International Journal of Computer Vision 56(1/2), 37–45, 2004

Constraint Based Region Matching for Image Retrieval

TAO WANG*
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, P.R. China
wangtao@cg.cs.tsinghua.edu.cn

YONG RUI
Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
yongrui@microsoft.com

JIA-GUANG SUN
Department of Computer Science and Technology, Tsinghua University, Beijing 100084, P.R. China
sunjg@mail.tsinghua.edu.cn

Received March 28, 2002; Revised April 11, 2003; Accepted April 24, 2003

Abstract. Objects and their spatial relationships are important features for human visual perception. In most existing content-based image retrieval systems, however, only global features extracted from the whole image are used. While they are easy to implement, they have limited power to model semantic-level objects and spatial relationship. To overcome this difficulty, this paper proposes a constraint-based region matching approach to image retrieval. Unlike existing region-based approaches where either individual regions are used or only first-order constraints are modeled, the proposed approach formulates the problem in a probabilistic framework and simultaneously models both first-order region properties and second-order spatial relationships for all the regions in the image. Specifically, in this paper we present a complete system that includes image segmentation, local feature extraction, first- and second-order constraints, and probabilistic region weight estimation. Extensive experiments have been carried out on a large heterogeneous image collection with 17,000 images. The proposed approach achieves significantly better performance than the state-of-the-art approaches.

Keywords: content-based image retrieval, region matching, probabilistic weight estimation, similarity model

1. Introduction

With the advances in both image analysis techniques and cheap digital storage, content-based image retrieval (CBIR) has become an active research area during the past decade. There exist rich literature on CBIR, including QBIC (Niblack et al., 1993), PhotoBook (Pentland et al., 1996), VisualSEEK (Smith and Chang, 1996), MARS (Rui et al., 1997), Netra (Ma and Manjunath, 1997), BlobWorld (Carson et al., 1997) and SIMPLIcity (Li et al., 2000) etc. For recent surveys on various CBIR systems and techniques please refer to Rui et al. (1999) and Smeulders et al. (2000). Despite years of extensive research, however, assisting users to find their desired images accurately and quickly is still an open problem. According to recent study results (Rodden et al., 2001), one of the main challenges is the semantic gap between users' high-level query concepts (e.g. an apple on a table) and low-level features which the

*Current Address: Statistical Computing Group, Intel China Research Center, # 06-01 Beijing Kerry Center, 1 GuangHua Road, Chaoyang District, Beijing 100020, P.R. China. E-mail: tao.wang@intel.com computer can extract (e.g. 30% red color and a bold with vertical edges).

While closing the semantic gap is still a far cry given today's technology, there are some ways to narrow it down. One is to introduce relevance feedback learning and take advantage of users' knowledge to guide the search. Many techniques have been developed along that line (e.g., Rui and Huang, 2000; Cox et al., 2000; Wang et al., 2001). The other is to improve the similarity models and features used in a CBIR system. In most existing systems, global features such as global color histogram, texture and shape are broadly used (e.g., Niblack, 1993; Smith and Chang, 1996; Rui and Huang, 2000). 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 correlations with objects, which are prospective to provide a big step towards the semantic-based retrieval.

Existing local-feature-based approaches fall into three categories: *fixed-layout-based* (e.g., Tian et al., 2000), *salient-point-based* (Tian et al., 2001; Gouet and Boujemaa, 2001) and *region-based* approaches (Carson et al., 1997; Kam et al., 2000; Moghaddam et al., 2000; Ko et al., 2000; Wang et al., 2002). Among various local-feature-based approaches, the region-based approach so far has been the winning approach since they have strong correlations with real-world objects. This approach first segments an image into multiple regions that have high correlations with real-world objects. Then the total similarity between two images is calculated based on all the corresponding regions. However, because of imperfect image segmentation (e.g., over or under segmentation), care must be taken when comparing two images. The Netra system (Deng and Manjunath, 1999; Ma and Manjunath, 1997) compares images based on individual regions. Although queries based on multiple regions are allowed, the retrieval is done by merging individual single-region query results. This approach is therefore less robust to imperfect segmentation. The SaFe system (Smith and Chang, 1997), on the other hand, uses a 2D-string approach. While this system is more robust than Netra, it is sensitive to region shifting and rotation. Xu et al. (2000) takes a different approach which represents visual objects (e.g., cars) as composite nodes in a hierarchical tree scheme. Compared with previous approaches, this approach is more robust, but less generalizable to other domains. A more recent technique, integrated region matching (IRM), is developed by Li et al. (2000). This approach reduces the influence of inaccurate segmentation by allowing a region to be matched to multiple regions from another image. However, it suffers from a greedy algorithm when estimating inter-region weights, and does not take into account the spatial relationship between regions in an image.

