Real-Time Salient Closed Boundary Tracking via Line Segments Perceptual Grouping

Xuebin Qin, Shida He, Camilo Perez Quintero, Abhineet Singh, Masood Dehghan and Martin Jagersand

***Abstract*— This paper presents a novel real-time method for tracking salient closed boundaries from video image sequences. This method operates on a set of straight line segments that are produced by line detection. The tracking scheme is coherently integrated into a perceptual grouping framework in which the visual tracking problem is tackled by identifying a subset of these line segments and connecting them sequentially to form a closed boundary with the largest saliency and a certain similarity to the previous one. Specifically, we define a new tracking criteria which combines a grouping cost and an area similarity constraint. This criteria makes the resulting boundary tracking more robust to local minima. To achieve real-time tracking performance, we use Delaunay Triangulation to build a graph model with the detected line segments and then reduce the tracking problem to finding the optimal cycle in this graph. This is solved by our newly proposed closed boundary candidates searching algorithm called "Bidirectional Shortest Path (BDSP)". The efficiency and robustness of the proposed method are tested on real video sequences as well as during a robot arm pouring experiment.**

I. INTRODUCTION

Closed contour boundaries are common elements in real world scenes. Hence, real-time closed boundary tracking is important in robot vision. As an example, consider Fig. 1, where the rim contour of a bowl is tracked. In real-world indoor images and robot applications, conventional trackers can fail, e.g. in the presence of texture-less target regions, non-rigid deformations, non-Lambertian surfaces, changing lighting conditions, clustered background and drastic contents changing of supportive regions.

Template-based high DOF trackers can estimate image transformations such as affine and homography of a planar object (or contour) region from one frame to the next [1]. Many template trackers have been proposed: Lucas and Kanade's iterative early image registration technique[2], methods using different appearance models, such as Normalized Cross Correlation (NCC) [3], Mutual Information (MI) [4], edges [5], [6] and hybrid of edge and texture information [7]. Different search methods include [8], efficient second-order minimization (ESM) [9], nearest neighbour [10], particle filter [11] and RANSAC based cascade method (RKLT) [12] have been proposed to achieve better performance. However, they are highly dependent on stable textures and sensitive to the presence of local minima. Keypoints based approaches [13], [14] are relatively robust to local minima, but they require many accurate feature points to be detected that can be difficult to achieve in practice.

Non-planar contours tracking can be approached as a pose estimation problem [15], [16] when 3D models are available. In unstructured, natural environments 3D models are seldom available. In these cases non-planar contours tracking can be approached as nonrigid tracking. Godec et al. propose a tracking-by-detection approach (HoughTrack) based on the generalized Hough-transform to track non-rigid objects [17]. HoughTrack's voting based detection and back-projection are coupled with a rough segmentation-based on graph-cuts. However, this class of methods do not work well in tracking targets whose appearances changes significantly or targets which are comprised of several regions with significant differences. Another class of non-rigid trackers are snakes or active contours [18], [19], [20], [21], [22], [23]. Given an initial contour from the previous frame, Contour-based tracking is performed by searching the target contour based on minimizing a suitable energy [1]. However, these methods are more likely to trap in local minima in cluttered environments [24].

Discrete edge fragment perceptual grouping methods are more robust to local minima. Elder and Zucker [25] define boundary grouping cost as the product of fragment/gap prominence along the boundary and then find the optimal cycle by the shortest-path algorithm [26]. Wang et al. [27] encode Gestalt laws [28] of closed bound-

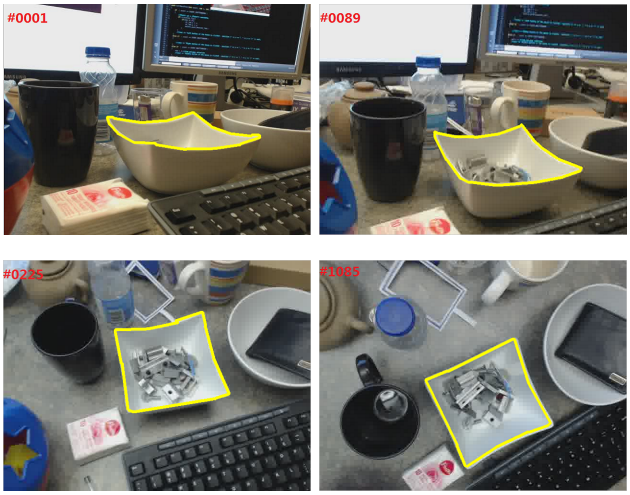
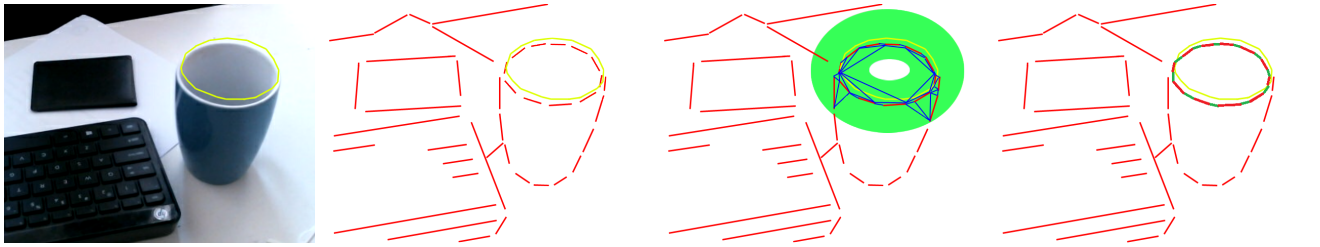


Fig. 1. Tracking the rim boundary of a bowl with the following characteristics: (1) the rim of the bowl is non-planar, (2) the bowl itself has no salient textures, (3) the viewpoint changes dramatically.

