10-2005

Common pattern discovery using earth mover's distance and local flow maximization

Hung-Khoon TAN

Chong-wah NGO
Singapore Management University, cwngo@smu.edu.sg

Common Pattern Discovery using Earth Mover's Distance and Local Flow Maximization

Hung-Khoon Tan and Chong-Wah Ngo
Department of Computer Science
City University of Hong Kong
83 Tat Chee Aveune, Kowloon, Hong Kong
{hktan, cwngo}@cs.cityu.edu.hk

Abstract

In this paper, we present a novel segmentation-insensitive approach for mining common patterns from 2 images. We develop an algorithm using the Earth Movers Distance (EMD) framework, unary and adaptive neighborhood color similarity. We then propose a novel local flow maximization approach to provide the best estimation of location and scale of the common pattern. This is achieved by performing an iterative optimization in search of the most stable flows' centroid. Common pattern discovery is difficult owing to the huge search space and problem domain. We intend to solve this problem by reducing the search space through identifying the location and a reduced spatial space for common pattern discovery. Experimental results justify the effectiveness and the potential of the approach.

1. Introduction

Huge amount of visual information in the form digital images and video database are generated everyday. Extracting common patterns from the images are becoming increasingly important for various applications. First, they can serve as entry points for classification and browsing of various data in a database. Secondly, they form the basic elements for video indexing, clustering and summarizing.

However, mining common pattern discovery is a difficult problem because the search scale and problem domains are huge. To perform a common pattern discovery, the common problems of rotation, scaling, translation, occlusion, shot angle and segmentation have to addressed. Existing approach often ignores some of the problems, hence limiting the target image categories. The problem can be significantly reduced if we can reduce the search domain by identifying the location of the common pattern and a reduced spatial space for the search to be further refined.

In this paper, we propose a novel approach for automatic discovery of common patterns between two images. We use Earth Mover's Distance (EMD) as our framework. Flows with small distances are extracted and decomposed into components. We encourage the mass flow between the common patterns by using adaptive neighborhood color similarity. The components are then analyzed using a novel Local Flow Maximization (LFM) approach to locate the common patterns. Our approach is especially robust to segmentation, rotation, shot angle and translation.

1.1. Common Pattern Discovery

Until recently, there is only a small amount of previous works on the common pattern discovery. There is also no widely accepted definition for the problem. [4] provides a good general definition of common pattern discovery. Given two images I and J, common pattern discovery is defined as finding the best transformation T, subimage I^* of I and subimage J^* of J which satisfy the following equation

$$T, I^*, J^* = \underset{T, I^*, J^*}{argmax}\ H(T(I^*), J^*) \tag{1}$$

where H is the similarity function of the features and size of the common pattern, and T is any similar transformation set.

Common Pattern Discovery can be regarded as a superset of pattern registration and common pattern detection. In common pattern detection, a known common pattern is used as an input to detect if the same pattern exist in the other image. For pattern registration, the geometrical transformation between the common patterns are discovered. Common pattern discovery is a harder problem because it mines for unknown common patterns from a set of images without any prior information and in the process, geometrical transformations are often retrieved. Even the fact if any common patterns exist in the images is also unknown.

Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05)
1550-5499/05 $20.00 © 2005 **IEEE**

Figure 1 illustrates the challenge of mining common pattern from two images in arbitrary backgrounds. The problem of discovery is difficult, particularly when the common patterns are occluded, scaled and shot from different views under varying lighting conditions.

(a) Two images (b) Result

Figure 1. Common pattern discovery

2. Previous Works

Related works on the problem of common pattern discovery include probabilistic model learning [2], subgraph mining [3], maximum weighted bipartite graph mining [4], multiple-instance learning [6] and visual model learning [10],

The studies in [2, 3, 4, 6, 10] are based on one-to-one mapping techniques, where a region or node is mapped to another node. In [6], images are grouped into positive and negative images (bags). The algorithm searches for the intersection of the positive bags, minus the union of negative bags. The method is limited because specific detectors are sometimes required for discovering certain patterns. In [10], pattern discovery is carried out by determining the features common to the positive and rare in the negative examples. A probabilistic visual model based on neighborhood-frequency descriptors and significance measure is used to model the common features. Both [6, 10] belongs to weakly supervised learning because they require manual labelling of positive and negative images.