In this paper, based on spatial constrains between image regions, a novel probabilistic region matching technique is presented. Different from other approaches, we take into account not only the first-order (e.g., region features) constraints but also the second-order ones (e.g., spatial relationship between them). More importantly, unlike other *ad hoc* approaches, our similarity model between two images is based on a principled probabilistic framework. The proposed approach achieves significantly better retrieval performance than the existing good approaches.

The rest of the paper is organized as follows. For related work in Section 2, we focus on the integrated region matching (IRM) approach (Li et al., 2000). It is one of the best approaches available and the one that we will compare against in this paper. In Section 3, we give detailed descriptions of our proposed constraint-based region matching (CRM) approach. 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. Related Work

After image segmentation, the overall similarity between two images can be calculated based on all the corresponding regions. 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. Let the similarity between regions r_i and r'_j be $S(r_i, r'_j)$, where r_i is the ith region in R_1 and r'_j is the jth region in R_2. Then the total similarity between two images can therefore be defined as the similarity $S(R_1, R_2)$ between the two region sets:

$$
\begin{aligned}
&S(R_1, R_2) = \sum_{i=1}^{M} \sum_{j=1}^{N} w_{ij} S(r_i, r'_j) \\
&\text{s.t.} \ \sum_{i=1}^{M} \sum_{j=1}^{N} w_{ij} = 1
\end{aligned}
\tag{1}
$$

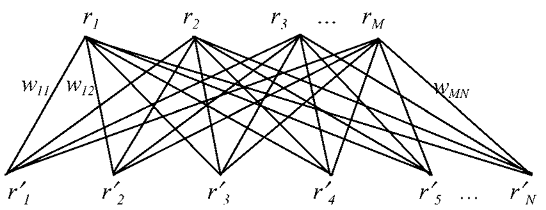

Figure 1. Region matching between region sets R_1, R_2.

where weight w_{ij} indicates the importance of region pair r_i and r'_j with respect to the overall similarity. This region-matching scheme is illustrated in Fig. 1. Every vertex in the graph corresponds to a region in an image. If two regions r_i and r'_j are matched, the two vertices are connected by an edge with a weight w_{ij}. It becomes clear that estimating w_{ij} is a key step for the similarity model.

Among various existing region-based approaches discussed in Section 1, IRM (Li et al., 2000) is one of the best. It uses the "most similar highest priority" (MSHP) principle to estimate w_{ij}, i.e., iteratively gives the largest priority to the region pair having minimum distance. Let the normalized importance of r_i and r'_j be p_i and p'_j, the following constraints hold:

$$\sum_{i=1}^{M} p_i = \sum_{j=1}^{N} p'_j = 1 \tag{2}$$

$$\sum_{i=1}^{M} w_{ij} = p'_j, \quad \sum_{j=1}^{N} w_{ij} = p_i \tag{3}$$

The IRM approach can be summarized as follows:

1. Initially, set the assigned region set $L = \{\ \}$. Let the unassigned region set $E = \{(i, j) \mid i = 1, 2, \ldots, M; j = 1, 2, \ldots, N\}$.
2. Choose the minimum d_{ij} for $(i, j) \in E\text{-}L$. Label the corresponding (i, j) as (i', j')
3. $\min(p_{i'}, p'_{j'}) \to w_{i'j'}$
4. If $p_{i'} < p'_{j'}$, set $w_{i'j} = 0, j \neq j'$; otherwise set $w_{ij'} = 0, i \neq i'$.
5. $p_{i'} - \min(p_{i'}, p'_{j'}) \to p_{i'}$
6. $p'_{j'} - \min(p_{i'}, p'_{j'}) \to p'_{j'}$
7. $L + \{(i', j')\} \to L$
8. If $\sum_{i=1}^{M} p_i > 0$ and $\sum_{j=1}^{N} p'_j > 0$, then go to Step 2; otherwise stop.

