REGION CORRESPONDENCE FOR IMAGE RETRIEVAL USING GRAPH-THEORETIC APPROACH AND MAXIMUM LIKELIHOOD ESTIMATION

*Chuech-Yu Li and Chiou-Ting Hsu**

Department of Computer Science, National Tsing Hua University, Taiwan
Email: cthsu@cs.nthu.edu.tw, Phone: +886-3-5742960, Fax: +886-3-5723694

ABSTRACT

This paper proposes employing an efficient graph-theoretic approach to estimate the region correspondence between two images. We represent each image as an attributed graph and transform the image matching problem into a graph matching problem. During the image retrieval process, we formulate the matching problem as a maximum likelihood estimation and propose an optimization technique to derive its closed-form solution. Hence, we are capable to measure the image distance in terms of both the estimated region correspondence and the low-level features. This paper has two main contributions. First, our proposed matching technique is efficient and applicable to the interactive process of image retrieval. Second, with the estimated region correspondence, the proposed matching criterion, which is defined in terms of matched regions and penalized with unmatched regions, achieve satisfactory performance for retrieval application.

1. INTRODUCTION

Region-based approaches [1-7] have been one of the important research issues in content-based image retrieval. Representing images in region level captures not only the regions' local variations but also their spatial organizations. Since image distance is often defined as a combination of region distances, estimation of region correspondence becomes a prerequisite for a region-based image matching problem. Estimation of region correspondence for the application of interactive image retrieval is expected to have the following properties. First, both the region attributes (i.e. low-level features) and the adjacent relationship should be involved into the estimation. Second, the estimation should have as few manual control parameters as possible. Third, the estimation should be efficient and feasible to interactive process. Last, the estimated region correspondence should be easily incorporated into the subsequent relevance feedback steps.

IRM [1] and EMD flow [2] employ linear programming approaches to find out region correspondence in terms of region attributes. Ko et al. [3] use region centroids as one of the region attributes and apply Hausdorff distance to measure the distance between two sets of regions. However, these works [1-3] takes no account of adjacent relationship between regions into their estimation. Consider the case when an object is segmented into multiple regions; employing region adjacency will greatly improve the accuracy of region correspondence.

Hsieh et al. [4] integrate several techniques to construct templates for representing regions' spatial organization. The major limitation of this work comes from their heuristic tuning of control parameters. When dealing with various types of databases, the retrieval performance will heavily rely on such tuning parameters.

Graph representations are also widely used in region correspondence estimation. To incorporate both region attributes and adjacent relationship into estimation, image is usually represented as an attributed graph. Hence, the image matching problem is transformed into an attributed graph matching problem. Huet et al. [6] proposed a MAP graph matching technique to estimate the region correspondence. Our previous work [7] extends [8] and employs the EM algorithm to solve the inexact graph matching problem. However, the estimation of region correspondence in [6-7] is an iterative procedure and computational demanding. When performing query on a large database using the inefficient matching algorithm, the execution time would become unacceptable.

Baeza-Yates et al. [5] also represent images as attributed graphs and adopt graph edit distance to calculate the image distance. The distance is measured by the cost of transforming from one graph to the other. However, this work produces no explicit correspondence result for further application in the relevance feedback steps.

The objective of this paper is twofold. First, we aim to employ a graph-theoretic approach to estimate the region correspondence with the above-mentioned four properties. Second, we aim to define an image distance measurement based on the estimated correspondence as well as the region attributes. Since a database image may also contain unmatched regions, these regions should be incorporated

0-7803-8554-3/04/$20.00 ©2004 IEEE.

into the matching criterion and used to penalize the overall distance.

The rest of this paper is organized as follows. Section 2 elaborates the proposed graph-theoretic image matching approach. Section 3 defines the image distance in terms of the estimated region correspondence as well as the region attributes. Several experiments and comparisons are shown in section 5. Finally, Section 6 summarizes our work.

