SOFT REGION CORRESPONDENCE ESTIMATION FOR GRAPH-THEORETIC IMAGE RETRIEVAL USING QUADRATIC PROGRAMMING APPROACH

*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 a graph-theoretic approach to estimate the region correspondence between two images. We represent each image as an attributed undirected graph and transform the image matching problem into an inexact graph matching problem. We formulate the estimation of the soft matching matrix between two graphs as a quadratic programming problem, and apply KKT (Karush-Kuhn-Tucker) conditions and the modified simplex algorithm to solve the constrained optimization problem. With the soft matching matrix, we are capable to integrate both the region correspondence and low-level visual features into an effective matching measurement for image matching. Experiments have been conducted on image retrieval to show the effectiveness of the proposed estimation algorithm.

1. INTRODUCTION

Region-based approaches have become an indispensable issue in content-based image retrieval (CBIR). Representing images in region level captures not only regions' local variations but also their spatial organizations. Study of distance measurements and learning strategies on region-based representation have been proven greatly improving the performance for image matching and relevance feedback [1,2,4,7,8].

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. Correspondence estimation for CBIR is expected to have the following three properties. First, both the region attributes and the adjacent relationship should be incorporated into the estimation process. Second, the estimation should deal with many-to-many mapping issues in case of imperfect segmentation. Consequently, the matching confidence between each pair of matched regions had better be determined with a soft decision. Finally, the estimated region correspondence should be easily incorporated into CBIR and the subsequent relevance feedback steps.

IRM [1] develops a greedy algorithm, MSHP (most similar highest priority) algorithm, to find out the region correspondence in terms of region attributes and region weights. EMD flow [2] adopts a similar formulation to IRM but solves the constrained optimization problem using a linear programming approach. Ko et. al [4] use region centroids as one of the region attributes and apply Hausdorff distance to measure the distance between two sets of regions. However, these methods [1,2,4] takes no account of adjacent relationship between regions into their estimation.

Graph-theoretic approaches have been widely used in correspondence estimation [6-10]. 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. Our work [7,8] employs maximal principal [9] to find out the matching matrix between two images. The main limitation in [7,8] is that the estimated matching matrix is restricted to be an orthogonal matrix and which may contain negative elements. Hence, we heuristically select the largest element in each row as the corresponding element and allow each node in one graph to match only one node in the other graph. We refer to this method as the hard matching method, which is incapable of mapping one region to multiple regions. 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.

In this paper, we aim to employ a graph-theoretic approach to find out the soft matching matrix between two images, i.e. each pair of regions between two images has a confidence value to reflect the importance of the mapping. Our goal is to deal with many-to-many mapping issues and then define the image distance based on the estimated soft matching matrix.

The rest of this paper is organized as follows. Section 2 formulates the graph-theoretic image matching as a quadratic programming problem. Section 3 applies KKT condition and a modified simplex method to find out the optimal soft matching matrix. In section 4, we define 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 the image matching problem, we use a graph (the data graph) to represent an image as $G_D = (V_D, E_D, \mathbf{f}, \mathbf{D})$, where V_D is the set of nodes corresponding to regions, and E_D is the set of edges. An edge $(x_a, x_b) \in E_D$ exists if the two regions x_a, $x_b \in V_D$ are spatially adjacent. The term $\mathbf{f}$ indicates the node attribute vector, and $\mathbf{D}$ is the edge attribute adjacency matrix. Similarly, we refer to another image in the matching problem as the model graph $G_M = (V_M, E_M, \mathbf{g}, \mathbf{M})$.