[2] uses graph matching and EM algorithm, where a common pattern is assumed as a linear combination of model components. Graph matching is used to align initial pattern model, and EM is used to discover the model components and their corresponding parameters. This approach is not flexible because the number of model components are normally unknown and require user input. [3] is a purely graph-based matching technique where images are segmented based on colors, and modelled as an Attributed Relational Graph (ARG). Common pattern is defined as the maximum common subgraph in the images. However, the algorithm is sensitive to segmentation and is limited to patterns with colorful and sharp-edged regions. [4] is a robust block-based mapping. It solve the segmentation problem by not performing any segmentation at all. Images are partitioned into small square blocks, and are represented as a node in a bipartite graph. An iterative maximum weighted bipartite graph matching using estimated transformation value obtained through procrustes analysis is used to find the matching nodes. All [2, 3, 4] could be considered as graph-based matching techniques.

There are still many areas of improvement to detect common patterns from images, especially for unsupervised learning. Most of the methods utilize the one-to-one mapping techniques where the number of nodes in all the common patterns in the different images must be similar. However, this is normally impossible for most real images, owing to occlusion, viewpoint difference, shadings and segmentation. Therefore, a many-to-many mapping technique is a more attractive alternative to solving the common pattern problem. In this paper, we propose to use the Earth Mover's Distance (EMD) as our framework, through a novel Local Flow Maximization approach. We will show that our approach are insensitive to image segmentation and can handle complex common patterns, which are difficult in one-to-one mapping techniques.

3. Algorithm Overview

Figure 2 shows an overview of the algorithm. The algorithm is divided into two portions: the initialization sequence and flow maximization sequence.

In the initialization stage, the images are converted into Attributed Relational Graphs (ARG). A node similarity matrix is created and used to calculate the node distances in EMD. An initial EMD is performed over the whole image and flows with small distances are extracted. Adjacent source and destination nodes are grouped into components. Each component is potential container of a common pattern, which would be mined in the next stage. If there are more than one common pattern in the images and they are grouped into different components in either one of the images in this initialization stage, Local Flow Maximization would be able to retrieve multiple common patterns from the images.

In the Local Flow Maximization stage, the components in the first image are matched with the components in the second image to determine if a possible match occurs. The Local Flow Maximization is based on the Expectation Maximization framework, which iteratively finds the location where the local flow maximizes. If a common pattern exists, the flow's centroid should converge at the centroid of the common pattern.

4. Earth Mover's Distance

The Earth Mover's Distance (EMD) is a multi-point matching technique, which is based on an old transportation problem. It is a flexible metric, which allows for par-

