Seventh International Conference on Control, Automation, Robotics And Vision (ICARCV'02), Dec 2002, Singapore

A Novel Quadrilateral-based Tracking Method

H. Y. Chung, N. H. C. Yung & P. Y. S. Cheung

Department of Electrical & Electronic Engineering
The University of Hong Kong
Pokfulam Road, Hong Kong SAR, China
Email: hychung@eee.hku.hk

ABSTRACT

This paper proposes a novel tracking method based on quadrilateral-based segmentation. The tracking method matches quadrilaterals in a region with those in a reference frame to construct region correspondences, from which trajectory of each region can be obtained. The Table Tennis sequence has been used to evaluate the tracking method. Experiment results show that the proposed tracking method can track the ping-pong ball and the racket reasonably well over a series of images, indicating matching quadrilaterals to track region is a viable approach.

1 INTRODUCTION

Object tracking is an essential task for several application areas including vision-based control, human-computer interfaces, surveillance and medical imaging. Object tracking involves two main steps: 1. Segmenting an image into distinct objects. 2. Tracking target regions in successive frames. These two steps, however, are non-trivial tasks, attracting a lot of researchers' attention. While considerable research and progress have been made, the robustness and generality of segmentation and tracking on a variety of image data have not been fully established. First of all, segmentation itself has already been a challenging task over decades. Traditional segmentation methods can segment images reasonably well into distinct regions or boundaries. However, these regions do not necessarily favor object tracking as those regions usually needed to be further interpreted to produce sufficient information for tracking. Secondly, tracking target regions between successive frames is difficult as objects may undergo various transformations such as rotations, scaling, translations and deformations.

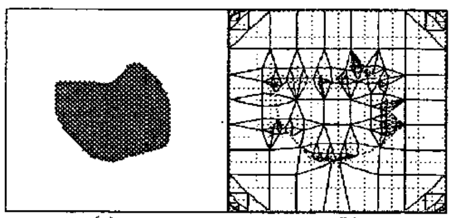
At present, there are two main tracking approaches: Model-based and Model-less tracking. In model-based tracking, either static [1-2] or dynamic [3-4] object model can be used to track specific targets. However, this requires the target objects to be known in advance and is highly application dependent as object models differ from one application to another. Model-less tracking [5-10], on the other hand, does not rely on object model to find correspondences between objects being tracked. Instead, it constructs correspondences between objects from motion information or generic object features like color, histogram, texture, etc. The advantage of this approach is that different kinds of object can be tracked and thus it is less application dependent. However, since less information is used to track objects, correspondences between objects may not be computed correctly and therefore, ambiguous correspondences may arise. To enable effective tracking, segmentation and tracking must be considered as integrated task where the purpose of segmentation is directed to making matching more successful, rather than simply resulting in a number of regions.

In this paper, we present a model-less tracking method based on quadrilateral-based segmentation [11]. Quadrilateral-based segmentation is used because of two reasons: First, quadrilateral representation offers a large data reduction similar to polygon approximation techniques except that it is far more flexible with lower approximation error. Second, inter-frame tracking of regions can be done by matching quadrilaterals [12-14] across frames. In principle, it employs a group of connected quadrilaterals, extracted from an edge-map, to represent a region. Model-less tracking is used because we are aiming at generic object tracking that is less application dependent. Inter-frame tracking of regions is based on matching of quadrilaterals, which is done by finding a quadrilateral in a search area in the reference frame that is most similar to the current quadrilateral. The region in the reference frame with largest number of quadrilaterals matched with the region in the current frame is considered to be the matched region. Based on the 'table tennis' sequence, the proposed tracking method was evaluated. It is found that our method is able to track both the ping pong ball and the racket, as well as other regions in the image.

This paper is organized as follows. Section 2 details our proposed quadrilateral based tracking method, section 3 presents tracking results while section 4 concludes the whole paper.