Preprint submitted to 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Received March 1, 2017.



(a) Current image and the tracked boundary (B_p) from the last frame (yellow polygon)

(b) Detected line segments

(c) Gap filling among line segments in the buffer (green) region of boundary B_p

(d) Tracked boundary (B_c) of the current frame

Fig. 2. Illustration of closed boundary tracking

aries into the fraction of the sum of gap pixels and curvatures with respect to the boundary length and optimize it by a minimum weights perfect matching algorithm [29]. In [30], Stahl and Wang substitute the closed boundary grouping cost in [27] for a ratio between the gap length along the boundary and the area enclosed by the boundary. These methods are designed for extracting salient closed boundaries from single images. Schoenemann and Cremers [31] perform contour tracking by integrating elastic shape matching and image segmentation into one solvable optimization problem. Unfortunately, this is not real-time without using GPU processing, ruling out light weight or low power robotics applications. In [32], rough contour tracking and shape context matching are performed separately to improve time efficiency and handle clustered background. However the use of a fixed shape template restricts the ability to track non-planar closed boundaries with out-of-plane motion.

Despite all different attempts, tracking of boundary targets in unstructured environment is still a challenging problem. This paper address this problem by presenting a novel line segment grouping based method for tracking salient closed boundaries. Our key contributions are:

- We define a salient closed boundary tracking criteria by combining a boundary grouping cost [30] and a regularization constraint on the contour's area variation. These criteria have a preference to track the structure with the smallest grouping cost and a similar area to the previous boundary, making the tracking method robust to noise.
- Our combinatorial tracking criteria are difficult to optimize in real-time with existing algorithms. To search the closed boundary that satisfies this criteria in incoming frames, we use Delaunay Triangulation to build a graph model $G(V, E)$ which is comprised of detected line segments, their endpoints and in-between gaps. Based on this graph, we reduce the tracking problem to a problem of searching for the optimal cycle according to the proposed criteria. Specifically, we propose a novel real-time searching method called bidirectional shortest path searching algorithm (BDSP) for optimal cycle searching.
- To evaluate the performance of our tracking scheme, we collected and annotated nine video sequences (9598 frames) of typical contours, which are challenging to track in real robot applications.

We implement our tracking method[1] and test it on our newly collected real world video sequences[2], and compare it against common robot vision trackers: RKLT [12], ESM [9], HoughTrack [17] and a tracker adapted from the contour grouping method RRC [30]. In addition, we also test the behavior of our method during a real robot arm experiment following a moving bowl and pouring cereal into it.

The reminder of this paper is organized as follows. Section II formulates the problem by introducing the tracking criteria that combines a previously proposed grouping cost and an area constraint. Section III presents the details of the graph modeling and the graph based optimization algorithm that is used for solving the tracking problem. Section IV describes experimental results on the newly built real world dataset and an application of robot arm pouring. A brief conclusion is given in Section V.

II. PROBLEM FORMULATION

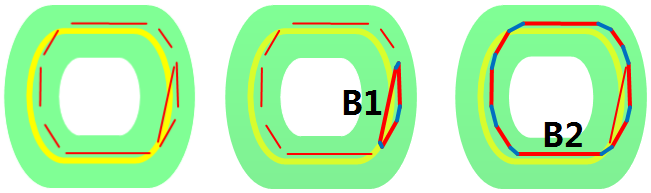
We refer to the process of extracting corresponding salient closed boundaries from sequential video frames as boundary tracking. For each frame, the main idea in boundary extraction is identifying a subset of line segments produced by line detector and connecting them to form a closed boundary which corresponds to the boundary tracked in the previous frame, see Fig. 2. In the first frame of a video sequence, we initialize the closed boundary by selecting a polygon manually.

To form an intact closed boundary from disconnected line segments, we generate another set of in-between gaps to connect them (see the blue lines in Fig. 2c). Line segments that are far from the prior boundary are filtered by distance.

The closed boundary tracking problem is now reduced to a prior shape constrained perceptual grouping problem.

[1] https://github.com/NathanUA/SlientClosedBoundaryTracking.git

[2] https://github.com/NathanUA/SalientClosedBoundaryTracking Dataset.git



(a) Lines within buffer region of prior boundary (b) Wrong grouping (c) Correct grouping

Fig. 3. Illustration of wrong boundary grouping: Yellow contours are prior boundaries. Line segments noise often results in wrong grouping of boundary (B_1) without using area constraint.

Hence, we define a tracking criteria which takes both grouping cost and shape constraint into consideration. The grouping cost Γ, introduced in [30], is given by:

$$\Gamma_B = \frac{|B_G|}{\iint_{R(B)} dx\,dy}, \tag{1}$$

where $|B_G|$ denotes the summation of the gap filling segments length along the boundary B. The denominator $\iint_{R(B)} dx\,dy$ is the area of region $R(B)$ which is enclosed by the boundary B.

