OBJECT RECOGNITION BY PARTIAL SHAPE MATCHING GUIDED SEARCH

Eli Saber[(1)], *Yaowu Xu*[(2)], *and A Murat Tekalp*[(3,4)]

1. Dept of Electrical Engineering, Rochester Institute of Technology, Rochester, NY 14623, USA
2. On2 Technologies, Inc., 21 Corporate Drive, Clifton Park, NY 12065, USA
3. Dept of Electrical and Computer Engineering, Univ. of Rochester, Rochester, NY 14627, USA
4. College of Engineering, Koc University, Sariyer, Istanbul, TURKEY

ABSTRACT

We propose a fast partial shape matching guided 2D object recognition algorithm to significantly accelerate recognition/matching of full (non-occluded) or partially occluded objects. The significant increase in speed comes from the fact that the search space is reduced to only those combinations of regions in the neighborhood of potential partial matches, as opposed to all combinations of regions as was done in our prior work [3]. Theoretical calculations and experimental results are provided to demonstrate effectiveness of the proposed algorithm on real images.

1. INTRODUCTION

Humans navigate through and retrieve samples from large image/video databases by means of semantic concepts, such as, objects, people, etc. However, most current multimedia systems can only process low level visual features, such as color, texture, shape, etc [1, 2] in an automatic fashion. "Learning by example" approaches have been proposed to automatically compute high-level semantic concepts from low-level visual features [3].

The concept of *learning by example* based on color and shape similarity for object based image labeling was proposed earlier by the authors [3]. "Learning" refers to storing combinations of regions, which correspond to meaningful objects, in the form of composite nodes with specific indexing information attached. As a result, subsequent searches using the same example template would immediately identify the composite node as a match using its shape and/or color attributes without re-processing its lower level nodes once again

The implementation of the learning process requires that we initially search all valid combinations of elementary regions (1 at a time, 2 at a time, etc.) in an image, as determined by the adjacency matrix, for shape and/or color similarity to a user provided example template. Evaluating shape similarity of all neighboring combinations of elementary regions to that of a user supplied object template is, in general, computationally expensive [4] for real life images whose segmentation possess a large number of homogeneous regions. This is because the order of the search grows exponentially with the number of elementary regions as will be demonstrated mathematically in Section IV. Hence, the need for effective guided search techniques that are capable of substantially reducing the initial search yielding practical results in near real time scenarios.

In this paper, we propose a partial matching guided full shape similarity search that is designed to only examine relevant combinations starting with the first combination which results in a successful partial match, and expand about its neighborhood instead of exploring all possible region combinations. To this effect, once an image region or a group of regions is identified to be a partial match to a given template, a priority will be given to the combination of the matched region with its immediate neighbors seeking the largest possible match. By doing so, unnecessary tests of region combinations can be avoided yielding significant improvements for the search speed of the full match.

The remainder of this paper is organized as follows. Section 2 reviews the concept of "learning by example" and a contour based shape similarity matching algorithm for partial object recognition proposed earlier in [5]. Section 3 introduces a search strategy of partial matching guided search. Section 4 analyzes the computational complexity of the proposed algorithm. Experimental results are presented in Section 5. Conclusions are drawn in Section 6.

2. BACKGROUND

2.1 Learning by Example

Here we provide a brief review of the concept of "*learning by examples*". We assume images are represented by a "scene graph" which consists of an object-region tree that indicates the parent-child relationships between high-level objects and low-level (elementary) image regions, and an adjacency matrix that captures the spatial relationships between these elementary regions. The root node of the tree corresponds to the whole image. Each leaf node (also called elementary node) represents a homogeneous image region with uniform color or texture. Elementary regions (nodes) are automatically constructed based on low-level color or texture segmentation, with each node denoting a uniform region [3]. Construction of composite nodes, which consist of groupings of elementary nodes or other composite nodes, is accomplished by learning from example(s). The implementation of the learning process requires searching all valid combinations of elementary regions. If a match is established, a composite node is formed containing the matching combination of elementary nodes. The composite node provides a level of semantic knowledge over and above the original scene graph containing only low-level nodes.

2.2 Partial Shape Matching

In [5], we propose a method for recognition of *partial* shape matches based on sub-matrix matching. The approach is comprised of two major steps: 1) feature extraction, and 2) similarity matching. Points on the contour with local maximum curvature are chosen as feature points and utilized to compute distance matrices for both the template and a candidate image region. The distance matrices serve to define correspondences between the template and the object under consideration using a sub-matrix matching algorithm. The correspondences are then employed to register the candidate region to the template resulting in a similarity measure between them. Finally, the similarity measure is utilized as a means to define and label the object semantically in a hierarchical segmentation fashion. The algorithm is invariant to translations, rotations in the image plane, reflections, and scale changes (a scale change introduces a scalar factor in the distance matrix). One important aspect of our technique is its low complexity, which is vital to its application in object-based image labeling. This was tested on many real life images yielding effective matches in cases of slight to severe occlusions.

