ENERGY MATCHING BASED ON DEFORMABLE TEMPLATES

Youssef S. Tawfik[1], Ahmed M. Darwish[2] and Samir I. Shaheen[2]

[1]Electronic Research Institute, Tahrir St., Dokki, Giza, 12411 EGYPT

[2]Computer Engineering Department, Cairo University, Giza, 12613 EGYPT

ABSTRACT

Snakes, or active contours, have been previously used in computer vision applications to locate and identify objects. However, problems associated with initialization, poor convergence to boundary concavities and high computational complexity, have limited their utility. In this paper, we present a scheme for object localization and classification based on inexact description of the shape. The approach is based on deformable templates and was tested and proven be very robust with respect to changes in object scale, position and orientation as well as noise and local deformations of shape. A coarse-to-fine algorithm was developed to reduce the computational complexity and achieve efficient implementation. Results of applying the algorithm for automatic identification and localization of industrial parts will be presented.

1. INTRODUCTION

Tracking deformable objects in the plane is a well-known problem in the computer vision domain. Its application has found its way in various areas such as image segmentation, industrial part identification, character recognition and medical imaging [1]-[2]. The problem under investigation here is as follows: given a rough contour (template) which gives an inexact description of the shape of a 2D object of interest, locate and identify all the objects in the image which resemble this template. The basis of the approach is to project instances of the template into the image until one which is well supported by the observed data is found.

There are several implementation requirements to complete the problem formulation. First the template must have a relatively small number of controlling parameters. For any particular set of parameter values it must be possible to reconstruct a feasible instance of the object it represents. Second, an objective function must exist which can assess the evidence of object existence when a model instance is projected into an image. Finally, a method of optimization should be developed which can search the parameter space of the model in order to identify the particular set of parameter values for which the objective function is maximized.

Feature-based methods cannot be applied when the objects of interest have not been segmented from the background. The deformable template matching scheme, on the other hand, provides an appealing solution to the recognition tasks because of its capability to model an overall shape and meanwhile accommodate shape variations. However, the generality of the approach and avoidance of segmentation are achieved at the cost of higher computational complexity. To reduce computations, we have adopted a hierarchical recognition scheme. In the first screening stage, the image is segmented and searched using some simple and efficient matching criteria. In the second stage, the deformable template matching is applied to the small set of choices obtained from the first stage. This hierarchical mechanism can improve both efficiency and accuracy of recognition.

Aside from this introductory section, the remainder of this paper is composed of five more sections. In section 2 we give a quick survey of similar approaches. Section 3 presents the necessary background for algorithm implementation, for which details are given is section 4. Samples of the results are presented in section 5. We finally conclude in section 6 with observations and recommendations for future work.

2. LITERATURE SURVEY

Dynamic contour models started to get more exposure after the introduction of the snake model [1]. In this approach, an energy-minimizing spline, called a "snake", is controlled by a combination of the three factors; the internal spline force which enforces the smoothness, the image force which attracts the spline to the desired features, and the external constraint force. Each force creates its own potential field and the spline actively adjusts its position to the shape until it reaches a local minimum of the potential energy.

This snake model provides a powerful interactive tool for image segmentation. However, the implementation of the original snake is vulnerable to image noise and the initial position. Several provisions have been made in the literature to improve the robustness and stability of the snakes. For example, a "balloon force" was introduced in [3], which can either inflate or deflate the contour. This force helps the snake to trespass spurious isolated weak image edges, and counters its tendency to shrink. The resulting snake is more robust to the initial position and image noise, but human intervention is needed to decide whether an inflationary or deflationary force is needed. Later, the use of dynamic programming was suggested, [4] & [5], to minimize the energy function. These methods exhaustively search all the admissible solutions, and each iteration results in a locally optimum contour. As a result, the method is guaranteed to converge in a finite number of iterations. This idea of active contour has been extended to perform other tasks subjective contour detection, motion tracking and stereo matching.

Deformable templates have been used to extract arbitrary objects directly from noisy images, [6] & [7]. The method combines a stable, invariant and unique contour model with Markov random field to yield prior distribution that exerts influence over an arbitrary global model while allowing for deformation. Under the

Bayesian framework, contour extraction turns into posterior estimation, which is in turn equivalent to energy minimization in a generalized active contour model.

Polygonal templates have been also used, [8], to characterize general models objects. A priori probability distribution was set to constrain the template to be deformed within a set of allowed shapes of a typical object. The likelihood function was, then, constructed based on motion information and edge directionality so that the deformed template is contained in the motion area and its boundary coincides with salient edges in the input image sequence. This approach was used to segment vehicles in traffic image sequences [8].