Although the distance based filtering (the green region shown in Fig. 2c), excludes many unrelated line segments, it can not handle noisy line segments. To illustrate this, consider the line segments shown in Fig. 3. It is clear that $\Gamma_{B_1} < \Gamma_{B_2}$ which results in an incorrectly grouped boundary B_1. To eliminate this kind of error without significantly increasing time complexity, we propose to use a simple area similarity $S_{Bp\_Bc} < Area\_Threshold$ between the searched boundary (B_c) and the prior shape (B_p) to constrain the grouping process as follows:

$$S_{Bp\_Bc} = min\left(\frac{\iint_{R(Bp)} dxdy}{\iint_{R(Bc)} dxdy}, \frac{\iint_{R(Bc)} dxdy}{\iint_{R(Bp)} dxdy}\right), \tag{2}$$

where $\iint_{R(Bp)} dxdy$ is the area of region $R(Bp)$ which is enclosed by the prior shape boundary Bp. Bp is the initialization or the tracked boundary from the last frame. (Bc is the to-be-tracked boundary).

To solve the grouping problem, we map the detected and generated (fill-in) line segments and their endpoints to a undirected graph $G = (V, E)$. Line segments and endpoints correspond to graph edges and graph vertices respectively, and thus closed boundaries correspond to graph cycles. Now, the problem of closed boundary grouping is converted into an optimal graph cycle searching problem. We developed a new graph based optimization method to find the optimal boundary in real-time.

III. METHOD

The workflow of our salient closed boundary tracking method is shown in Fig. 4. As mentioned above, we approximate smooth boundaries by polygons which are comprised of straight line segments. First, we introduce the line detection and filtering. Then, the details of gap filling are presented, followed by graph modeling and optimization.

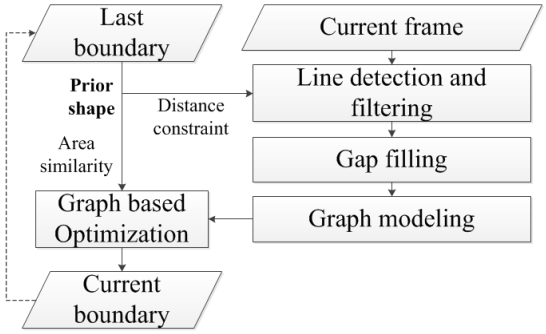


Fig. 4. Workflow of boundary tracking

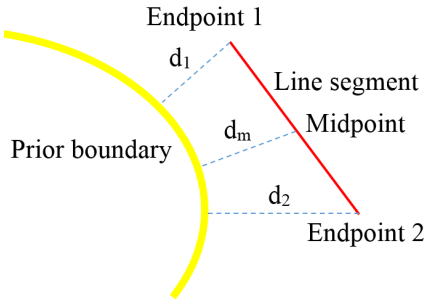


Fig. 5. Line segments filtering: d_1, d_2 and d_m are the smallest distances from corresponding points to the prior boundary.

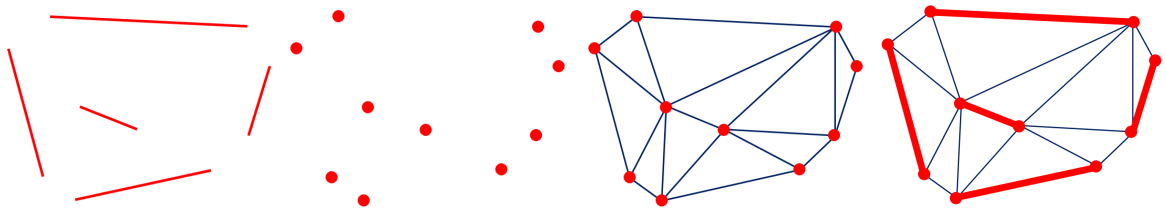
A. Line Detection and Filtering

Straight line segments are fundamental elements in our tracking method. High accuracy and recall are basic requirements for line detectors. For tracking, two other factors are also important: (1) real-time performance; and (2) automatic parameters tuning. We first use EDLines [33], a linear time line segments detector which satisfies these requirements, for automatic detection of boundary line segments. Detected line segments are represented by pairs of endpoints. In each incoming frame, line detection is conducted on the whole frame. Then, lines of interest are filtered by a distance constraint from the previous boundary. Green buffer regions shown in Fig. 2c and Fig. 3 roughly illustrate the distance constraint. We filter a line based on its distance from prior contour. To this end, we use three distances, (two end-points and the mid-point) as shown in Fig. 5. If the average of these distances d_1, d_m and d_2 is smaller than certain threshold (30 pixels), it will be preserved.

B. Gap Filling

After obtaining the line segments of interest, Delaunay Triangulation (DT) [34] is introduced to generate virtual fragments and fill the gaps among disconnected segments,

Preprint submitted to 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Received March 1, 2017.



(a) Line segments (b) Endpoints (c) Delaunay Triangulation (d) Final gap filling

Fig. 6. Gap filling process: The detected line segments are connected by Delauny Triangulation which is performed on the endpoints of these line segments.