In IRM, p_i, p'_j are respectively defined as the area percentage of region r_i in R_1, region r'_j in R_2 and d_{ij} is the distance between region r_i and r'_j. While IRM makes a significant step in the integrated region matching, its accuracy and robustness suffers from the fact that it does not take into account the spatial relationships between regions for the weight estimation. In addition, it uses a greedy algorithm to estimate inter-region weight w_{ij}, which again reduces the robustness of the similarity model. For example, the greedy algorithm can easily be trapped into local minimum, and all the subsequent weight estimation becomes un-reliable (see Section 3.5).

3. Constraint Based Region Matching

To improve the robustness of region matching, we propose a new technique that depends on not only the region properties but also the spatial relationships between regions. More importantly, it formulates the weight estimation problem in a probabilistic framework and solves the problem in a principled way. In the rest of this section, we present detailed description of CRM approach.

3.1. Image Segmentation

Image segmentation tries to divide an image into regions that have strong correlations with real-world objects. We use a region-growing approach to segment images based on HSV color model (Jain, 1989). An example of segmentation result is shown in Fig. 2.

More sophisticated segmentation techniques can be used (e.g., Pavlidis and Liow, 1990; Felzenszwalb and Huttenlocher, 1998). But that is beyond the discussing scope of this paper.

3.2. Region Features

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

Figure 2. A landscape image (a) and its segmentation (b).

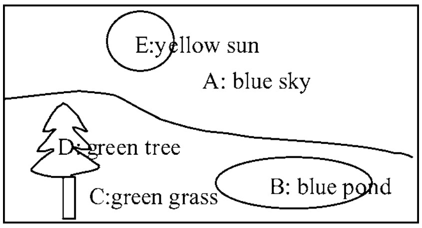

Figure 3. Spatial relationship between objects.

the content of an image. In our implementation, the color feature $\vec{C}$ is the mean color in HSV color model, the size feature ρ is the area percentage, the shape feature is the eccentricity e (Leu, 1991) and the position feature $\vec{O}$ is the normalized region center.

3.3. Regional and Spatial Constraints

Object and relationship modeling has been extensively studied for the past three decades in spatial layout planning (Pfefferkorn, 1975; Kazuyoshi and Fumio, 1995). Pfefferkorn defines three important components (Kazuyoshi and Fumio, 1995): (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 about the individual objects, we call them *first-order* constraints. Similarly, we call the spatial relationship constraints *second-order* constraints, as it involves two objects. Even though the above concepts were first developed in spatial layout planning, they apply equally well in image content modeling. For example, in Fig. 3, the image is characterized with objects A, B, C, D and their spatial relationship (see Table 1). If another image contains both similar regions and similar spatial relationships, then the two images should be similar in content.

Table 1. Spatial relationship of objects in Fig. 3.

	A	B	C	D	E
A	N/A	Top	top	top	outside
B	bottom	N/A	inside	right	bottom
C	bottom	outside	N/A	outside	bottom
D	bottom	left	inside	N/A	bottom
E	inside	top	top	top	N/A

Let r_i, r_k be regions in Image 1 and r'_j, r'_l are matched regions to them respectively in Image 2. The constraints and their associated similarities are defined as follows:

1. *Region property constraint*
   - *Color*:

$$S_c(r_i, r'_j) = \exp\left(-\|\vec{C}_i - \vec{C}'_j\|^2 \Big/ 2\sigma_c^2\right) \quad (4)$$

   - *Shape*:

$$S_e(r_i, r'_j) = \exp\left(-\|e_i - e_j\|^2 \Big/ 2\sigma_e^2\right) \quad (5)$$

