *Proc. the 15th International Conference on Vision Interface,* Calgary, Canada, May 27–29, 2002.

[18] Xu Y, Duygulu P, Saber E, Tekalp A M, Yarman-Vural F T. Object based image retrieval based on multi-level segmentation. In *Proc. IEEE ICASSP,* Istanbul, Turkey, 2000, 6: 2019–2022.

[19] Deng Y, Manjunath B S. An efficient low-dimensional color indexing scheme for region-based image retrieval. In *Proc. IEEE Intl. Conf. Acoustics, Speech, and Signal Processing* (*ICASSP*)*,* 1999, pp.3017–3020.

[20] Smith J R, Chang S F. SaFe: A general framework for integrated spatial and feature image search. *IEEE First Workshop on Multimedia Signal Processing,* 1997, pp.301–306.

[21] Pfefferkorn C. A heuristic problem solving design system for equipment or furniture layouts. *Communications of the ACM,* 1975, 18(5): 286–297.

[22] Kazuyoshi H, Fumio M. Constraint-based approach for automatic spatial layout planning. In *Proceedings of the 11th Conference on Artificial Intelligence for Applications,* Los Angeles, CA, USA, 1995, pp.38–45.

[23] Anil K. Jain, Fundamentals of Digital Image Processing. Prentice-Hall press, 1989, pp.412–413.

[24] Pavlidis T, Liow Y T. Integrating region growing and edge detection. *IEEE Trans Pattern Analysis and Machine Intelligence,* 1990, 12(3): 225–133.

[25] Felzenszwalb P F, Huttenlocher D P. Image segmentation using local variation. In *Proc. of IEEE Computer Vision and Pattern Recognition,* Santa Barbara, CA, USA, 1998, pp.98–104.

[26] Leu J G. Computing a shape moment from its boundary. *Pattern Recognition,* 1991, 10: 949–957.

[27] Gorman J W, Mitchell O R, Kuhl F P. Partial shape recognition using dynamic programming. *IEEE Trans. PAMI,* 1988, 10: 257–266.

[28] Cox I J, Miller M L, Minka T P, Papathomas T V, Yianilos P N. The bayesian image retrieval system, pichunter: theory, implementation, and psychophysical

experiments. *IEEE Trans. on Image Processing,* 2000, 9(3): 524–524.

[29] Zhou X, Hunag T S. Comparing discriminate transformations and SVM for learning during multimedia retrieval. *ACM Multimedia'2001,* 2001, pp.137–146.

[30] Anil J, Douglas Z. Feature selection: Evaluation, application, and small sample performance. *IEEE PAMI,* 1997, 19(2): 153–158.

[31] Pudil P, Novovicova J, Kittler J. Floating search methods in feature selection. *Pattern Recognition Letters,* 1994, 15: 1119–1125.

[32] Wang T, Rui Y, Hu S M. Optimal adaptive learning for image retrieval. In *Proc. of IEEE Computer Vision and Pattern Recognition* (*CVPR2001*), Kauai, Hawaii, December 11–13, pp.1140–1147, 2001.

[33] Rodden K, Basalaj W, Sinclair D, Wood K. Does organization by similarity assist image browsing. In *Proc. ACM Compute-Human Interaction* (*CHI*), 2001, pp.190–197.

[34] Salton G, McGill M J. Introduction to modern information retrieval. McGraw-Hill Book Company, New York, 1982.

Wang Tao is a Ph.D. candidate in Department of Computer Science and Technology at Tsinghua University. He received the B.S. degree in the University of Science and Technology of China in 1996, the M.S. degree from Chinese Academy of Sciences in 1999. His current research interests are in content-based image retrieval, pattern recognition, image processing and computer graphics.

Yong Rui received his Ph.D. cadidate in Electrical and Computer Engineering in 1999 from the University of Illinois at Urbana-Champaign. Since March 1999, he has been a researcher in Microsoft Research, Redmond. His research interests include multimedia systems, distance learning and distributed meetings, image/video/audio processing, computer vision and machine learning. He has published five book chapters, six journal papers, and over forty referred conference papers in the above areas. Dr. Rui holds six U. S. pending patents and is a member of ACM and IEEE.

Jia-Guang Sun is a professor in Department of Computer Science and Technology at Tsinghua University. He is also the Director of National CAD Engineering Center at Tsinghua University and Academician of Chinese Academy of Engineering. He received the B.S. degree in Computer Science from Tsinghua University in 1972. During 1985 to 1986, he was a visiting scholar in UCLA. His current research interests are in computer-aided geometric design, computer graphics and product data management.