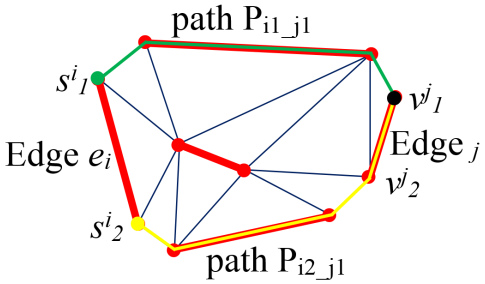


Fig. 7. Cycle candidates construction by BDSP: Edge e_i is the current edge and v_1^j is the third vertex. Edge e_i, shortest paths P_{i1_j1} and path P_{i2_j1} construct a closed cycle candidate.

as illustrated in Fig. 6. First, line segments are degenerated to endpoints (see Fig. 6a and Fig. 6b). Then, DT is conducted on these endpoints (Fig. 6c). Finally, we superimpose the detected line segments in Fig. 6a over the generated DT edges in Fig. 6c. Generated DT edges that overlap the detected line segments are removed. The final result of gap filling is shown in Fig. 6d. To distinguish two kinds of line segments, we refer to the detected line segments (the red lines in Fig. 6d) as *detected* segments and the gap filling segments (the blue lines in Fig. 6d) as *generated* segments.

C. Graph Modeling

Assuming the gap filling is done, we now describe the construction of a undirected graph $G = (V, E)$. We map endpoints and segments (both *detected* and *generated*) to graph edges E and vertices V respectively. The graph has the same structure as the gap filling result (Fig. 6d). We define the edge-weight function for each edge $e \in E$ similar to [30]. Given a graph edge e_i, we set its weight to

$$w(e_i) = \begin{cases} 0 & e_i \text{ is a detected segment} \\ |P_1^i P_2^i| & e_i \text{ is a generated segment} \end{cases} \quad (3)$$

where $|P_1^i P_2^i|$ is the length of the corresponding line segments $P_1^i P_2^i$ of the graph edge e_i.

D. Optimization

Now, our goal is to find the optimal graph cycle which has the minimum boundary cost, based on (1), and satisfies the similarity constraint (2) simultaneously. The key problem is that both the boundary cost and the similarity constraint cannot be determined when the boundary itself is unknown. Furthermore, they are difficult to be integrated into one cost function. So we developed a novel heuristic search method to obtain the optimal cycle. Our method has two steps: generating boundary candidates and finding the optimal one from those candidates.

Given a graph, exhaustive searching is time-consuming and unfavorable. In an attempt to avoid an exhaustive search, we propose a method called "Bidirectional Shortest Path (BDSP)" to generate cycle candidates. This method is based on the hypothesis that cycles with smaller total weights are more likely to contain the optimal boundary. As shown in Fig. 7, for each *detected* edge e_i, we search shortest paths from its two vertices s_1^i and s_2^i to the same third vertex v_1^j using Dijkstra [26]. The weight of e_i is set to infinity other than its original weight zero during searching. The edge e_i, shortest paths P_{i1_j1} and P_{i2_j1} construct a cycle candidate C_{i_j1}. Vertex v_1^j and Vertex v_2^j belong to the same edge e_j and the weight of e_j is zero, thus, taking v_1^j or v_2^j as the third vertex usually produce the same cycle. Given a graph which contains n *detected* line segments, we can generate $(n-1)$ (the current segment is excluded) cycle candidates for each *detected* segment. Hence, the total number of the cycle candidates is $n \times (n-1)$. The number of edges mapped from *generated* line segments determines the time cost of shortest path searching but has no contribution to the number of cycle candidates.

Having obtained these cycle candidates, we can search for the optimal closed boundary by computing their boundary costs (1) and similarity constraints (2) easily. The optimal boundary searching algorithm is shown in Algorithm. 1.

IV. EXPERIMENTAL RESULTS

In this section, we validate our tracking scheme using a number of comparative experiments and a real robot arm experiment.

A. Dataset and Evaluation Measure

We collected nine video sequences of salient close boundaries, as shown in Fig. 9. Each sequence is about 30 seconds (30 fps) and the frame size is 640×480

Algorithm 1 Optimal cycle searching

Input: Undirected graph $G = (V, E)$ and a shape prior represented by a set of ordered points

Output: The optimal cycle

```
Store the optimal cycle C_opt
for i = 0; i < n; i++ do
    Use BDSP to search cycle candidates C_i•
    for j = 0; j < 2(n−1); j+= 2 do
        if i == 0&&j == 0 then
            C_opt = C_ij
        end if
        if S_{C_ij_Bp} > S_e then
            if Γ_{C_ij} < Γ_{C_opt} then
                C_opt = C_ij
            end if
        end if
    end for
end for
return C_opt
```

Notes: 1) the similarity measure threshold $S_e = 0.9$; 2) $C_{i\bullet}$ are $(n-1)$ cycle candidates related to edge e_i.

(width×height). There are 9598 frames in total. In each sequence, different styles of motion such as translation, rotation, zooming and viewpoint changing are all performed. We annotate them by drawing polygons which are well matched with the salient closed boundaries in human vision.

To evaluate our proposed method quantitatively, we define the error metric as the alignment error (E_{AL}) of tracked closed boundary and ground truth as the following:

$$E_{AL} = max\{\frac{B_i \otimes Dist_{B_{gt}}}{P_{B_i}}, \frac{B_{gt} \otimes Dist_{B_i}}{P_{B_{gt}}}\}, \qquad (4)$$

where $\otimes$ represents convolution, Fig. 8, B_i(B_{gt} represents ground truth) denotes the boundary binary image, $Dist_{B_i}$ denotes the distance transform image of B_i and P_{B_i} indicates the perimeter of boundary B_i.

We use the success rate to measure a tracker's overall accuracy [35]. The success rate on a sequence is defined as the ratio of frames where the tracking error E_{AL} is less than a threshold of e_p pixels and the total frames.