2. *Region position constraint*
   - *Position*:

$$S_p(r_i, r'_j) = \exp\left(-\|\vec{O}_i - \vec{O}'_j\|^2 \Big/ 2\sigma_p^2\right) \quad (6)$$

3. *Spatial relationship constraint*
   - *Orientation*:

$$S_o(r_i, r_k, r'_j, r'_l) = \left(\frac{(\vec{O}_i - \vec{O}_k)\cdot(\vec{O}'_j - \vec{O}'_l)}{\|\vec{O}_i - \vec{O}_k\|\cdot\|\vec{O}'_j - \vec{O}'_l\|} + 1\right)\Big/ 2 \quad (7)$$

   - *Inside/outside*:

$$S_i(r_i, r_k, r'_j, r'_l) = \overline{(r_i \text{ in/out } r_k)\,\mathrm{XOR}(r'_j \text{ in/out } r'_l)} \quad (8)$$

   - *Size Ratio*:

$$S_s(r_i, r_k, r'_j, r'_l) = \left(\frac{(\rho_i, \rho_k)\cdot(\rho'_j, \rho'_l)}{\|(\rho_i, \rho_k)\|\cdot\|(\rho'_j, \rho'_l)\|} + 1\right)\Big/ 2 \quad (9)$$

where σ_c, σ_e and σ_p control the penalty of different variation respectively. The total similarity based on the *first-order* constraints, e.g., Eqs. (4)–(6), for two regions r_i, r'_j is defined as:

$$S_1(r_i, r'_j) = w_c S_c(r_i, r'_j) + w_e S_e(r_i, r'_j) + w_p S_p(r_i, r'_j) \quad (10)$$
$$\text{s.t.} \quad w_c + w_e + w_p = 1$$

Similarly, the total similarity based on the *second-order* constraints, e.g., Eqs. (7)–(9), is

$$\begin{aligned} S_2(r_i, r_k, r'_j, r'_l) &= w_o S_o(r_i, r_k, r'_j, r'_l) \\ &\quad + w_i S_i(r_i, r_k, r'_j, r'_l) \\ &\quad + w_s S_s(r_i, r_k, r'_j, r'_l) \qquad (11) \\ \text{s.t.} \quad & w_o + w_i + w_s = 1 \end{aligned}$$

where $w_{c,} w_{e,} w_{p,} w_{o,} w_{i,}$ and w_s are proper weights for corresponding constraints. These weights can either be specified through experiments, or be dynamically adjusted via relevance feedback (e.g., Wang et al., 2001). 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. In our experiments, the constraints of Eqs. (4), (5), (8) and (9) are translation, rotation and scaling invariant. Equation (7) is translation and scaling invariant and Eq. (6) is only scaling invariant.

3.4. *Probabilistic Weight Estimation*

The overall similarity between two region sets, thus two images, is defined as follows:

$$\begin{aligned} S(R_1, R_2) &= \sum_{i=1}^{M} \sum_{j=1}^{N} w_{ij} S_1(r_i, r'_j) \qquad (12) \\ \text{s.t.} \quad & \sum_{i=1}^{M} \sum_{j=1}^{N} w_{ij} = 1 \end{aligned}$$

As discussed in Section 2, there are several problems with how IRM estimates the weights. To address these issues, we cast the weight estimation problem into a probabilistic framework and solve the problem in a principled way.

Note that the similarity between two entities can be interpreted as the probability of the two entities being similar. All the similarities defined in Eqs. (4)–(11) are in the range of [0, 1], and can readily be used to estimate the probabilities. Let $x \sim y$ and $x \approx y$ denote that x matches y based on the *first-order* (region feature) and *second-order* (spatial relationship) constraints respectively. We therefore have:

$$P(r_i \sim r'_j) \doteq S_1(r_i, r'_j) \qquad (13)$$

$$\begin{aligned} & P(r_i \approx r'_j, r_k \approx r'_l \mid r_i \sim r'_j, r_k \sim r'_l) \\ & \quad \doteq S_2(r_i, r_k, r'_j, r'_l) \qquad (14) \end{aligned}$$

where $\doteq$ stands for "can be estimated by". That is, the probability that r_i matches r'_j in terms of the first-order constraints can be estimated by $S_1(r_i, r'_j)$. Similarly, the probability that r_i and r_k match r'_j and r'_l in terms of the second-order constraints, given r_i matches r'_j and r_k matches r'_l in term of the first-order constraints, can be estimated by $S_2(r_i, r_k, r'_j, r'_l)$.