When the data graph is matched to the model graph, we assume that the node attribute vector $\mathbf{g}$ of the model graph is the permutation of $\mathbf{f}$ with additive noise [10], i.e.

$$\mathbf{g} = \mathbf{P}\mathbf{f} + \varepsilon_1 \mathbf{n}, \tag{1}$$

where $\mathbf{P}$ is a $|V_M| \times |V_D|$ permutation matrix, and ε_1 is the noise magnitude. The elements of the noise vector $\mathbf{n}$ are assumed to be drawn from a zero-mean and unit-variance Gaussian distribution. Similarly, we assume that the two adjacency matrices are related by

$$\mathbf{M} = \mathbf{P}\mathbf{D}\mathbf{P}^T + \varepsilon_2 \mathbf{N}. \tag{2}$$

The goal of our image matching approach is to find out the optimal $|V_D| \times |V_M|$ soft matching matrix $\hat{\mathbf{S}}$ that maximizes the probability

$$\hat{\mathbf{S}} = \arg\max_{\mathbf{S}} P(G_M \mid G_D, \mathbf{S}) \cdot \tag{3}$$

Each element $s_{a\alpha}$ in $\mathbf{S}$ represents the confidence that the node $x_a \in V_D$ matches $y_\alpha \in V_M$. If we restrict elements of $\hat{\mathbf{S}}$ to be binary and only allow one-to-one mapping, then $\hat{\mathbf{S}}^T$ is the same as the permutation matrix $\mathbf{P}$. Hence, in this work, $\hat{\mathbf{S}}^T$ can be treated as a generalization of the permutation matrix.

In (3), the probability $P(G_M \mid G_D, \mathbf{S})$ is factorized according to the definition of conditional probability

$$\begin{aligned} P(G_M \mid G_D, \mathbf{S}) &= P(\mathbf{g}, \mathbf{M} \mid \mathbf{f}, \mathbf{D}, \mathbf{S}) \\ &= P(\mathbf{g} \mid \mathbf{f}, \mathbf{D}, \mathbf{S}) P(\mathbf{M} \mid \mathbf{f}, \mathbf{g}, \mathbf{D}, \mathbf{S}). \end{aligned} \tag{4}$$

Using the assumption defined in (1), the probability of the node attribute vector $\mathbf{g}$ is independent of the edge attribute adjacency matrix $\mathbf{D}$. From (1), each element in $\mathbf{g}$ is defined by $g_\alpha = \sum_{a \in V_D} s_{a\alpha} f_a + \varepsilon_1$. Hence, we define

$$P(\mathbf{g} \mid \mathbf{f}, \mathbf{D}, \mathbf{S}) = P(\mathbf{g} \mid \mathbf{f}, \mathbf{S}) = \exp\left\{ -\sum_{a \in V_D} s_{a\alpha} d(f_a, g_\alpha) \right\} \cdot \tag{5}$$

For the second term $P(\mathbf{M} \mid \mathbf{f}, \mathbf{g}, \mathbf{D}, \mathbf{S})$ in (4), we first assume that elements in the edge attribute adjacency matrix $\mathbf{M}$ are statistically independent, given $\mathbf{f}$, $\mathbf{g}$, $\mathbf{D}$, and $\mathbf{S}$. Hence we have

$$P(\mathbf{M} \mid \mathbf{f}, \mathbf{g}, \mathbf{D}, \mathbf{S}) = \prod_{\alpha \in V_M} \prod_{\beta \in V_M} P(M_{\alpha\beta} \mid \mathbf{f}, \mathbf{g}, \mathbf{D}, \mathbf{S}) \cdot \tag{6}$$

Furthermore, by the assumption defined in (2) and from $M_{\alpha\beta} = \sum_{a \in V_D} \sum_{b \in V_D} s_{a\alpha} D_{ab} s_{b\beta}$, we define

$$P(M_{\alpha\beta} \mid \mathbf{f}, \mathbf{g}, \mathbf{D}, \mathbf{S}) = \exp\left\{ -\sum_{a \in V_D} \sum_{b \in V_D} s_{a\alpha} d(D_{ab}, M_{\alpha\beta} \mid \mathbf{f}, \mathbf{g}) s_{b\beta} \right\}. \tag{7}$$