Other interesting approaches and applications could be found in [9]-[11]

3. BACKGROUND

The idea of representing the contour as a vector containing an ordered set of points, $V = [\, v_1, v_2, \ldots, v_n \,]$ is not new, where each v_i is defined on the finite grid:

$$v \in E = \{(x, y) : x, y = 1,2,..., M\} \qquad (1)$$

For modeling of highly variable but locally predictable contours, each $u_i = v_i - g$ is expressed as a linear combination of its two neighboring points:

$$u_i = \alpha_i u_{i\alpha} + \beta_i u_{i\beta} \qquad (2)$$

where the basis indices are given by:

$$i_\alpha = \begin{cases} i-1; & i>1 \\ 3; & i=1 \end{cases} \qquad i_\beta = \begin{cases} i+1; & i<n \\ n-2; & i=n \end{cases} \qquad (3)$$

and g is an arbitrary reference point.

The use of such model to represent the characteristic salient curves of the object makes the modeling very general and flexible.

The use of potential functions to influence template deformations towards image features is akin to the work in [1]. However, our work is different in that we relate the potential to the nearest input edge pixels through the use of a Chamfer distance to edge points, and this potential is further modified based on the angle between the contour points and the closest point on the image edge.

We have also adopted a multi-resolution coarse-to-fine algorithm to automatically locate objects of interest in a given image. This greatly reduces the amount of computations and makes real-time implementations feasible.

It is necessary to build background about two techniques that we use in cascade in the algorithm. These are invariant moments and deformable contours. As for invariant moments we use the first seven. For space restrictions, readers are referred to [12] for detailed equations. The remainder of this section will concentrate on deformable contours.

A deformable contour is a planar curve, which has an initial position, and an objective function associated with it. A special class of deformable contours called snakes was introduced in [1] in which the initial position is specified interactively by the user and the objective function is referred to as the energy of the snake. This energy of the snake (E_{snake}) is expressed as:

$$E_{snake} = E_{internal} + E_{external} \qquad (4)$$

The internal energy term imposes a regularization constraint on the contour as follows:

$$E_{internal} = \int_s (w_1(s)\|v'(s)\|^2 + w_2(s)\|v''(s)\|^2)ds \qquad (5)$$

where s is arc length, w_1 and w_2 are weighting parameters that control the snake's tension and rigidity, respectively, and $v(s)$ stands for the ordered pair $(x(s),y(s))$ which denotes a point along the contour.

The external energy term is responsible for attracting the snake to interesting features in the image, such as boundaries. The exact expression for $E_{external}$ depends on the characteristics of the features of interest.

Finding local minima for E_{snake} corresponds to solving the following Euler-Lagrange equation for v:

$$-(w_1 v')' + (w_2 v'')'' + \nabla E_{external}(v) = 0 \qquad (6)$$

To find a solution to (6), the snake is made dynamic by treating v as function of time t as well as s, i.e., $v(s,t)$. Then, the partial derivative of v with respect to t is then set equal the left-hand side of (6) as follows:

$$v_t(s,t) = w_1 v''(s,t) - w_2 v''''(s,t) - \nabla E_{external}(v) \qquad (7)$$

When the solution $v(s,t)$ stabilizes, the term $v_t(s,t)$ vanishes and we achieve a solution of (6).

4. ALGORITHM

The algorithm for object localization and classification is summarized as follows:

Step 1. Preprocess the image and calculate the edge map and gradient direction (we used the Least Square method) at two different resolutions (coarse and fine).

Step 2. Perform the coarse-level matching. Use a contour following procedure to recover objects' boundaries. For each of the detected objects, the matching process is as follows:

- Calculate the object's invariant moments.
- For each template in the database, compare the precomputed invariant moments to those of the detected object. The distances are sorted and the top N templates are marked.

Step 3. Perform fine level matching using the deformable template approach. This step is initialized by candidate templates generated by the coarse-level search of step 2. The templates are projected on the image at the position, scale, and orientation determined by the objects' features.

Step 4. The energy of the snake is minimized using dynamic programming. The object is labeled to belong to the class of the lowest energy snake.