2 PROPOSED TRACKING METHOD

As we are aiming at generic object tracking that is less application dependent, our approach is based on model-less tracking. Since we do not require an accurate model of an object, it can be represented as a group of regions instead of a set of features. To make tracking possible, we have to extract the regions first, using the quadrilateral-based segmentation proposed in [11]. The segmentation method is built upon a network of quadrilaterals to represent regions. Each region is completely described by a set of quadrilaterals, which approximates the region boundary. The following diagram shows an example of using quadrilaterals to approximate object boundaries with object boundaries shown in thick gray lines.



(a) (b)

Fig. 1: Quadrilateral approximation result (a) Object (b) Boundaries approximated

2.1 OVERVIEW

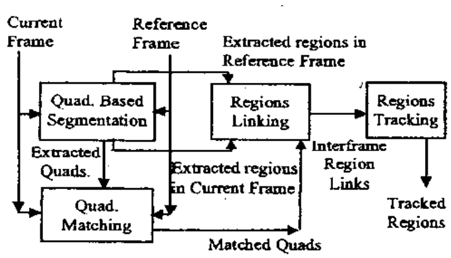


Fig. 2: Tracking method overview

Fig. 2 shows the block diagram of the tracking method proposed. To track regions between two frames, the current frame and the reference frame are first segmented into regions by quadrilateral-based segmentation method. Right after that, regions extracted in the two frames would be represented by a group of quadrilaterals. Inter-frame quadrilateral matching will then be carried out to match the quadrilaterals across frames.

2.2 MATCHING OF QUADRILATERALS

To do this, each quadrilateral in the current frame is first projected orthographically onto the reference frame. A rectangular search area is then built upon the centroid of the projected quadrilateral as shown in Fig. 3. The quadrilateral in the current frame is then matched with the quadrilaterals within the search area in the reference frame. The matching criteria can be shape, color, texture, geometric invariants [12-13], etc. In our proposed method, color is considered, i.e., the quadrilateral with color most similar to the quadrilateral to be matched would be chosen.

Let $S(q)$ be the set of quadrilaterals in the search area built upon the centroid of the projected quadrilateral q,
$R(q)$ be the average red color component of the quadrilateral q,
$G(q)$ be the average green color component of the quadrilateral q,
$B(q)$ be the average blue color component of the quadrilateral q.

Denote the color difference between two quadrilaterals p and q as

$$D(p,q) = [R(p)-R(q)]^2+[G(p)-G(q)]^2+[B(p)-B(q)]^2;$$

Then, for each quadrilateral q in the current frame, the matched quadrilateral r is:

$$r = \arg\min_{p} \min_{p\in S(q)} D(p,q)$$

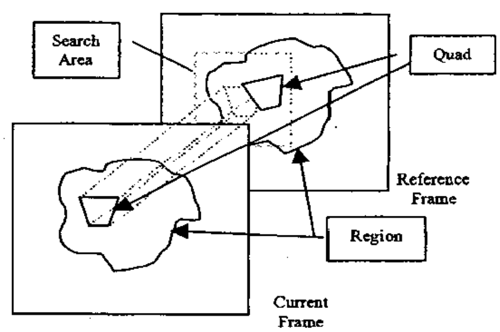


Fig. 3: Matching of quadrilaterals across frames.

After each quadrilateral in the current frame is matched, Inter-frame Region Links (region links across two frames) can be built. The Inter-frame Region Links are then used to perform the region tracking.

2.3 REGION LINKING

Let us define the following:

$Q_c(l)$ – The set of quadrilaterals in current frame within the region with region label l;

$Q_r(l)$ – The set of quadrilaterals in reference frame within the region with region label l ;
$M(q)$ – Quadrilateral in the reference frame that matches the quadrilateral q in the current frame.
$L_c(q)$ – Region label of the quadrilateral q in the current frame ;
$L_r(q)$ – Region label of the quadrilateral q in the reference frame ;
$T(l)$ – Region label in reference frame that matches the region in current frame with region label l ;

Region Linking is used to match regions between two frames. Essentially, region linking is to determine the $T(l)$ (Inter-frame Region Links), which is derived as follows:

Let $R(l, l') = \{ q : L_c(q) = l, L_r(M(q)) = l', q \bullet Q_c(l)\}$

Then, $T(l) = \arg\max_{l'}\{\|R(l,l')\|\}$.