3. PARTIAL MATCHING GUIDED SEARCH

We propose a partial matching guided full shape similarity search in images given a user specified template. The approach is designed to reduce the initial computational complexity of having to search all possible combinations for a suitable match. For a given image, the proposed search strategy can be summarized as follows:

1) Generate the initial content tree and adjacency matrix based on low-level image analysis [3].

2) Start with the largest elementary region (in terms of number of pixels) whose centroid is closest to the center of the image and examine it for a partial match.

3) If a partial match is found, expand about its neighborhood by considering combinations of neighboring regions until the largest matching combination of regions has been established. For occluded objects, the search stops once the visible portion of the object has been identified. For a non-occluded object, this typically results in a full match at a significant reduction in computational complexity. On the other hand, if an initial partial match is not found, then we proceed to examine other elementary regions and valid combinations of them sorted in decreasing size and increasing proximity to the center of the image until we find an initial partial match. Then, we repeat the steps described above to achieve a suitable match.

4) Once a match has been established with the largest possible group of regions, a composite node is formed capturing the "semantic" notion of the object in the image. As a result, subsequent searches can be done at a significant improvement in speed due to the fact that the top down hierarchy searching algorithms encounter a matching composite node prior to having to search combinations of its elementary nodes.

4. COMPLEXITY OF PARTIAL MATCHING GUIDED SEARCH

To quantify the benefits of our proposed algorithm, we provide an analysis of the computational complexity of the partial match guided search compared to the algorithm described in [3]. Assume we have an image with N elementary nodes in the initial content hierarchy where m of the N (m<=N) elementary nodes represent a given object of interest (i.e. a car is made up of m regions). In the case where each of the N regions is connected to all other regions (worst case scenario) in the adjacency matrix, the total number of valid combinations to be tested using the approach described in [3] is:

$$C = \sum_{k=1}^{N} \frac{N!}{k!(N-k)!} = \sum_{k=1}^{N} \binom{N}{k} = 2^N - 1$$

However, by utilizing the partial matching approach described above in an "ideal" sense where a partial match is identified at the singleton region level, then the number of combinations to be tested to find a full match that made up of the m neighboring regions is:

$$C_1' = N + (N-1) + \cdots + (N-m+1)$$

Once the full match is found, the m elementary nodes are grouped under a single composite node in the content hierarchy. Hence, the outstanding combinations to be tested are made up of the (N-m) remaining elementary nodes. For the worst case scenario, this is computed as:

$$C_2' = \sum_{k=1}^{N-m} \frac{(N-m)!}{k!(N-m-k)!}$$

Therefore, the total number of combinations that are tested for the image is:

$$C' = C_1' + C_2' = [N + (N-1) + \cdots + (N-m+1)] + \sum_{k=1}^{N-m} \frac{(N-m)!}{k!(N-m-k)!} \le mN + 2^{N-m}$$

where the computational complexity is reduced from $O(2^N)$ to $O(2^{N-m})$.

In a more realistic scenario, assuming the first candidate region that partially matches to the template contains L elementary nodes, the total combinations tested for the image will be:

$$C'' = \sum_{k=1}^{L} \frac{N!}{k!(N-k)!} + [(N-L) + \cdots + (N-m+1)] + \sum_{k=1}^{N-m} \frac{(N-m)!}{k!(N-m-k)!}$$

where the first and last summation terms represent, in this respect, the number of combinations tested prior to establishing an initial partial match; and the number of combinations to be tested after a composite node is formed for the full match.

For illustration purposes, the following table provides the numbers computed from the above equations for three example cases, where the reduction in combinations is clearly demonstrated.

	N=10, m=2, L=1	**N=10,m=5, L=3**	**N=15,m=10, L=3**
C	1023	1023	32767
C'	276	82	182
C''	276	214	669

For practical applications, the N regions found in a given image are not all connected to each other as assumed above. Hence, the total number of combinations to be examined is significantly smaller than what was portrayed above for the worst case scenario. This is clearly demonstrated in the results.

5. EXPERIMENT RESULTS

In this section, we demonstrate the effectiveness of the proposed algorithm in reducing the search and labeling complexity on images that contain full objects which are typically more computationally expensive than occluded objects. The images shown are of RGB type, where each color channel is quantized to 8 bits/pixel.

The advantage of utilizing the partial matching algorithm to perform a guided search for full match is illustrated in Figure 1. The proposed techniques in Section 3 are applied on a set of color car images that display no occlusions. A car template is utilized to search the images shown in Figure 1a. For each image, Figure 1b provides the corresponding segmentation maps. Figures 1c & 1d show the first partial match and the final match respectively.

Table 1 demonstrates the effectiveness of the algorithm in reducing the computational burden when searching real life images. In this table, we provide the number of regions in the segmentation map for the images found in Figure 1, where image 1 is the car shown on top. For each image, the table also provides, for a given template, a performance comparison between the initial search using the original scheme [3] and the partial matching guided search to full match. From the table, it is easily seen for the three images that the number of combinations examined using partial matching guided search for full match represents a significant reduction over the search without the use of partial matching. For all the images, the reduction of combinations to be tested is several orders of magnitude as seen from Table 1.