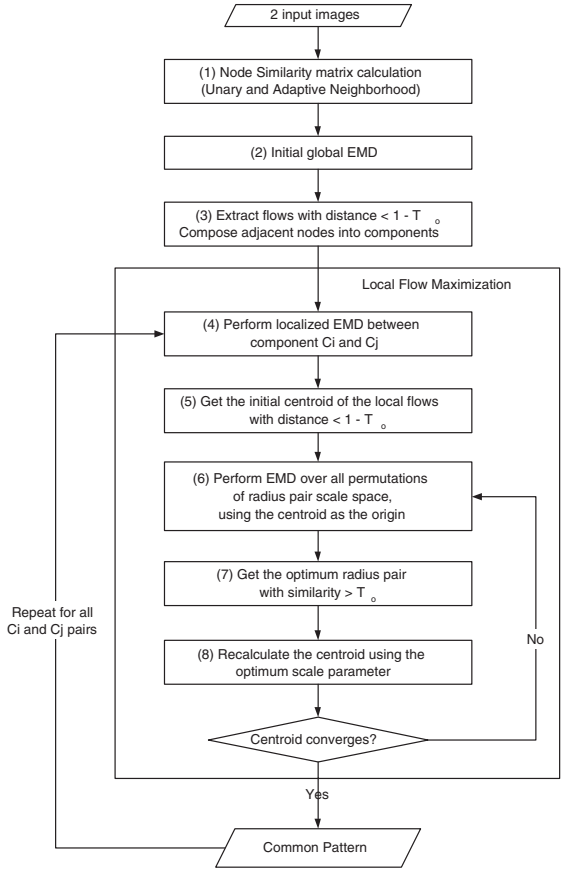


Figure 2. Overview of our approach

tial matching, and can be applied to variable-length representations of distributions. It allows huge information to be compressed into signatures and still yield better results compared to the conventional histogram methods for image comparison. It has been successful in image retrieval [8] and image database navigation and visualization [7]. Lately, it has been extended for low-level image processing for corner, junction and edge detection [9].

In EMD, the result is the minimum amount of work which minimizes the overall cost of moving the mass from a signature to another.

$$EMD(\mathbb{S}, \mathbb{D}) = \min WORK(\mathbb{S}, \mathbb{D}, \mathbf{F}) \qquad (2)$$

$$WORK(\mathbb{S}, \mathbb{D}, \mathbf{F}) = \sum_{i=1}^{m} \sum_{j=1}^{n} dist_{ij} \times flow_{ij} \qquad (3)$$

where flow $\mathbf{F} = [flow_{ij}]$, with $flow_{ij}$ representing the flow from node s_i in the source signature $\mathbb{S}$ to node d_j in the destination signature $\mathbb{D}$, subject to the constraints specified in [8]. $dist$ is the dissimilarity measure between the two nodes. Since EMD's flows are based on the minimum distance to minimize the total cost, we can prioritize the flow between the common patterns if we could provide a higher similarity value for the nodes within the common patterns. We achieve this by using an adaptive neighborhood similarity method, which we will elaborate in section 4.2.

4.1. Data Representation and Settings

We use Attributed Relational Graph (ARG) as our data representation because it can be readily converted into the EMD signatures. The image is initially segmented by its color distribution and then converted into an ARG model. Each ARG node is a color cluster, containing the following information: node weight (normalized number of pixels for the cluster), centroid position and average color value. The edges represent the spatial neighborhood information between the ARG nodes. Each ARG Nodes can be conveniently used as the input for generating the EMD's signatures.

The EMD color signatures is a data structure consisting of a set of ordered pairs $\{(c_1,w_1), (c_2,w_2), ..., (c_n,w_n)\}$, where c_i is the cluster representation and w_i is the total number of pixels for the cluster normalized by the image size. For simplicity, we refer the ordered pair as s_i for the source $\mathbb{S}$ and d_i for the destination $\mathbb{D}$.

To compute the perceptual distance between the two signatures, CIE-L*a*b color space is used because the Euclidean distance between two nearby colors in the space is more equivalent to their perceptual distance, compared to the Red, Green and Blue (RGB) and Hue, Saturation and Value (HSV) color space. Having defined the requirement for the ground distance, we can define the distance metric between s_i and d_j as

$$dist_{node}(i,j) = 1 - H_{node}(i,j) \qquad (4)$$

where $H_{node}(i,j)$ is specified by equation (5). The distance $dist_{node}$ ranges from 0 to 1, with 0 being the smallest distance between two nodes.

4.2. Node to Node Similarity

The fundamental problem in our method is how to encourage the flows between the nodes of the common patterns. We achieve this by tapping into the neighborhood information of each nodes. Kim et. al. [5] utilizes a modified nested EMD to calculate the similarity between 2 nodes using the neighborhood information of a central node for his perceptual 3-D shape descriptor. Sivic et. al. [11] utilizes common neighborhood structure and extent to mine for common patterns from the key frames of a video sequence. Candidate matching objects are obtained when both nodes have M matching neighbors. Their results prove

that structural information in terms of neighborhood similarity is a powerful discriminating feature that could be tapped for common pattern mining. For our case, we use the neighborhood node's color information.