With $T(l)$, each region in the current frame will be associated with a region in the reference frame. By using this information, each region in the current frame can be tracked.

2.4 REGION TRACKING

After inter-frame region links have been obtained, region tracking is performed. The inter-frame region links are used to determine which region(s) in the reference frame correspond to the regions in the current frame. The challenge of this is that we have to determine whether or not a region requires splitting or merging. At present, the proposed method employs simple tracking rule (one-to-one) mapping. That is, a region in the current frame corresponds to exactly one region in the reference frame.

Consider an n frame video sequence, let the inter-frame region link with the current frame c and reference frame r be $T_{c,r}(\bullet)$ where $c, r \in \{1, 2, \dots n\}$.

Then, for each region l_1 in frame 1, the corresponding region l_k in frame k, where $k \in \{1, 2, \dots n\}$ can be obtained as follows:

$l_2 = T_{1,2}(l_1)$
$l_3 = T_{2,3}(l_2)$
....
$l_k = T_{k-1,k}(l_{k-1})$

which results
$$l_k = T_{k-1,k} \circ T_{k-2,k-1} \circ T_{k-3,k-2} \circ \cdots \circ T_{2,3} \circ T_{1,2}(l_1)$$

In our current work, as shown in above, only the inter-frame region links between consecutive frames is considered. This simple tracking rule may not be able to deal with cases like occlusion, birth and death of regions, etc. More sophisticated tracking rule will be needed to cater for the aforementioned situations and this will be one of our future directions.

3 TRACKING RESULTS

The first 40 frames of the 'table tennis' sequence have been used to see how well the tracking technique performs. The ping pong ball and the racket are the targets to be tracked in the table tennis sequence. It is found that the proposed method can track the ping pong ball and the racket in all the 40 frames. The following shows the boundaries of the tennis ball and the racket extracted in the first frame.

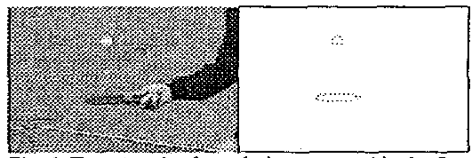
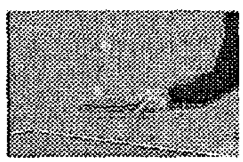
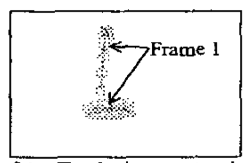


Fig. 4: Target region boundaries extracted in the first frame of "Table Tennis"

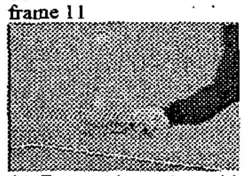
The following shows the tracked targets. The tracked targets are presented in 4 sequence segments: Frames 1 to 11, 11 to 21, 21 to 31 and 31 to 40. The beginning and the ending frame in each sequence segment are overlaid to show the corresponding positions of each target (as shown in (a), (c), (e) and (g)). The trajectory of the center of mass and boundary of each target (as shown in (b), (d), (f), (h)) are drawn with solid and gray lines respectively.