Table 1: Experiment results of partial matching guided search

Image name	Image 1	Image 2	Image 3
Number of Regions	14	12	21
Number of total valid region combinations tested without partial match guided search	4793	1363	50072
Number of total combination tested using partial match guided search for full match	193	92	6007

6. CONCLUSIONS

In this paper, we discussed the utilization of a partial matching search technique – invariant to translations, rotations, scale and reflections – to search for full non-occluded objects in real life images. The technique is effective in reducing the initial search time by focusing on the relevant neighboring regions instead of exploring all possible neighboring combinations once an initial partial match has been established. This was clearly demonstrated using theoretical calculations and experiment results. The technique was tested on a database of images containing a variety of objects such as cars, humans, outdoor scenes, etc. yielding similar results to the ones shown above. In each case, the object that matches the user specified template was found with significant savings in computational complexity. The above technique is incorporated into a system called "Voogle" designed to search images using visual queries.

REFERENCES

[1] Y. Rui, T. S. Huang, and S.-F. Chang, “Image retrieval: current techniques, promising directions and open issues,” Journal of Vis. Comm. & Image Rep., 10 (4) (1999) 39-62.
[2] A.W.M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, “Content-based image retrievals at the end of the early years,” IEEE Trans. on PAMI, 22 (12) (2000) 1349-1320.
[3] Y. Xu, E. Saber, and A. M. Tekalp, “Object Segmentation and Labeling by Learning from Examples,” *IEEE Trans. on Image Processing*, 12(6) (2003) 627-638.
[4] E. Saber and A. M. Tekalp, “Region-based shape matching for automatic image annotation and query by example,” J. Vis. Comm. Image Rep. 8 (1) (1997) 3-20.
[5] Y. Xu, E. Saber, A. M. Tekalp, “Shape matching of partially occluded objects for image retrieval using hierarchical content description”, Proc. Of SPIE 4315 (2000) 86-96.

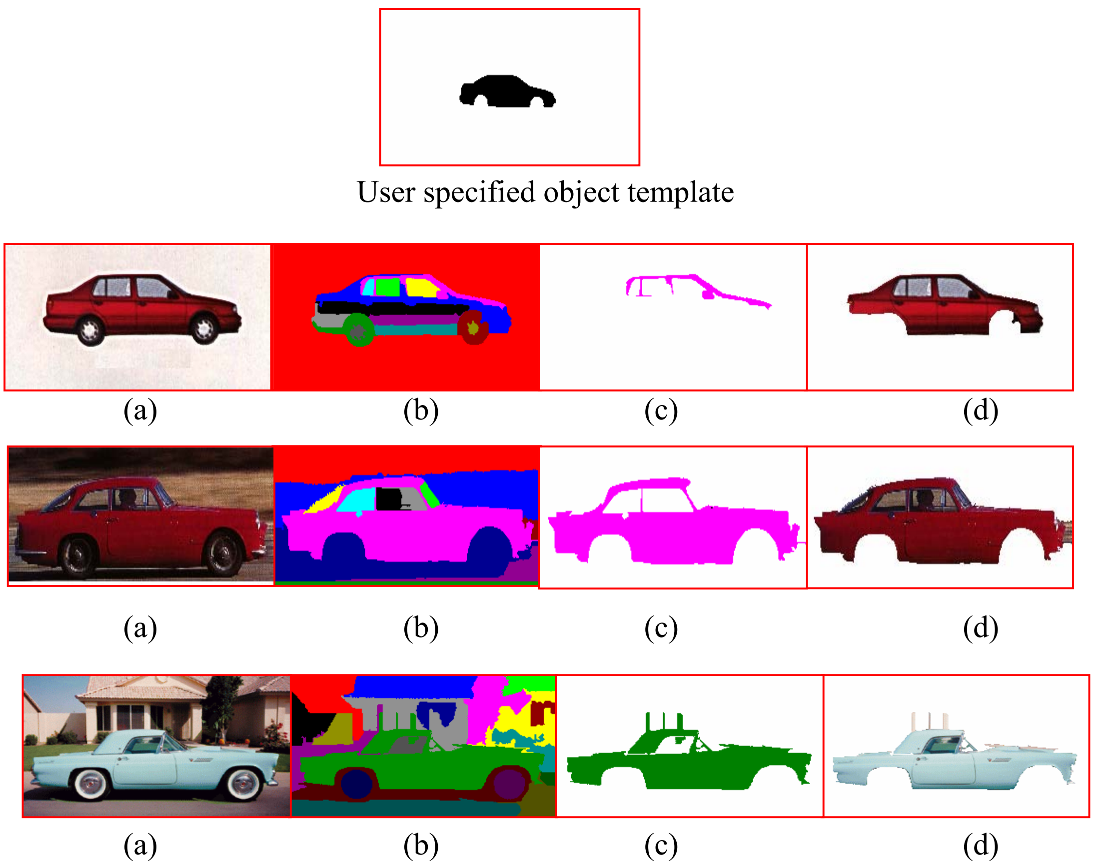

Figure 1 – Experimental results: a) Original Images, b) Segmentation maps; c) The first region found to be a partial match to the template shown above, and d) The combination of regions that represent a full match to the template.