By substituting (4), (5), (6), and (7) into (3) and defining

$$\phi_{a\alpha} = d(f_a, g_\alpha), \ \psi_{a\alpha b\beta} = d(D_{ab}, M_{\alpha\beta} \mid \mathbf{f}, \mathbf{g}), \tag{8}$$

we have

$$\begin{aligned} &\arg\max_{\mathbf{S}} P(G_M \mid G_D, \mathbf{S}) \\ &= \arg\max_{\mathbf{S}} \log P(G_M \mid G_D, \mathbf{S}) \\ &= \arg\max_{\mathbf{S}} \{ \log P(\mathbf{g} \mid \mathbf{f}, \mathbf{D}, \mathbf{S}) + \log P(\mathbf{M} \mid \mathbf{f}, \mathbf{g}, \mathbf{D}, \mathbf{S}) \} \\ &= \arg\max_{\mathbf{S}} \left\{ -\sum_{a \in V_D} \sum_{\alpha \in V_M} s_{a\alpha} \phi_{a\alpha} - \sum_{a \in V_D} \sum_{\alpha \in V_M} \sum_{b \in V_D} \sum_{\beta \in V_M} s_{a\alpha} s_{b\beta} \psi_{a\alpha b\beta} \right\}, \end{aligned} \tag{9}$$

subject to the constraints

$$\sum_{a \in V_D} s_{a\alpha} \le 1 \ \text{ for } \alpha \in V_M; \tag{10}$$

$$\sum_{\alpha \in V_M} s_{a\alpha} \le 1 \ \text{ for } a \in V_D; \tag{11}$$

$$\sum_{a \in V_D} \sum_{\alpha \in V_M} s_{a\alpha} = \min(|V_D|, |V_M|); \tag{12}$$

$$s_{a\alpha} \ge 0 \ \text{ for } \alpha \in V_M \text{ and } a \in V_D. \tag{13}$$

The constraints (10-12) are defined based on the observation that the numbers of nodes in the data graph and in the model graph are usually different. Note that, the constraint defined in (12) will be removed by including slack variables in the original problem. Suppose $|V_D| > |V_M|$, then we have $\sum_{a=1}^{|V_D|} s_{a\alpha} = 1$, $\sum_{\alpha=1}^{|V_M|} s_{a\alpha} \le 1$, and (12) will hold accordingly. If we include the slack variables $\zeta_a \ge 0$ in (11), we can transform the inequality constraint defined in (11) into the equality constraints $\sum_{\alpha=1}^{|V_M|} s_{a\alpha} + \zeta_a = 1$ and no longer need to define the constraint (12). In the following derivation, we will assume $|V_D| > |V_M|$. The case for $|V_D| < |V_M|$ can be similarly derived.

Note that in (9), if we drop the term $P(\mathbf{M} \mid \mathbf{f}, \mathbf{g}, \mathbf{D}, \mathbf{S})$, then this formulation turns out to be the same as the EMD flow method [2]. Thus, EMD flow method can be treated as a special case of our formulation.

3. OPTIMAL SOLUTION FOR THE SOFT MATCHING MATRIX

In (9), the objective function includes the term $s_{a\alpha}s_{b\beta}$ and thus becomes a quadratic programming problem. In order to derive the optimal matching matrix $\mathbf{S}$, we follow the procedure in [11] to solve (9). Let μ_α ($\alpha \in V_M$), ν_a ($a \in V_D$) represent the corresponding dual variables of (10) and (11), respectively. The KKT (Karush-Kuhn-Tucker) conditions of (9) are as follows.