2. GRAPH-THEORETIC IMAGE MATCHING

In our approach, we represent an image as an attributed and undirected graph $G=(V,E)$. Each node $x_a \in V$ corresponds to a region in the image and the node's attribute corresponds to the region's low-level features. An edge $(x_a, x_b) \in E$ exists if the two corresponding regions are spatially adjacent. In image retrieval, the query image is denoted by the data graph $G_D=(V_D,E_D)$, while a database image is denoted by a model graph $G_M=(V_M,E_M)$. We use the two sets $\{a,b,\ldots\}$ and $\{\alpha,\beta,\ldots\}$ to index the regions in the data graph G_D and the model graph G_M respectively, and use adjacency matrices to represent the spatial organization of regions within each image. The data graph's adjacency matrix $\mathbf{D}$ with dimension $|V_D|\times|V_D|$ is defined as

$$D_{ab}=\begin{cases}1, \text{if } (x_a,x_b)\in E_D; \\ 0, \text{otherwise.}\end{cases} \quad (1)$$

Similarly, the model graph's adjacency matrix $\mathbf{M}$ with dimension $|V_M|\times|V_M|$ is defined as

$$M_{\alpha\beta}=\begin{cases}1, \text{if } (y_\alpha,y_\beta)\in E_M; \\ 0, \text{otherwise.}\end{cases} \quad (2)$$

The purpose of our image matching is to find out the best matching matrix to represent the region correspondence between the data graph (query image) and the model graph (database image). A matching matrix $\mathbf{S}$ with dimension $|V_D|\times|V_M|$ is defined as

$$S_{a\alpha}=\begin{cases}1, \text{if } x_a \text{ matches } y_a; \\ 0, \text{otherwise.}\end{cases} \quad (3)$$

From (3), we see that each node in the data graph matches only one region in the model graph, while a region in the model graph could be matched by more than one region in the data graph. Hence, the model graph may contain several unmatched regions, which indicate the database image's irrelevant regions to the query image.

Given the data graph G_D and the model graph G_M, we have the best matching matrix $\hat{\mathbf{S}}$ when the maximum a posteriori probability $P(\mathbf{S}\mid G_D,G_M)$ is achieved, i.e.

$$\begin{aligned}\hat{\mathbf{S}} &= \arg\max_{\mathbf{S}} P(\mathbf{S}\mid G_D,G_M) \\ &= \arg\max_{\mathbf{S}} P(G_D,G_M\mid\mathbf{S})P(\mathbf{S}).\end{aligned} \quad (4)$$

In (4), the second equality holds because of the Bayes' formula. If we further assume that all possible matching matrices have equal possibility, we obtain the following maximum likelihood formulation:

$$\hat{\mathbf{S}} = \arg\max_{\mathbf{S}} P(G_D,G_M\mid\mathbf{S}). \quad (5)$$

To simplify (5) into a more tractable form, we assume that given a matching matrix, all possible correspondence between nodes of G_D and G_M are conditionally independent. Then we have

$$P(G_D,G_M\mid\mathbf{S})=\prod_{a\in V_D}\prod_{\alpha\in V_M} p(x_a,y_\alpha\mid\mathbf{S}). \quad (6)$$

In (6), the matching probability $p(x_a,y_\alpha\mid\mathbf{S})$ depends on the similarity between the region attributes of x_a and y_α. Hence we define $p(x_a,y_\alpha\mid\mathbf{S})$ in terms of the region distance $d(x_a,y_\alpha)$:

$$p(x_a,y_\alpha\mid\mathbf{S})=\exp\{-w_{a\alpha}d(x_a,y_\alpha)\}. \quad (7)$$

We will detail the definition of the region distance $d(x_a,y_\alpha)$ in the next section. Note that, in (7), we introduce a new variable $w_{a\alpha}$ to weight the region distance. Given a matching matrix $\mathbf{S}$, we propose using the term $w_{a\alpha}$ to measure the neighboring dissimilarity of x_a and y_α and thus use this term to penalize the region distance $d(x_a,y_\alpha)$. We define $w_{a\alpha}$ as follows:

$$w_{a\alpha}=\sum_{b\in V_D}\sum_{\beta\in V_M} D_{ab}M_{\alpha\beta}S_{b\beta}d(x_b,y_\beta). \quad (8)$$

In (8), the multiplication of D_{ab}, $M_{\alpha\beta}$ and $S_{b\beta}$ accumulates the distance $d(x_b,y_\beta)$ for all the matched nodes between x_a's and y_α's neighboring nodes.