To overcome missing or faulty neighborhood assignment caused by segmentation, we perform an adaptive neighborhood search to perform a recursive search on a neighboring node if it is similar in color to the parent nodes, and mark only nodes with low color similarity as a neighbor. In this scheme, nodes with more matching neighbor nodes are assigned a higher similarity value. During segmentation, a region could be segmented into different color regions with low distance value. Applying a direct neighborhood measure results in a missing neighborhood information. In the adaptive neighborhood search technique, the nodes which belong to the same region will retrieve the same set of correct neighbor node set by performing recursive search on neighboring nodes until all neighboring nodes have different color from the node's color.

Finally, the node-to-node similarity H_{node} is a weighted similarity metric between the unary color similarity H_{unary} and the adaptive neighborhood similarity $H_{neighbor}$. Let α denotes the weighting coefficient, A_{node} denotes the total number of matching neighbors and β denotes the minimum number of matching neighbor pairs for a perfect neighbor similarity value. H_{node} is defined as

$$H_{node}(i,j) = \alpha H_{unary}(i,j) + (1-\alpha) H_{neighbor}(i,j) \quad (5)$$

where

$$H_{unary}(i,j) = exp\left[-(\frac{D(i,j)}{\gamma})^2\right] \quad (6)$$

and

$$H_{neighbor} = \begin{cases} \frac{1}{\beta} \times A_{node} & A_{node} < \beta \\ 1 & A_{node} \geq \beta \end{cases} \quad (7)$$

D is the Euclidean distance of the colors in the CIE-L*a*b color space given by

$$D(i,j) = [(\triangle L)^2 + (\triangle a)^2 + (\triangle b)^2]^{\frac{1}{2}} \quad (8)$$

Both α and β are determined empirically from our experiments, where β has a lower bound value of 2. From our experiments, we manage to produce consistently accurate results by setting α to 0.6 and β to 5.

The similarity is formulated so that $H_{unary}(i,j)$ is an exponential measure with the steepness of the slope governed by γ, which we set to 30. The final similarity node-to-node matrix $\mathbf{H}_{node}$ is a $M_a \times N_a$ matrix, where M_a and N_a are the total ARG nodes in each image.

5. Local Flow Maximization

5.1. Centroid-based Similarity Matrix

A centroid is generally the center point of a system of masses. We define the flows' centroid as the center of mass for the points involved in the EMD flows having distance smaller than a threshold value $1 - T_o$, where each node is assigned a weight of one. T_o is a similarity threshold value over which two nodes are assumed to be similar. T_o is determined empirically, and the rule of thumb is to set it to a value larger than the weighting coefficient α. We obtain consistently accurate results by setting T_o to 0.8 in our experiments.

Let $\mathbb{S} = \{s_1, s_2, \ldots, s_M\}$ and $\mathbb{D} = \{d_1, d_2, \ldots, d_N\}$, where M and N are the total number of source and destination nodes, contributing to EMD flows with $dist_{node} < 1 - T_o$. The flows' centroid for the source and destination images can be formulated as

$$C_s(x,y) = \frac{1}{M}\sum_{i=1}^{M} s_i(x,y) \quad (9)$$

$$C_d(x,y) = \frac{1}{N}\sum_{i=1}^{N} d_i(x,y) \quad (10)$$

where $s_i(x,y)$ and $d_i(x,y)$ denote the centroid of the cluster s_i and d_i.

Having defined the flows' centroid, we can now define the similarity metric used in Local Flow Maximization. The metric must fulfil the following requirements: (i) Similarity increases when the centroid of the EMD flow approaches the centroid of the common patterns in both image. (ii) The highest similarity happens when the centroid of the flow overlaps with the centroid of both common patterns. (iii) Provides the best estimate of the matching area using the current centroid.