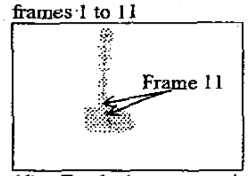
(a) Frame 1 overlay with frame 11



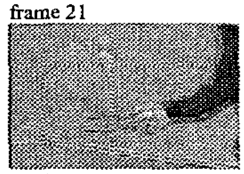
(b) Tracked targets in frames 1 to 11



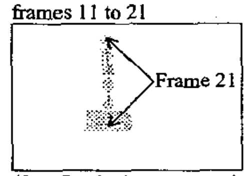
(c) Frame 11 overlay with frame 21



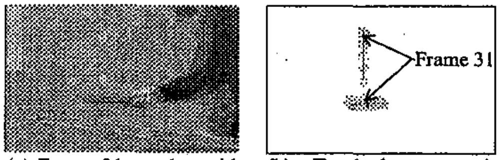
(d) Tracked targets in frames 11 to 21



(e) Frame 21 overlay with frame 31



(f) Tracked targets in frames 21 to 31



(g) Frame 31 overlay with frame 40 (h) Tracked targets in frames 31 to 40

Fig. 5: Tracked targets in frames 1 to 40

As shown in the above diagrams, the ping pong ball and the racket can be tracked but the boundary of the ping pong ball is not very exact. This is due to the fact that the ball is not homogenous in color due to the ambient illumination effect. This indicates that the segmentation approach can be further improved. On the other hand, even with imperfect segmentation, tracking is still viable.

4 CONCLUSION

To conclude, we have presented a novel tracking method which is built upon quadrilateral-based segmentation. Although the tracking method is far from perfect, it shows that tracking based on quadrilateral presents a feasible research direction. Further research effort is needed to improve the proposed tracking method. Future directions would be focused on two things: 1. Improve the quadrilateral matching by considering more attributes of quadrilaterals instead of solely rely on color, which is sensitive to intensity variation between frames. 2. Improve Region Linking so that more region correspondences can be obtained to deal with the cases like occlusion, birth and death of regions, etc.

REFERENCES

[1] H. Wang and M. Brady, "Real-time corner detection algorithm for motion estimation," *Image Vision Comput.*, vol. 13, Nov. 1995, pp.695-703.

[2] D. K. Park, H. S. Yoon and C. S. Won, "Fast object tracking in digital video," *ICCE*2000, 2000 , pp.356-357.

[3] S. Malassiotis and M.G. Strintzis, "Tracking textured deformable objects using a finite-element mesh," *IEEE Trans. CSVT*, vol. 8, no. 6, Oct. 1998, pp.756-774.

[4] Y. Zhong, A. K. Jain and M. P. Dubuisson-Jolly, "Object tracking using deformable templates," *IEEE Trans. PAMI*, vol. 22, no. 5, May 2000, pp.544-549.

[5] N. K. Paragios and R. Deriche, "A PDE-based level-set approach for detection and tracking of moving objects," *ICCV98*, pp.1139-1145.

[6] E. Oron, "Motion estimation and image difference for multi-object tracking," *IEEE Proc. Aerospace Conference 1999*, vol. 4, 1999, pp.401-409.

[7] Yining Deng and B. S. Manjunath, "Spatio-temporal relationships and video object extraction," Conference Record of the Thirty-Second Asilomar Conference on Signals, Systems & Computers, vol. 1, 1998, pp.895-899.

[8] L. Favalli, A. Mecocci and F. Moschetti, "Object tracking for retrival applications in MPEG-2," *IEEE Trans. CSVT*, vol. 10, no. 3, Apr. 2000, pp.427-432.

[9] F. Bremond and M. Thonnat, "Tracking multiple nonrigid objects in video sequences," *IEEE Trans. CSVT*, vol. 8, no. 5 Sept. 1998, pp.585-591.

[10] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Trans. PAMI*, Mar. 2000, pp.266-280.

[11] N. H. C. Yung, H. Y. Chung and P. Y. S. Cheung (2002), "Quadrilateral based region segmentation for tracking", to be published in *Optical Engineering – the Journal of SPIE*, Nov. 2002.

[12] U. Uenohara and T. Kanade, "Geometric invariants for verification in 3-D object tracking," *IEEE Proc. IROS96*, 1996, pp.785-790.

[13] S. W. Lee, B. J. You, G. D. Hager, "Model-based 3-D object tracking using projective invariance," *IEEE Proc. Robotics & Automation*, May 1999, pp.1589-1594.

[14] G. Lei, "Recognition of planar objects in 3-D space from single perspective views using cross ratio," *IEEE Trans. Robotics and Automation*, vol. 6, no. 4, Aug 1990, pp.432-437.