In addition, since perfect segmentations rarely happen in nature images, we can further refine (8) by including the similarity between neighboring nodes. If x_a and x_b are two regions belonging to the same object in the image, then the weight $w_{a\alpha}$ of x_a would be highly dependent on the node x_b. Otherwise, if the two nodes x_a and x_b correspond to different objects, then x_b will have less influence on $w_{a\alpha}$. Hence when the edges (x_a,x_b) and (y_α,y_β) exist, we set $D_{ab}=u_{ab}$ and $M_{\alpha\beta}=v_{\alpha\beta}$, where u_{ab} is a function of the similarity between x_a and x_b, and $v_{\alpha\beta}$ is a function of the similarity between y_α and y_β.

Substituting (7) and (8) into (6) results in the final likelihood function. To simplify the expression, we define 2 matrices $\mathbf{F}$ and $\mathbf{H}$, where $F_{a\alpha}=d(x_a,y_\alpha)$ and

$H_{b\beta} = S_{b\beta} d(x_b, y_\beta)$. The log-likelihood function then becomes

$$
\begin{aligned}
\log P(G_D, G_M \mid \mathbf{S}) &= -\sum_{a\in V_D}\sum_{\alpha\in V_M}\sum_{b\in V_D}\sum_{\beta\in V_M} D_{ab} M_{\alpha\beta} F_{a\alpha} H_{b\beta} \\
&= -\sum_{a\in V_D}\sum_{\alpha\in V_M}\sum_{b\in V_D}\sum_{\beta\in V_M} F_{a\alpha} M_{\alpha\beta} H^T_{\beta b} D^T_{ba} \\
&= \mathrm{Tr}\left[-\mathbf{FMH}^T\mathbf{D}^T\right] \\
&= \mathrm{Tr}\left[-\mathbf{D}^T\mathbf{FMH}^T\right] \\
&= \mathrm{Tr}\left[\mathbf{KH}^T\right],
\end{aligned} \tag{9}
$$

where $\mathbf{K} = -\mathbf{D}^T\mathbf{FM}$.

Our goal is to obtain the optimal $\mathbf{H}$ by maximizing (10). Let $\mathbf{P}\Sigma\mathbf{Q}^T$ be the singular value decomposition of $\mathbf{K}$, and let Δ be a $|V_D|\times|V_M|$ matrix with $\Delta_{ii}=1$ for all i and $\Delta_{ij}=0$ for all $i \neq j$. From the extremum principle [8], the maximum value of $\mathrm{Tr}\left[\mathbf{KH}^T\right]$ occurs when $\mathbf{H} = \mathbf{P}\Delta\mathbf{Q}^T$; i.e. the maximum value is the sum of singular values of $\mathbf{K}$.

After solving $\mathbf{H}$, we obtain $\mathbf{S}$ by setting $S_{b\beta} = H_{b\beta} / d(x_b, y_\beta)$. So far the elements of the derived matching matrix $\mathbf{S}$ are not binary. We set the maximum element to be 1 and the other elements to be 0 for each row of $\mathbf{S}$. Note that if $d(x_b, y_\beta) = 0$, we set $S_{b\beta} = 1$ and $S_{b\gamma} = 0$ for $\gamma \neq \beta$.

In summary, we simplify the maximum formulation (5) into an optimization problem (9) and obtain its closed-form solution via the singular value decomposition. Hence, given the data graph and the model graph, we first calculate all pairs of region distances to obtain the matrix $\mathbf{F}$ and then perform singular value decomposition to derive $\mathbf{H}$ and thus $\mathbf{S}$. Since there is no iterative operation in our estimation, the overall region correspondence estimation is rather efficient.