We use the C_s and C_d as the centers of the area encompassed by a circle with radius r_i and r_j respectively. All ARG nodes whose centroid fall within the radius are selected to generate the EMD's signatures. The resolution of the radius scale space is $\triangle r_{s,d} = L_{s,d}/R$ where $L_{s,d}$ are the diagonals of the images and R is the total number of radius steps. For all our experiments, we set R to 20 which provide a reasonable resolution and run time. We perform an EMD for all permutations of signature pair generated from the different radius settings. Thus, we end up with a $R \times R$ similarity matrix $\mathbf{H}_{centroid}$. This could be formulated as

$$\mathbf{H}_{centroid}(r_i, r_j) = 1 - \mathrm{EMD}(\mathbb{S}_{r_i}, \mathbb{D}_{r_j}) \quad (11)$$

where

$$\mathbb{S}_{r_i} = \bigcup_{k=1}^{M} s_k(x,y), \quad \forall_k\, s_k(x,y) \leq C_s(x,y) + r_i \quad (12)$$

and

$$\mathbb{D}_{r_j} = \bigcup_{k=1}^{N} d_k(x,y), \quad \forall_k d_k(x,y) \leq C_d(x,y) + r_j \quad (13)$$

We define the centroid-based similarity value H_{object} as the total number of elements in $\mathbf{H}_{centroid}$ which have similarity value exceeding the object similarity threshold T_o. H_{object} can be formulated as

$$H_{object} = \sum_i \sum_j \begin{cases} 1 & \mathbf{H}_{centroid}(r_i, r_j) > T_o \\ 0 & \text{otherwise} \end{cases} \quad (14)$$

Intuitively, H_{object} meets the requirement (i) and (ii) because the similarity value will display the highest similarity value when the flows' centroid is correlated with the common pattern's centroid. For requirement (iii), all radius pairs with similarity value larger than T_o are possible candidates for the matching area. For our case, the best estimate of matching area should be the value where the radius scale pair maximizes to allow more space for EMD to find the next best centroid space. Thus, we have

$$r_i^*, r_j^* = \underset{r_i, r_j}{\operatorname{argmax}}(r_i + r_j), \quad (15)$$

where we consider only the radius pairs with similarity value larger than T_o.

5.2. Local Flow Expectation Maximization

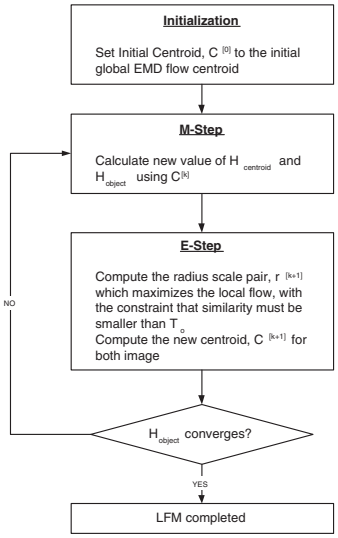


Figure 3. Overview of the LFM algorithm

The Local Flow Maximization (LFM) is the EMD's local flow maximization method based on the Expectation Maximization (EM) framework. The hidden data is the centroid and the radius pairs when the local flows maximizes with the constraint that the area within radius pair has similarity higher than T_o. The observed values are the EMD flows over the radius scale space, from which we obtain the centroid similarity matrix $\mathbf{H}_{centroid}$ and the object similarity value H_{object}. LFM aims to find the centroid and radius pairs by maximizing H_{object}. Figure 3 shows an overview of the LFM algorithm.