It is intuitive that a region pair (r_i, r'_j) should receive higher weight w_{ij} in the similarity model if they are a better matching pair. In Eq. (11), 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), $P(r_i \sim r'_j, r_i \approx r'_j) = \sum_{k=1}^{M} \sum_{l=1}^{N} P(r_i \sim r'_j, r_k \sim r'_l, r_i \approx r'_j, r_k \approx r'_l)$.

The $P(r_i \sim r'_j, r_i \approx r'_j)$ is the probability that 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 the Eqs. (13) and (14), we have:

$$\begin{aligned} W_{ij} &\doteq P(r_i \sim r'_j, r_i \approx r'_j) \\ &= \sum_{k=1}^{M} \sum_{l=1}^{N} P(r_i \sim r'_j, r_k \sim r'_l, r_i \approx r'_j, r_k \approx r'_l) \\ &= \sum_{k=1}^{M} \sum_{l=1}^{N} P(r_i \approx r'_j, r_k \approx r'_l \mid r_i \sim r'_j, r_k \sim r'_l) \\ &\quad P(r_i \sim r'_j, r_k \sim r'_l) \\ &\doteq \sum_{k=1}^{M} \sum_{l=1}^{N} S_2(r_i, r_k, r'_j, r'_l) P(r_i \sim r'_j) P(r_k \sim r'_l) \\ &\doteq \sum_{k=1}^{M} \sum_{l=1}^{N} S_2(r_i, r_k, r'_j, r'_l) S_1(r_i, r'_j) S_1(r_k, r'_l) \qquad (15) \end{aligned}$$

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

$$w_{ij} = q_i * \frac{W_{ij}}{\sum_{j=1}^{N} W_{ij}} \qquad \text{s.t.} \quad \sum_{i=1}^{M} q_i = 1 \tag{16}$$

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

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.

The complete CRM approach is summarized as follows:

Input: Image 1 and Image 2
Output: similarity between Image 1 and Image 2

(1) Obtain region sets $R_1 = \{r_1, r_2, \ldots, r_M\}$ for Image 1 and $R_2 = \{r'_1, r'_2, \ldots, r'_N\}$ for Image 2 from the image segmentation module.
(2) Extract region features of each region r_i, r'_j.
(3) Compute the 1st order constraint $S_1(r_i, r'_j)$ and the 2nd order constraint $S_2(r_i, r_k, r'_j, r'_l)$ via Eqs. (4)–(11).
(4) Calculate the probability weights w_{ij} based on both 1st order and 2nd order constraints using Eqs. (15) and (16).
(5) Calculate the total similarity $S(R_1, R_2)$ between two region sets R_1, R_2 using Eq. (12).

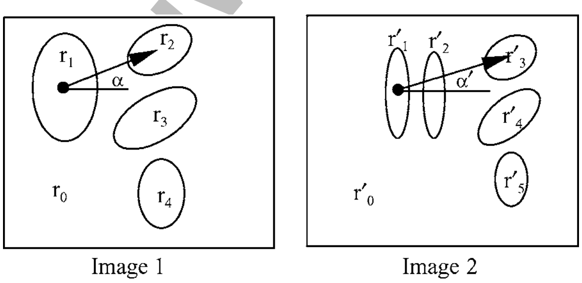

Figure 4. Two images with segmented region sets R_1, R_2.