5. EXPERIMENTAL RESULTS

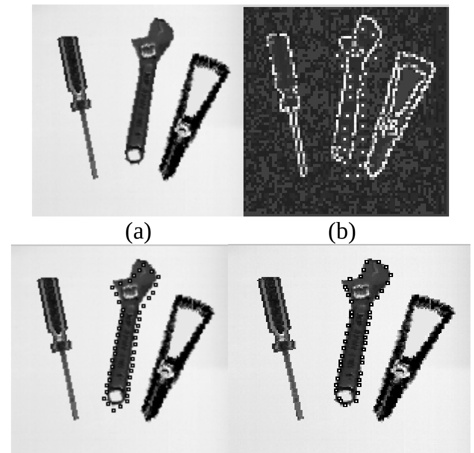
The system was tested using a total of 10 objects. Fig. 1 shows the objects used to in different images. In these experiments, both the presence and the number of desired objects in image, their position, pose and scale are unknown. Two modes of operation were tested. The first is object localization given a query template. The system should report all objects similar to the template along with its pose (orientation) and scale. The second is the query mode where the system should report what are the objects in the image along with pose and scale.

Several intermediate steps of the applying the algorithm are shown in Fig. 2. The image is segmented at the coarse level, using a contour following procedure. The query contour is compared to the segmented objects based on invariant moment shape features. Only few objects are passed to the next level to initialize the matching at a higher accuracy. At the finer level, the deformable template matching is applied. Localization is reported where the objective function has the smallest value.

Fig. 3 illustrates that system can localize objects independent of their location, and orientation in the image. Objects of different shapes are retrieved using different prototype templates. The figurer shows the localization of a screwdriver and a wire cutter separately.

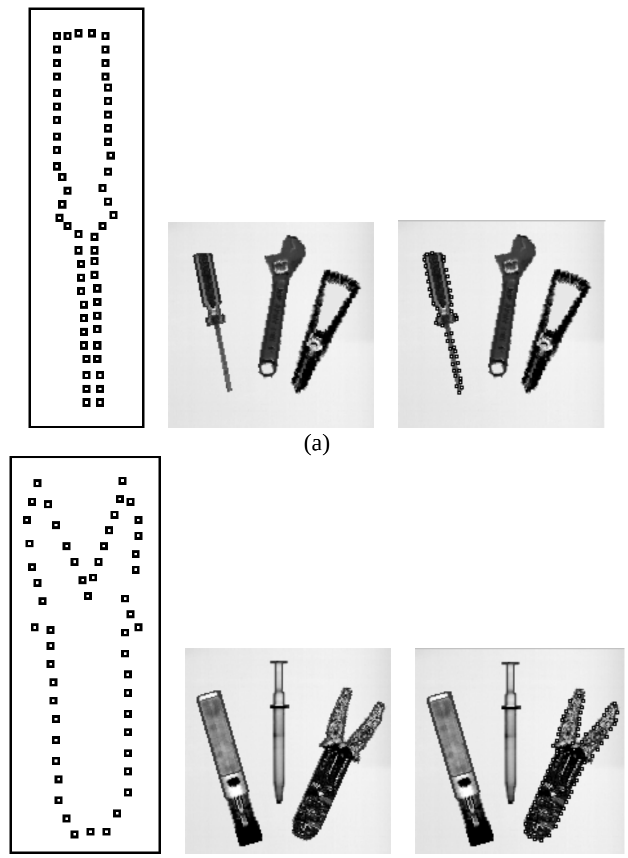
In Fig. 4 and 5 it is demonstrated that the system is able to retrieve all the objects in an input image that resemble the prototype template even if they are of different sizes, different orientations, and have local variations. In Fig. 5 it is clear that all pliers were located correctly despite sever deformation of one of them and that the wire cutter was not mistaken as a pliers.

Figure 1. Objects used to test the system



(c) (d)

Figure 2. Object localization using multiscale. (a) Image at coarse scale. (b) Edge magnitude with initial contour. (c) Intermediate result at fine scale. (d) Final result.



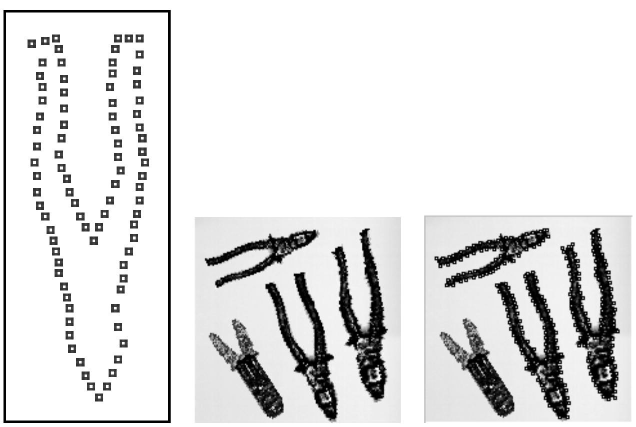
(b)

Figure 3. Automatic localization of desired objects. (a) Retrieval of a screwdriver using multiresolution deformable template matching. (b) Retrieval of a wire cutter using multiresolution deformable template matching. (From left to right: template, input image, retrieved deformed template.)