In the Expectation stage, the best matching local area are determined from equation (15) to maximize the local flow. By calculating the flows in the area, a new value of the centroid value is obtained from equations (9) and (10). For simplicity, we summarize the sequence as a function E, which takes in the centroid-based similarity matrix $\mathbf{H}_{centroid}^{(k)}$ as input and produces the centroid and radius pair as outputs. Therefore, the maximization process could be formulated as

$$\begin{bmatrix} C_s \\ C_d \\ r_s \\ r_d \end{bmatrix}^{(k+1)} = E(C_s, C_s, r_s, r_d | \mathbf{H}_{centroid}^{(k)}) \quad (16)$$

In the Maximization stage, the new centroid value is used to perform a series of EMD matching over all radius scale space to retrieve a new $\mathbf{H}_{centroid}$, from which we calculate a new object similarity value H_{object}, which are obtained from equations (13), (11) and (14). We denote the sequence as a function Q, which takes in $C_{s,d}$ as input and produces the $\mathbf{H}_{centroid}$ and H_{object} as output. Therefore, the expectation process could be formulated as

$$\begin{bmatrix} \mathbf{H}_{centroid} \\ H_{object} \end{bmatrix}^{(k+1)} = Q(C_s^{(k+1)}, C_d^{(k+1)}) \quad (17)$$

The two steps in LFM are iterated until the object similarity value H_{object} converges when $\|H_{object}^{(k+1)} - H_{object}^{(k)}\| = 0$. As in EM, convergence will be to a local maximum which depends on the initial starting point. In our algorithm, the initial case is set to the centroid of the global EMD flow. In the event where there are two common patterns, it will converge to the object with more features. If the common pattern has similar number of features, LFM will converge into a point between the two common patterns. However, the centroid will not converge to a global maximum, but rather to the nearest local maximum. Therefore, the result would be affected by the location of the initial global centroid.

The noise encountered by LFM is the background nodes with low distance in the EMD flow. For example, if a lot of the background nodes are similar to the common pattern's internal or boundary node's color, we will notice an increase of noise. If the background noise level exceeds the common pattern's flow, Local Flow Maximization will not

Table 1. Result of Common Pattern Discovery

CATEGORY	HIT_RATIO	BG_RATIO
Card1	0.9994	0.0269
Doll	0.8054	0.0609
Dustbin	0.8466	0.0000
Big Cat	0.8210	0.0715
Map	1.0000	1.8668
Milo	0.9843	0.0690
Milo2	0.9670	1.2100
RoadSign	1.0000	0.0000
Average	0.9280	0.4648

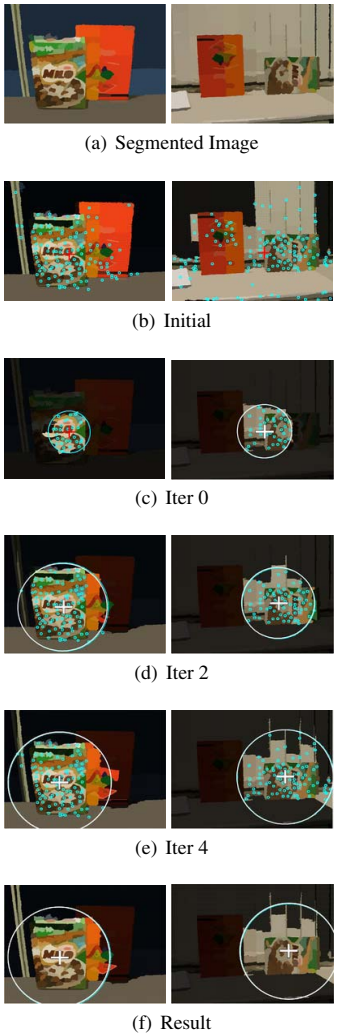
converge to the centroid of the common pattern. The level of background noise depends on how efficient the node-to-node similarity matrix H_{node} can prioritize the EMD flows. In our case, we use the adaptive neighborhood similarity to provide a higher similarity value for the common patterns.