$$-\phi_{a\alpha} - s_{a\alpha}\psi_{a\alpha a\alpha} - \sum_{b=1}^{|V_D|}\sum_{\beta=1}^{|V_M|} s_{b\beta}\psi_{a\alpha b\beta} - \mu_\alpha - \nu_a \le 0 ; \tag{14}$$

$$s_{a\alpha}\left[-\phi_{a\alpha} - s_{a\alpha}\psi_{a\alpha a\alpha} - \sum_{b=1}^{|V_D|}\sum_{\beta=1}^{|V_M|} s_{b\beta}\psi_{a\alpha b\beta} - \mu_\alpha - \nu_a\right] = 0 ; \tag{15}$$

$$\sum_{a=1}^{|V_D|} s_{a\alpha} - 1 = 0 ; \tag{16}$$

$$\sum_{\alpha=1}^{|V_M|} s_{a\alpha} + \zeta_a - 1 = 0 ; \tag{17}$$

$$\mu_\alpha\left[\sum_{a=1}^{|V_D|} s_{a\alpha} - 1\right] = 0 ; \tag{18}$$

$$\nu_a\left[\sum_{\alpha=1}^{|V_M|} s_{a\alpha} + \zeta_a - 1\right] = 0 ; \tag{19}$$

$$s_{a\alpha} \ge 0 ,\ \mu_\alpha \ge 0 ,\ \nu_a \ge 0 ,\ \zeta_a \ge 0 . \tag{20}$$

We introduce nonnegative slack variables $\gamma_{a\alpha} \ge 0$ to convert (14) into equalities:

$$-\phi_{a\alpha} - s_{a\alpha}\psi_{a\alpha a\alpha} - \sum_{b=1}^{|V_D|}\sum_{\beta=1}^{|V_M|} s_{b\beta}\psi_{a\alpha b\beta} - \mu_\alpha - \nu_a + \gamma_{a\alpha} = 0 . \tag{21}$$

Furthermore, we have the complementarity constraint

$$\sum_{a\in V_D}\sum_{\alpha\in V_M} s_{a\alpha} r_{a\alpha} = 0 , \tag{22}$$

which ensures that only one of the two variables in $(s_{a\alpha}, r_{a\alpha})$ can be nonzero.

In order to determine the initial basic feasible solution, we introduce artificial variables p_α, q_a, and $z_{a\alpha}$ to revise the problem. The revised constraints for (16), (17), and (21) are as follows:

$$\sum_{a=1}^{|V_D|} s_{a\alpha} - 1 + p_\alpha = 0 ; \tag{23}$$

$$\sum_{\alpha=1}^{|V_M|} s_{a\alpha} + \zeta_a - 1 + q_a = 0 ; \tag{24}$$

$$-\phi_{a\alpha} - s_{a\alpha}\psi_{a\alpha a\alpha} - \sum_{b=1}^{|V_D|}\sum_{\beta=1}^{|V_M|} s_{b\beta}\psi_{a\alpha b\beta} - \mu_\alpha - \nu_a + \gamma_{a\alpha} + z_{a\alpha} = 0 ; \tag{25}$$

Finally, we have the following linear programming problem:

$$\text{Minimize} \quad \sum_{a=1}^{|V_D|}\sum_{\alpha=1}^{|V_M|} z_{a\alpha} + \sum_{\alpha=1}^{|V_M|} p_\alpha + \sum_{a=1}^{|V_D|} q_a \tag{26}$$

subject to (23), (24), (25)

$$\text{and } s_{a\alpha}, \mu_\alpha, \nu_a, \zeta_a, \gamma_{a\alpha}, z_{a\alpha}, p_\alpha, q_a \ge 0 . \tag{27}$$

Using phase 1 of the two-phase method [11], we iteratively derive the optimal solution for all the elements $s_{a\alpha}$ in the matching matrix $\mathbf{S}$. The optimal solution achieves when all the artificial variables are zero, according to the restricted entry rule [11].

In summary, by assuming that the nodes attributes and the edge attribute adjacency matrix of the model graph are permutations of that of the data graph, we transform the inexact graph matching problem (3) into a quadratic programming problem (9). The optimal soft matching matrix can be derived by the KKT conditions and the modified simplex method.