Table 2. Estimated weights using CRM.

	r'_0	r'_1	r'_2	r'_3	r'_4	r'_5
r_0	0.171	0.000	0.000	0.000	0.000	0.000
r_1	0.000	0.087	0.060	0.025	0.024	0.014
r_2	0.000	0.035	0.040	0.070	0.037	0.012
r_3	0.000	0.028	0.042	0.034	0.056	0.044
r_4	0.000	0.011	0.022	0.017	0.053	0.106

3.5. *Illustrations and Simulations*

In this subsection, we illustrate the effectiveness of the proposed CRM approach by using two synthetic images. Shown in Fig. 4, Image 2 is an over-segmented version of Image 1, shifted 25% to the right and scaled 25% down. To highlight CRM's probabilistic weight estimation framework and the second-order constraints, we assign the same red color to all the regions in the two images. Let r_0 and r'_0 represent the white background regions in the two images, Tables 2 and 3 show the estimated weights, based on CRM and IRM respectively. The following observations can be made:

- Because CRM takes into account the second-order constraints, it is much more robust to translation and scaling changes. For example, in CRM, r_1 is well matched to both r'_1 and r'_2. But in IRM, r_1 is not matched to r'_2 at all.
- Because CRM estimates the weights using a probabilistic framework, it is much less likely to be trapped into a local minimum. For example, in CRM, even though other regions' pose distractions, r_3 can still match well with r'_4. However, in IRM, r_3 is matched to r'_2 instead of r'_4 because of the local minimum created by the background regions.

Similar results are obtained when we assign different colors to different corresponding regions, e.g., assign red to regions r_1, r'_1, and r'_2, and assign yellow to

Table 3. Estimated weights using IRM.

	r'_0	r'_1	r'_2	r'_3	r'_4	r'_5
r_0	0.790	0.000	0.000	0.000	0.000	0.000
r_1	0.043	0.027	0.000	0.000	0.015	0.000
r_2	0.014	0.000	0.000	0.014	0.000	0.000
r_3	0.000	0.000	0.032	0.000	0.019	0.000
r_4	0.000	0.000	0.000	0.012	0.000	0.024

regions r_3 and r'_4. The above simulations clearly demonstrate the advantage of CRM over IRM due to CRM's principled probability weights.

4. Experiments

In the previous section, we have shown the advantages of CRM by both theory and simulation. In this section, we will use large-scale real-world images to examine and compare the retrieval performance of CRM against existing approaches.

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 due to 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 professional-annotated by Corel professionals. The whole images have been classified into 170 categories and there are 100 images in each category.

The Corel data sets have been used in other systems in which relatively high retrieval performance has been reported (Carson, 1997; Cox, 2000; Li, 2000; Rui, 2000; Tian, 2001 etc.). However, those systems only use pre-selected categories with distinctive visual characteristics (e.g., red cars vs. green mountains). In our experiments, no pre-selection is made in 17,000 images. Since average users want to retrieve images based on high-level concepts, not low-level visual features (Rodden, 2000), 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 techniques, we randomly generated 400 queries in the whole 17,000 images for each retrieval condition. For all the experiments reported in this section, they are the average of all the 400 query results.

4.3. Performance Measures

The most widely used performance measures for information retrieval are precision (*Pr*) and recall (*Re*) (Salton and McGill, 1982). *Pr*(*Sc*) is defined as the number of retrieved relevant objects over the number of total retrieved objects, say the top 20 images. *Re*(*Sc*) is defined as the number of retrieved relevant objects over the total number of relevant objects in the image collection (in the case of Corel data set, 99). 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 used to better characterize the performance of a retrieval system.

4.4. CRM vs. Existing Approaches

Tested on a PC with PIII 1G CPU and 256 M memory, the average time for one query is 1.66 second for IRM and 3.61 second for CRM over 17,000 images. Table 4 compares the image retrieval results using CRM, IRM (Li, 2000) and the global features, e.g. color histogram (Swain, 1991) and wavelet texture (Rui, 1997). When the top 20, 100, and 180, most similar images are returned. In IRM, we use same color, shape and position features as CRM. For the wavelet-based texture, the original image is decomposed into the third level, resulting 10 de-correlated 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.

To better compare these approaches, we also plot their *Pr*(*Re*) curves in Fig. 5. Based on the Table 4 and Fig 5, the following observations can be made:

Table 4. Comparison CRM with IRM, color histogram and wavelet texture.