B. Results

We compare our tracking method BDSP against following methods: ESM[9] which is a popular registration based homography tracker, RKLT [12] which is a RANSAC based cascade registration based tracker that can handle partial appearance changing and occlusion, HoughTrack [17] which is a state-of-the-art segmentation based tracker that can provide us accurate contour and a tracker adapted from edge grouping method RRC [30]. Both ESM and RKLT were tested with appearance model NCC, which were implemented in a modular tracking framework (MTF) [35]). We initialize them by selecting a bounding box of the target boundary at the first frame. Then boundaries of following frames are computed by homography transformations with respect to the first frame (all of these boundaries are assumed to be planar). We modified RRC by substituting its line detector for EDlines [33] and added a buffer search region as shown in Fig. 2c.

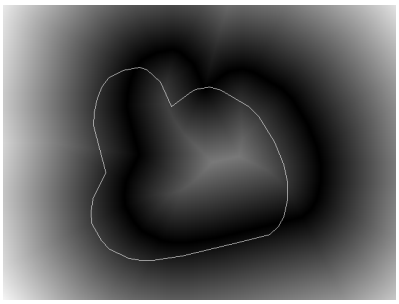


Fig. 8. Convolution of a boundary and the distance transform image of its ground truth

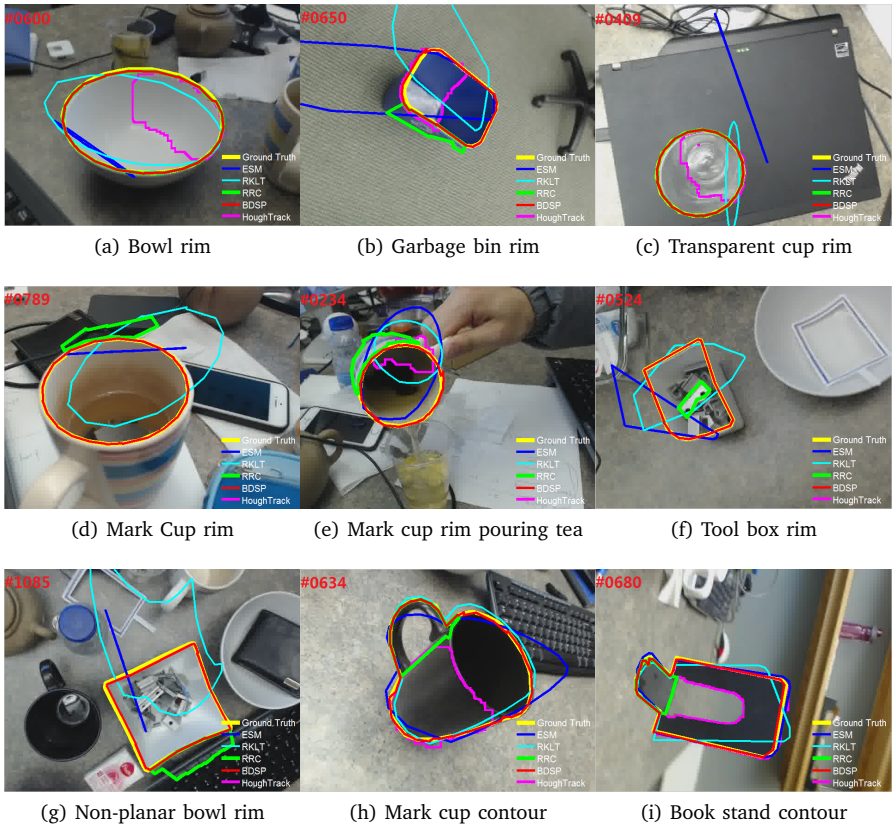
The success rate curves of the above four methods and our method are illustrated in Fig. 10. As we can see, the proposed method (BDSP) performs better than others in almost all of these sequences. Both registration based trackers ESM and RKLT fail quickly because they are sensitive to appearance changing, as blue and cyan boundaries shown in Fig. 9a to Fig. 9g. Although HoughTrack performs better than registration based trackers due to its model updating, it corrupts quickly when the content of target region are heterogeneous as pink contours shown in Fig. 9. Fig. 9a and Fig. 9c show an empty bowl and an empty transparent cup with relative clean background. The corresponding tracking results illustrated in Fig. 10a and Fig. 10c show that RRC tracker produces almost the same success rate with our method. But RRC trackers are very susceptible to noise, as the tracked green contours shown in Fig. 9. The intact video results are included in the supplementary video[3].

In addition, we measured the average processing speed of our method for each of the nine video sequences on a machine with a quad core 3.10 GHz Intel Core i5 processor, 16 GB RAM and Ubuntu 14.04 64-bit OS. Our method is implemented in C++ using OpenCV and Boost library. Table. I illustrates the average graph size, which is indicated by numbers of edges and nodes, the average time costs of line detection (LineTime) and grouping time (GroupTime) and the average frequency per second (fps). The total tracking time per each frame contains line detection time and grouping time. As we can see our method is acceptable for real-time tracking.

C. Robot Arm Pouring Experiment

The salient closed boundary tracker has been used successfully in a real robot arm pouring experiment. The task is to track and follow a moving bowl and then pour

[3]https://youtu.be/RXjD0yHkukI



(a) Bowl rim (b) Garbage bin rim (c) Transparent cup rim

(d) Mark Cup rim (e) Mark cup rim pouring tea (f) Tool box rim

(g) Non-planar bowl rim (h) Mark cup contour (i) Book stand contour

Fig. 9. Results of sample frames. It is noteworthy that there are no HoughTrack contours in (d), (f) and (g) because it has already been corrupted before these certain frames.