4. IMAGE REPRESENTATION AND DISTANCE MEASUREMENT

This section elaborates our distance measurement for node attributes and edge attributes in (8). Although in section 2, we mainly deal with nodes with single attribute, extending our formulation to be with multi-valued node attributes is very intuitive. Let f_a and g_α represent feature vectors of nodes x_a and y_α respectively, we define

$$\phi_{a\alpha} = d(f_a, g_\alpha) = \|f_a - g_\alpha\|_2^2 , \tag{28}$$

where $\|\cdot\|_2$ denotes the Euclidean distance.

Next, according to the edge consistency between two graphs, we define

$$\psi_{a\alpha b\beta} = D_{ab} M_{\alpha\beta} \phi_{a\alpha}\phi_{b\beta} . \tag{29}$$

Two edges are said to be consistent if both $(x_a, x_b) \in E_D$, $(y_\alpha, y_\beta) \in E_M$. If two edges are inconsistent, we define the distance as the product of node attributes' distance.

Based on the estimated soft matching matrix, we then define the image distance as

$$d(G_D, G_M) = \sum_{a\in V_D}\sum_{\alpha\in V_M} s_{a\alpha}\phi_{a\alpha} . \tag{30}$$

5. 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 [12] 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. Since our formulation can be regarded as a generalization of EMD flow [2], and EMD flow is also a soft matching approach, we compare our work with EMD flow to show that employing region adjacency indeed improves the accuracy of region correspondence estimation. We perform experiments over all 1500 test queries and use the averaged precision-recall curve to measure the retrieval performance. Figure 1 shows that our method outperforms EMD flow.

6. CONCLUSIONS

In this paper, we develop a graph-theoretic approach for the image matching problem. The region correspondence is represented by a matching matrix. Our framework incorporates adjacent relationship between regions as a weight to penalize the region distance. The first contribution of this work is to formulate the estimation of soft matching matrix into an inexact graph matching problem. The second contribution of this work is to solve the constrained optimization problem for inexact graph matching using a quadratic programming approach. The experimental results demonstrate that the proposed estimation method indeed achieve satisfactory result for image retrieval.

7. REFERENCES

[1] J. Li, J. Z. Wang, and G. Wiederhold, "IRM: integrated region matching for image retrieval," *Proc. ACM 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] D. H. Kim, I. D. Yun, and S. U. Lee, "A new attributed relational graph matching algorithm using the nested structure of Earth Mover's distance," *Proc. International Conference on Pattern Recognition*, 2004.

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

[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 and C.T. Hsu, "Region correspondence for image retrieval using graph-theoretic approach and maximum likelihood estimation," *Proc. International Conference on Image Processing*, 2004.

[8] 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.

[9] 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.

[10] B. J. van Wyk, M. A. van Wyk, "Kronecker product graph matching," *Pattern Recognition Journal*, vol. 36, pp. 2019-2030, 2003.

[11] F. S. Hillier and G. J. Lieberman, *Introduction to mathematical programming*, McGraw-Hill, 1990.

[12] 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.

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

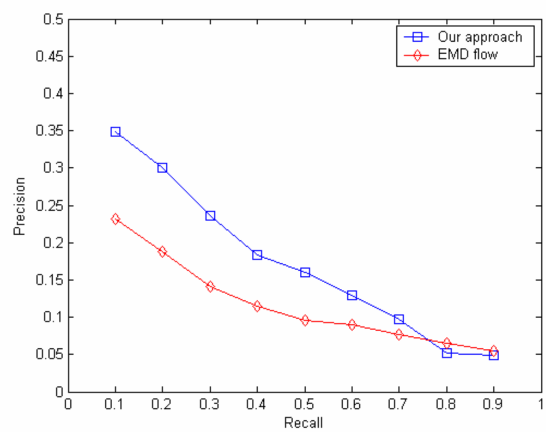

Figure 1. The averaged precision-recall curve.