3. IMAGE REPRESENTATIONS AND DISTANCE MEASUREMENT

This section elaborates our definition of image distance in terms of the estimated region correspondence and the low-level features. Recall that each node (region) associates its low-level features as its attribute. We will first define the region distance in terms of low-level features and then define the image distance in a hierarchical manner.

Let N_f be the number of low-level feature types and the attributes of nodes $x_a \in V_D$ and $y_\alpha \in V_M$ are represented by the sets of feature vectors $\{\mathbf{q}_{a1}, \mathbf{q}_{a2}, \ldots, \mathbf{q}_{aN_f}\}$ and $\{\mathbf{f}_{\alpha 1}, \mathbf{f}_{\alpha 2}, \ldots, \mathbf{f}_{\alpha N_f}\}$ respectively. We define the region distance between x_a and y_α as follows:

$$
d(x_a, y_\alpha) = \sum_{l=1}^{N_f} \left\|\mathbf{q}_{al} - \mathbf{f}_{\alpha l}\right\|_2^2, \tag{10}
$$

where $\|\cdot\|_2$ denotes Euclidean distance. In section 2, we redefine D_{ab} and $M_{\alpha\beta}$ using the two functions u_{ab} and $v_{\alpha\beta}$, which are defined as follows:

$$
u_{ab} = \exp\{-d(x_a, x_b)\}, \quad v_{\alpha\beta} = \exp\{-d(y_\alpha, y_\beta)\}. \tag{11}
$$

In section 2, we have determined the matching matrix $\mathbf{S}$. As aforementioned, the model graph G_M may contain unmatched regions. These unmatched regions are often dissimilar to any of the query regions. We have to penalize the image distance by the dissimilarity between query regions and unmatched regions. Hence, we define the image distance as

$$
d(G_D, G_M) = d_M(G_D, G_M) + d_U(G_D, G_M). \tag{12}
$$

Our image distance consists of two parts. $d_M(\cdot,\cdot)$ represents the distance between query regions and matched regions, while $d_U(\cdot,\cdot)$ stands for the distance between query regions and unmatched regions. We use the estimated matching matrix $\mathbf{S}$ to define $d_M(\cdot,\cdot)$ as

$$
d_M(G_D, G_M) = \sum_{a\in V_D}\sum_{\alpha\in V_M} r(x_a) S_{a\alpha} d(x_a, y_\alpha), \tag{13}
$$

where the function $r(\cdot)$ denotes the region area normalized by the image size. A larger query region will have more weight in the calculation of $d_M(\cdot,\cdot)$. For the distance between query regions and unmatched regions, let $V_{UM} \subset V_M$ be the set of unmatched regions in G_M. Then $d_U(\cdot,\cdot)$ is defined as

$$
d_U(G_D, G_M) = \sum_{\gamma\in V_{UM}} r(y_\alpha) \sum_{a\in V_D} d(x_a, y_\gamma). \tag{14}
$$

Since an unmatched region y_γ has no corresponding regions in G_D, we penalize the image distance by the sum of distances between y_γ and all query regions. If an unmatched region occupies larger area, the image distance should be penalized more to reflect that many irrelevant regions exist in this image.

4. EXPERIMENTAL RESULTS

We select 30 categories from the Corel photo gallery as our database. Each category contains 100 images. We first perform the mean-shift based approach [9] to segment images into regions and extract four low-level features from each region. We average the colors of pixels in $L^*u^*v^*$ space to obtain color features. For texture features, we compute the normalized co-occurrence matrix and then extract 5 numerical features, including energy, entropy, contrast, homogeneity, and correlation. For shape features, we measure 7 moment variants by representing the luminance variation along the location change as a probability distribution. The coordinates of the region center is extracted as spatial features.

Images in the same category are defined as relevant. We select 15 categories from our database as test queries to perform the experiment. This experiment has two purposes. First, we compare our work with IRM [1] to show that employing region adjacency indeed improves the accuracy of region correspondence estimation. We restrict the elements of the estimated matching matrix in IRM to be binary. In order to have a fair comparison, we do not use penalization for unmatched regions in this part. Second, we will show that the proposed image matching criterion which incorporates penalization for unmatched regions will improve the retrieval performance. We perform experiments over all 1500 test queries and show the averaged precision-recall curve in Figure 1. GRPM and GPM stand for our work with and without penalization for unmatched regions respectively.