Video	Fig.9a	Fig.9b	Fig.9c	Fig.9d	Fig.9e	Fig.9f	Fig.9g	Fig.9h	Fig.9i
Edges	445	508	550	672	753	362	419	505	286
Nodes	80	88	96	112	124	62	70	89	53
LineTime(ms)	4.34	8.78	4.15	4.69	5.17	4.33	5.09	5.07	3.79
GroupTime(ms)	14.45	18.85	19.00	26.10	32.13	8.68	16.93	17.72	6.56
fps	54.12	37.55	43.21	32.48	26.80	76.81	59.74	43.88	96.57

TABLE I. GRAPH SCALE AND TIME EFFICIENCY

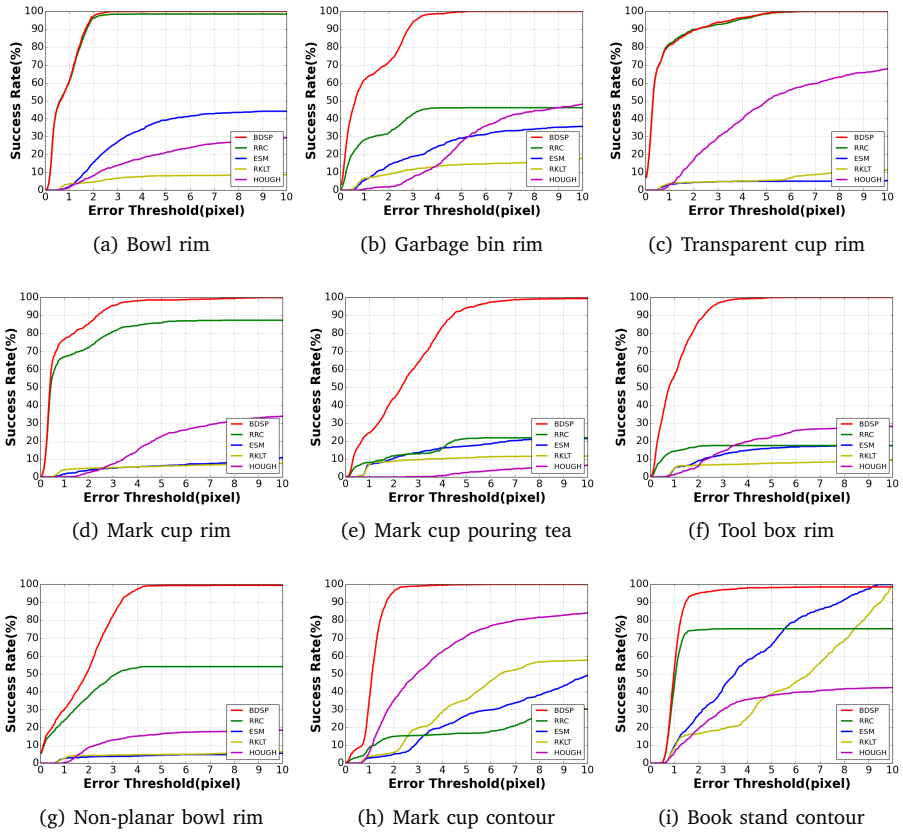
cereal into it. The difficulties of this experiment are that the bowl is non-Lambertian and has no salient textures.

The setup of our experiment is shown in Fig. 11a. The system includes a set of WAM arm and a kinect. The 3D coordinates of the WAM arm and the kinect are registered. We initialize the bowl rim and track it in RGB video stream captured by the kinect. Meanwhile, we map the tracked closed boundary, which is represented by a set of 2D image points, to 3D points acquired by the kinect depth sensor, as shown in Fig. 11b and Fig. 11c. The centroid of the bowl is computed based on these mapped 3D contour points and is taken as the pouring target position. We pre-compute the shifting of the WAM hand to the bowl centroid. When the distance $Dist_{r\_b}$ between the centroid of the tracked bowl and the robot hand satisfies certain thresholds($e_{low} < Dist_{r\_b} < e_{high}$), the WAM arm will pour the cereal into the bowl, as shown in Fig. 11d. Without having the tracking points of the contour provided by our tracker it will be very difficult to find the 3D center of the bowl with any other types of tracker.The experiment shows that our tracker is stable and efficient in real robot application. A demonstration of the pouring task can be seen in the accompanying video.

V. CONCLUSIONS

We presented a novel real-time method for salient closed boundary tracking. By combining a saliency measure and an area constraint as tracking criteria, the proposed method improves the tracking performance greatly. A bidirectional shortest path based boundary candidates searching algorithm enables the real-time solvability of the combined tracking criteria. We validated it quantitatively on various real-world video sequences. Since it is robust and fast enough, it has been used successfully in real robot pouring experiment where other trackers have failed.

Preprint submitted to 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems. Received March 1, 2017.



(a) Bowl rim (b) Garbage bin rim (c) Transparent cup rim

(d) Mark cup rim (e) Mark cup pouring tea (f) Tool box rim

(g) Non-planar bowl rim (h) Mark cup contour (i) Book stand contour

Fig. 10. Success rates of BDSP, RRC, ESM, RKLT and HoughTrack on collected video sequences corresponding to those in Fig. 9.

ACKNOWLEDGMENT

This work has in part been sponsored by the China Scholarship Council and University of Alberta. The support is gratefully acknowledged.