6. Results and Discussion

Common pattern discovery is still a new area of research and we are not aware of any widely accepted method to quantify the quality of the common pattern algorithm result. For this paper, we compute the performance by using two parameters, hit_ratio and bg_ratio. We manually label the precise area of the common patterns and use it to calculate the ratio of the correct matched area versus the total correct area, hit_ratio and the ratio of the incorrect match area versus the total correct area, bg_ratio. Table 1 shows the result of common pattern discovery for several test cases in our test case. Our algorithm manages to detect around 93% of the common pattern area, with an average of 46% background noise.

To verify our Local Flow Maximization portion, we configure our experiment by capturing an image with two objects: one with a lot of color features and the other with few color features. The centroid of the flow should converge to the second common pattern. The result of our algorithm is shown in Figure 4. The large circle marks the best radius pair r, the small circles marks the flows with distance smaller than T_o and the cross (+) shows the centroid of the similar EMD flows for each iterations. The darkened areas are the areas with nodes involved in EMD flows with large distances. Therefore, they are marked as background nodes during the iteration. We could see how the centroid converges to the common pattern with significant color features and and the radius scale changes during each iteration. In the experiment, the centroid converges after the fifth iteration.

We also perform several experiments to show that our algorithm is insensitive to segmentation, translation, rotation, shot angle and partial occlusion. From the results in Figure 4, we can observed that the algorithm could handle translation and rotation with relative ease, since the algorithm is designed to find the centroid of the common patterns and matching is done over radius scale space.



(a) Segmented Image

(b) Initial

(c) Iter 0

(d) Iter 2

(e) Iter 4

(f) Result

Figure 4. Local Flow Maximization

We show that our approach is insensitive to segmentation, a difficult problem especially for one-to-one mapping techniques. We verify this by performing a series of matching using the same image with different segmentation settings. We use the segmentation algorithm proposed in [1] and set the minimum number of pixels per-cluster to 100, 150, 200, 250, 300, 350 and 400. As shown in Figure 5, our algorithm manages to detect the position and the matching area with good accuracy for all cases. This is owing to the

adaptive neighborhood methodology and the many-to-many mapping technique.

The result in Figure 6 shows that the algorithm is capable of performing multiple common pattern matching when there are more than two common pattern coexisting in the images. This is possible if the patterns are distinctive during component generation. This is affected by how well the node similarity, H_{node} can distinguish between the common patterns. In this setting also, we show that the algorithm can handle a certain degree of occlusion as long as the major color features are still visible.

(a) Minimum 100 pixel per cluster segmentation settings

(b) Minimum 150 pixel per cluster segmentation settings

(c) Minimum 250 pixel per cluster segmentation settings

(d) Minimum 350 pixel per cluster segmentation settings

(e) Minimum 400 pixel per cluster segmentation settings

Figure 5. Results using different segmentation settings

Figure 7, 8 and 9 shows some of the other experiments that we have conducted. The images in 7 and 8 are difficult for one-to-one mapping techniques because the number of nodes are different in both images. Our approach manages to extract them with good accuracy and speed. Most other algorithms based on subgraph mining are NP-Hard, and consumes exponential time with respect to N, the total number of nodes. For example, the algorithm used in [3] requires $2 \times (N!)$ time and execution time amounts to several hours when N grows to more than 200 on a 3GHz Pentium 4 with 512MB memory. With our algorithm, we completed the task within a time frame of 5-10 minutes. Figure 7 also

(a) Segmented image

(b) First match

(c) Second match

Figure 6. Multiple common pattern detection

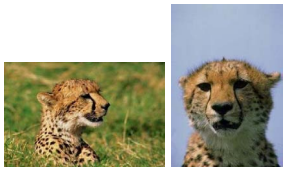
shows that our approach is robust to shot angle. The result in Figure 8 shows a rough matching area because the backgrounds contain a lot of nodes which are similar in color.