An example of query experiments is shown in Fig. 6. In this case no specific template is supplied. It is required to classify and label all objects in the image. In the first stage, all templates were applied and invariant features method was used to select the top best categories. The deformable template method was then applied to refine the classification. The system was able to classify all the objects in the image correctly.

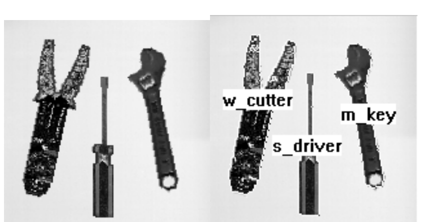
(a) (b) (c)

Figure 4. Detecting multiple objects with different scale, orientation, and local deformation. (a) Template. (b) Input image. (c) Detected objects.



(a) (b) (c)

Figure 5. Detecting multiple objects with different scale, orientation, and local deformation. (a) Template. (b) Input image. (c) Detected objects.



(a) (b)

Figure 6. Classifying objects. (a) Input image. (b) Correctly classified objects.

6. CONCLUSION

We have presented a complete, fast and general object localization and classification system. The image was first segmented so that the physical boundaries of the objects could be recovered. A two stage, coarse-to-fine matching algorithm was used to correctly classify each object. The first stage used Invariant moments to prune the search space. The second stage used deformable template matching algorithm to find the best-matching one of the surviving templates. The system was found to be accurate with respect to reporting position, orientation and scale of the objects as well as being robust against noise and local shape distortions. The use of deformable template matching is usually avoided because of its high computational complexity. A major contribution is the fast efficient implementation of the approach with the introduction of the fine coarse matching. Our future work plans includes looking at generalizing the approach to cases where portion of the object is occluded.

REFERENCES

[1] Kass, M., Witkin, A. and Terzopoulos, D., “Snake: Active Contour Models”, *International Journal of Computer Vision*, 1:321-331, 1987.

[2] Leymarie, F. and Levine, M., “Tracking deformable objects in the plane using an active contour model”, *IEEE Trans. on Pattern Anal. Machine Intell*., 17(2):158-175, 1995.

[3] Cohen, L and Cohen, I., “Finite-element Methods for Active Contour Models and Balloons for 2-D and 3-D Images,” *IEEE Trans. Patt. Anal. and Mach. Intell*., 15(11):1131-1147, 1993.

[4] Amini, A., Tehrani, S. and Weymouth, T., “Using Dynamic Programming for Minimizing the Energy of Active Conoturs in the Presence of Hard Constraints,” *Proc. Second Int. Conf. Computer Vision*, Tarpon Springs, FL, pp. 95-99, Dec. 1988.

[5] Geiger, D., Gupta, A., Costa, L. and Vlontzos, J., “Dynamic Programming for Detecting, Tracking, and Matching Deformable Contours”, *IEEE Trans. on Pattern Anal. Machine Intell*., 17(3):294-302, 1995.

[6] Lai, K. and Chin, R., “Deformable Contours: Modeling and Extraction”, *IEEE Trans, on Pattern Anal. Machine Intell.*, 17(11):1084-1090, 1995.

[7] Lai, K. and Chin, R., “On Regularization, Formulation and Initialization of the Active Contour Models (Snakes)”, *Assian Conf. Computer Vision*, 1993, pp. 542-545.

[8] Dubuisson M., Lakshmanan, J. and Jain, A., “Vehicle Segmentation Using Deformable Templates,” *IEEE Trans. Pattern. Anal. and Machine Intell*., 18(3):293-308, 1996.

[9] Nakagawa, S., Watanabe, T. and Kuno, Y., “A framework for feature extraction of images by energy minimization”, *IEICE Trans. on Info. and Syst*, E77-D(11):1213-1218, Nov. 1994.

[10] Tsai, C., Sun, Y. and Chung, P., “Minimizing the energy of active contour model using a Hopfield network”, *IEE Proceedings E (Computers and Digital Techniques)*, 140(6):297-303, Nov. 1993.

[11] Herpers, R. and Sommer, G., “Discrimination of facial regions biased on dynamic grids of point representations”, *International Journal of Pattern Recognition and Artificial Intelligence*, 12(4):381-405, June 1998.

[12] Jain, A, *Fundamentals of Digital Image Processing*, Prentice Hall 1989.