REFERENCES

[1] A. Yilmaz, X. Li, and M. Shah, "Object contour tracking using level sets," in *Asian Conference on Computer Vision*, 2004.

[2] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proceedings of the 7th International Joint Conference on Artificial Intelligence, IJCAI '81, Vancouver, BC, Canada, August 24-28, 1981*, 1981, pp. 674–679.

[3] G. G. Scandaroli, M. Meilland, and R. Richa, "Improving ncc-based direct visual tracking," in *Computer Vision - ECCV 2012 - 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part VI*, 2012, pp. 442–455.

[4] N. D. H. Dowson and R. Bowden, "Mutual information for lucas-kanade tracking (MILK): an inverse compositional formulation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 1, pp. 180–185, 2008.

[5] S. C. N. N. Hofhauser, A., "Edge-based template matching and tracking for perspectively distorted planar objects," in *International Symposium on Visual Computing*, 2008.

[6] G. Zhu, Q. Zeng, and C. Wang, "Efficient edge-based object tracking," *Pattern Recognition*, vol. 39, no. 11, pp. 2223–2226, 2006.

[7] M. Pressigout and É. Marchand, "Real-time hybrid tracking using edge and texture information," *I. J. Robotics Res.*, vol. 26, no. 7, pp. 689–713, 2007.

[8] S. Baker and I. A. Matthews, "Lucas-kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, vol. 56, no. 3, pp. 221–255, 2004.

[9] S. Benhimane and E. Malis, "Real-time image-based tracking of planes using efficient second-order minimization," in *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, Sendai, Japan, September 28 - October 2, 2004*, 2004, pp. 943–948.

[10] T. Dick, C. P. Quintero, M. Jägersand, and A. Shademan, "Real-time registration-based tracking via approximate nearest neighbour search," in *Robotics: Science and Systems IX, Technische Universität Berlin, Berlin, Germany, June 24 - June 28, 2013*, 2013.

[11] J. Kwon, H. S. Lee, F. C. Park, and K. M. Lee, "A geometric particle filter for template-based visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 625–643, 2014.

[12] X. Zhang, A. Singh, and M. Jägersand, "RKLT: 8 DOF real-time robust video tracking combing coarse ransac features and accurate fast template registration," in *12th Conference on Computer and Robot Vision, CRV 2015, Halifax, NS, Canada, June 3-5, 2015*, 2015, pp. 70–77.

[13] J. Shi and C. Tomasi, "Good features to track," in *Conference on Computer Vision and Pattern Recognition, CVPR 1994, 21-23 June, 1994, Seattle, WA, USA*, 1994, pp. 593–600.

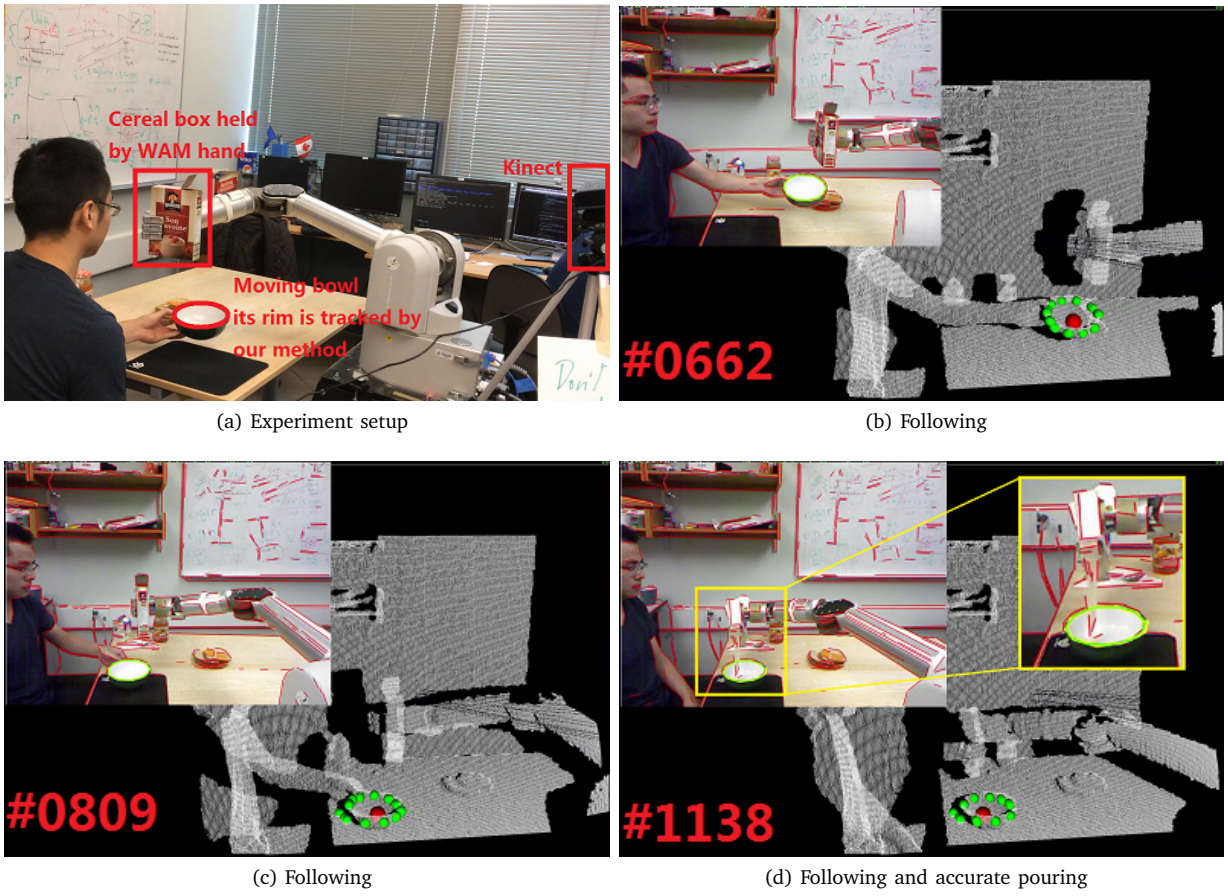
[14] S. Gauglitz, T. Höllerer, and M. Turk, "Evaluation of interest point detectors and feature descriptors for visual tracking," *International Journal of Computer Vision*, vol. 94, no. 3, pp. 335–360, 2011.