7. Summary and Conclusions

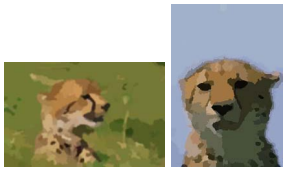
In this paper, we present a hierarchical approach to solve the problem of common pattern discovery, where the positions and radius perimeters of the common pattern are found. We use adaptive neighborhood and unary color similarity as a measure of node-to-node similarity. A many-to-many mapping technique is endorsed where EMD is used to perform similarity matching. We propose a novel local flow maximization approach to retrieve the location and best radius perimeter where the common pattern is located.

Experimental results shows that the algorithm is effective to detect even complicated common patterns, and is insensitive to segmentation, which is a hard problem especially for one-to-one matching techniques. We also show that the algorithm is robust to rotation, translation, shot angle and certain degree of occlusion. We do not consider objects at different scale, although the results shows that it can handle certain degree of scale difference in the common patterns. For all experiments, the sizes of the common patterns are actually different and can be handled successfully to certain extent. This is owing to the radius scale space specified by Sec 5.1 where the image is matched at various scale, and the best radius pair is selected in the process.

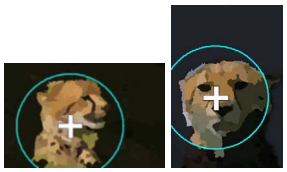
In our future works, we would improve our algorithm on scale invariant. We also plan to investigate ways to improve our adaptive neighborhood techniques so that the algorithm could cover a wider variety of images. For this paper, the background used are of different colors compared



(a) Original images



(b) Segmented images



(c) Results

Figure 7. Experiment Big Cat

to the common pattern.

Acknowledgements

The work described in this paper was fully supported by two grants from the Research Grants Council of the Hong Kong Special Administrative Region, China (CityU 1072/02E and CityU 118905).

References

[1] P. Felzenszwalb and D. Huttenlocher. Image segmentation using local variation. *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, pages 98–104, 1998.

[2] P. Hong and T. S. Huang. Spatial pattern discovery by learning a probabilistic parametric model from multiple attributed relational graph. *Discrete Applied Mathematics*, 139(1-3):113–135, April 2004.

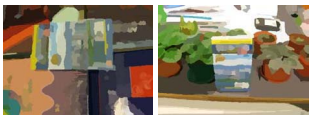
[3] H. Jiang and C. W. Ngo. Image mining using inexact maximum common subgraph of multiple arg. *International Conference on Visual Information System*, 2003.

[4] H. Jiang and C. W. Ngo. Graph based image matching. *International Conference on Pattern Recognition*, 2004.

[5] D. H. Kim, D. Yun, and S. U. Lee. A comparative study on attributed relational graph matching algorithms for perceptual 3-d shape descriptor in mpeg-7. *ACM Conf. on Multimedia*, 2004.

[6] A. Ratan, O. Maron, and T. Lozano-Perez. A framework for learning query concepts in image classification. *Int. Conf. Computer Vision and Pattern Recognition*, 1:429, 1999.

(a) Original images



(b) Segmented images

(c) Results

Figure 8. Experiment Map

(a) Original images

(b) Segmented images

(c) Results

Figure 9. Experiment Map

[7] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance, multi-dimensional scaling, and color-based image retrieval. *Proceedings of the ARPA Image Understanding Workshop*, pages 661–688, May 1997.

[8] Y. Rubner, C. Tomasi, and L. J. Guibas. The earth mover's distance as a metric for image retrieval. *International Journal of Computer Vision*, 40(2):99–121, November 2000.

[9] Y. Rubner, C. Tomasi, and L. J. Guibas. Edge, junction and corner detection using color distributions. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(11):1281–1295, November 2001.

[10] C. Schmid. Weakly supervised learning of visual models and its application to content-based retrieval. *Int. Journal of Computer Vision*, 56(1-2):7–16, January-February 2004.

[11] J. Sivic and A. Zisserman. Video data mining using configurations of viewpoint invariant regions. *Computer Vision and Pattern Recognition*, July 2004.