Precision (percentage)	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

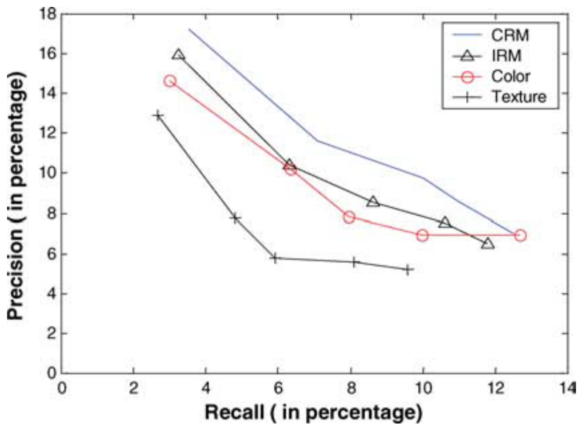

Figure 5. Retrieval performance $Pr(Re)$ comparison.

- 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 correlations 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 the 1st and 2nd order constraints for the probabilistic weight estimation, the computational complexity of CRM is bigger than IRM. However the gain is that the CRM performs 10% better than IRM in terms of retrieval precision.

5. Conclusion and Future Direction

In this paper, we have proposed a novel constraint-based region matching approach to image retrieval. This approach is based on a principled probabilistic framework and models both first-order region properties and second-order spatial relationships. Simulations and real-world experiments show that its retrieval performance is better than a state-of-the-art technique, i.e., IRM.

As a general future direction, we believe local-feature-based approach, e.g., IRM and CRM, is a significant step towards narrowing down the semantic gap. For our future work, we are integrating relevance feedback into CRM to provide users with a robust and adaptive retrieval experience.

Acknowledgment

The first author wants to thank Prof. Wenping Wang of Hongkong University for his helpful comments and discussions about the paper. The first author is supported in part by China NSF grant 69902004 and the National 973 Program (No. 1998030600). The Corel image set was obtained from Corel and used in accordance with their copyright statement.

References

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

Cox, I.J., Miller, M.L., Minka, T.P., Papathomas, T.V., and Yianilos, P.N. 2000. The Bayesian image retrieval system, Pichunter: Theory, implementation, and psychophysical experiments. *IEEE Trans. on Image Processing*, 9(3):524–524.

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

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

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

Jain, A.K. 1989. *Fundamentals of Digital Image Processing*. Prentice-Hall Press, pp. 412–413.

Kam, A.H., Ng, T.T., Kingsbury, N.G., and Fitzgerald, W.J. 2000. Content based image retrieval through object extraction and querying, 2000. In *Proc. IEEE Workshop on Content-based Access of Image and Video Libraries (CBAIVL)*, pp. 91–95.

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

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

Leu, J.-G. 1991. Computing a shape moments from its boundary, *Pattern Recognition*, 10:949–957.

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

Ma, W.Y. and Manjunath, B.S. 1997. Netra: A Toolbox for navigation Large Image Database. In *Proc. of IEEE ICIP*, pp. 568–571.

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

Niblack, W. et al. 1993. 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, vol. 1908, pp. 173–187.

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

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

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

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

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

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

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

Salton, G. and McGill, M. J. 1982. *Introduction to Modern Information Retrieval*. McGraw-Hill Book Company: New York.

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

Smith, J.R. and Chang, S.F. 1996. VisualSEEK: A Fully Automated Content-Based Image Query System. *ACM Multimedia 1996*, Bostan MA.

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

Swain, M.J. and Ballard, D.H. 1991. Color indexing. *International Journal of Computer Vision*, 7(1):11–32.

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

Tian, Q., Sebe, N., Lew, M.S., Loupias, E., and Huang, T.S. 2001. Image Retrieval Using Wavelet-Based Salient Points. *Journal of Electronic Imaging, Special Issue on Storage and Retrieval of Digital Media*, 10(4):835–849.

Wang, T., Rui Y., and Hu, S.M. 2001. Optimal Adaptive Learning for Image Retrieval. In *Proc. of IEEE Computer Vision and Pattern Recognition (CVPR2001)*, Kauai, Hawaii, pp. 1140–1147.

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

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