[15] G. Klein and D. W. Murray, "Full-3d edge tracking with a particle filter," in *Proceedings of the British Machine Vision Conference 2006, Edinburgh, UK, September 4-7, 2006*, 2006, pp. 1119–1128.

[16] C. Choi and H. I. Christensen, "3d textureless object detection and tracking: An edge-based approach," in *2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, IROS 2012, Vilamoura, Algarve, Portugal, October 7-12, 2012*, 2012, pp. 3877–3884.

[17] M. Godec, P. M. Roth, and H. Bischof, "Hough-based tracking of non-rigid objects," *Computer Vision and Image Understanding*, vol. 117, no. 10, pp. 1245–1256, 2013.

[18] M. Kass, A. P. Witkin, and D. Terzopoulos, "Snakes: Active contour



(a) Experiment setup (b) Following

(c) Following (d) Following and accurate pouring

Fig. 11. Robot pouring experiment: (a) A human is moving the bowl continuously under the surveillance of a kinect. (b)(c) Our algorithm tracks the bowl rim and maps its 2D image points to 3D points in the robot coordinates through the kinect which is registered with the robot coordinates, then, makes the robot arm follow the moving bowl. (d) The robot hand pours the cereal into the bowl accurately according to our tracking result.

models," *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[19] N. Paragios and R. Deriche, "Geodesic active contours and level sets for the detection and tracking of moving objects," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 3, pp. 266–280, 2000.

[20] A. Yilmaz, X. Li, and M. Shah, "Contour-based object tracking with occlusion handling in video acquired using mobile cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 11, pp. 1531–1536, 2004.

[21] Y. Rathi, N. Vaswani, A. Tannenbaum, and A. J. Yezzi, "Tracking deforming objects using particle filtering for geometric active contours," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 8, pp. 1470–1475, 2007.

[22] C. Bibby and I. D. Reid, "Real-time tracking of multiple occluding objects using level sets," in *The Twenty-Third IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2010, San Francisco, CA, USA, 13-18 June 2010*, 2010, pp. 1307–1314.

[23] X. Sun, H. Yao, S. Zhang, and D. Li, "Non-rigid object contour tracking via a novel supervised level set model," *IEEE Trans. Image Processing*, vol. 24, no. 11, pp. 3386–3399, 2015.

[24] M. Pressigout and É. Marchand, "Real time planar structure tracking for visual servoing: a contour and texture approach," in *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems, Edmonton, Alberta, Canada, August 2-6, 2005*, 2005, pp. 251–256.

[25] J. H. Elder and S. W. Zucker, "Computing contour closure," in *Computer Vision - ECCV'96, 4th European Conference on Computer Vision, Cambridge, UK, April 15-18, 1996, Proceedings, Volume I*, 1996, pp. 399–412.

[26] E. W. Dijkstra, "A note on two problems in connexion with graphs," *Numerische mathematik*, vol. 1, no. 1, pp. 269–271, 1959.

[27] S. Wang, T. Kubota, J. M. Siskind, and J. Wang, "Salient closed boundary extraction with ratio contour," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 546–561, 2005.

[28] G. Kanizsa, *Organization in Vision*. New York: Praeger 1979.

[29] J. Edmonds, "Path, trees, and flowers," *Canadian J. Math*, vol. 17, pp. 449–467, 1965.

[30] J. S. Stahl and S. Wang, "Edge grouping combining boundary and region information," *IEEE Trans. Image Processing*, vol. 16, no. 10, pp. 2590–2606, 2007.

[31] T. Schoenemann and D. Cremers, "A combinatorial solution for model-based image segmentation and real-time tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 7, pp. 1153–1164, 2010.

[32] Z. Liu, H. Shen, G. Feng, and D. Hu, "Tracking objects using shape context matching," *Neurocomputing*, vol. 83, pp. 47–55, 2012.

[33] C. Akinlar and C. Topal, "Edlines: A real-time line segment detector with a false detection control," *Pattern Recognition Letters*, vol. 32, no. 13, pp. 1633–1642, 2011.

[34] F. P. Preparata and M. I. Shamos, *Computational Geometry - An Introduction*, ser. Texts and Monographs in Computer Science. Springer, 1985.

[35] A. Singh, A. Roy, X. Zhang, and M. Jägersand, "Modular decomposition and analysis of registration based trackers," in *13th Conference on Computer and Robot Vision, CRV 2016, Victoria, BC, Canada, June 1-3, 2016*, 2016, pp. 85–92.