We use a PC with 2.4GHz CPU and 512 RAM to perform experiments. For each query, our method takes about 5 seconds to estimate the region correspondences for all database images and rank the retrieved results. The response time is applicable for the interactive requirement in image retrieval. In addition, since our approach incorporates the adjacent relationship to estimate the region correspondence, Figure 1 shows that the retrieval performance of our method (GRM) outperforms that of IRM. We also show that with penalization for unmatched regions (GPRM), the proposed image matching criterion further improve the performance.

5. CONCLUSIONS

In this paper, we develop a graph-theoretic approach for the image matching problem. We propose a maximum likelihood formulation to estimate the region correspondence. The region correspondence is represented by a matching matrix. Our framework incorporates regions' adjacent relationship as weight to penalize the region distance. Note that our work doesn't include any tuning parameters. We use the extremum principle [8] to derive the closed-form solution for the matching matrix. Thus the estimation in our approach is very efficient. In addition, we define image distance based on estimated region correspondence. Our matching criterion penalizes the image distance by unmatched regions to take account of irrelevant parts in database images.

6. REFERENCES

[1] J. Li, J. Z. Wang, and G. Wiederhold, "IRM: integrated region matching for image retrieval," *Proc. ACM International Conference on Multimedia*, 2000.

[2] H. Greenspan, G. Dvir, and Y. Rubner, "Region correspondence for image matching via EMD flow," *CVPR 2000 Workshop on Content-Based Access of Image and Video Libraries*, 2000.

[3] B. Ko and H. Byun, "Integrated region-based image retrieval using region's spatial relationships," *Proc. International Conference on Pattern Recognition, 2002*.

[4] J. W. Hsieh and W. E. L. Grimson, "Spatial template extraction for image retrieval by region matching," *IEEE Trans. Image Processing*, vol. 12, no. 11, pp. 1404-1415, Nov. 2003.

[5] R. Baeza-Yates and G. Valiente, "An image similarity measure based on graph matching," *Proc. International Symposium on String Processing Information Retrieval*, 2000.

[6] B. Huet and E. R. Hancock, "Inexact graph retrieval," *Proc. IEEE Workshop on Content-Based Access of Image and Video Libraries*, 1999.

[7] C. Y. Li, M. C. Shih and C. T. Hsu, "Image retrieval with relevance feedback based on graph-theoretic region correspondence estimation," *Proc. International Conference on Pattern Recognition*, 2004.

[8] B. Luo and E. R. Hancock, "Structural graph matching using the EM algorithm and singular value decomposition," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 10, Oct. 2001.

[9] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 23, no. 5, pp. 603-619, May 2002.

[10] L.G. Shapiro and G.C. Stockman, *Computer vision*, Prentice Hall, 2001.

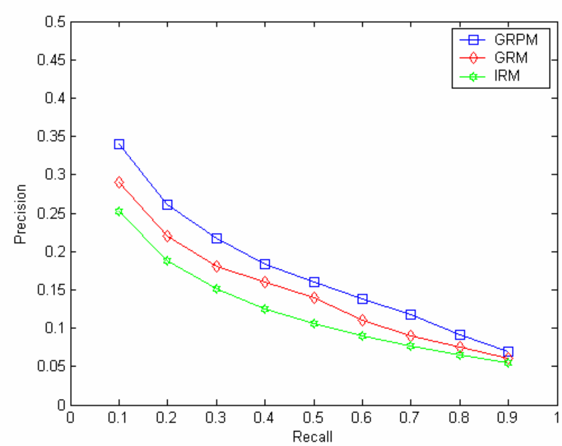

Figure 1. The average precision-recall curve. The image matching criterion with and without penalization are abbreviated to GRPM and GRM respectively.