ICIC Express Letters
Part B: Applications
Volume **7**, Number **5**, May **2016**

ICIC International ©2016 ISSN 2185-2766
pp. 1135–1140

REAL-TIME IMPLEMENTATION OF A KEYPOINT-BASED AUTOMATIC TARGET SELECTION AND TRACKING SYSTEM FOR APPLICATIONS OF RANDOM BIN PICKING

Chi-Yi Tsai*, Cheng-Jui Yu, Chao-Chun Yu and Ching-Chang Wong

Department of Electrical Engineering
Tamkang University
No. 151, Ying-zhuan Road, Tamsui District, New Taipei City 25137, Taiwan
*Corresponding author: chiyi_tsai@mail.tku.edu.tw

Received October 2015; accepted January 2016

Abstract. *Object recognition and detection play important roles in various computer vision applications. When images contain different types of objects, detection of an object-of-interest (OOI) from multiple stacking objects becomes a difficult task to be handled using a monocular camera. In this paper, a novel automatic target selection and tracking algorithm is proposed to address this issue efficiently. The proposed method first uses keypoint correspondences to compute control points of targets appearing in incoming images. Next, each OOI in captured images is separated from multiple stacking objects randomly placed in a box using mean shift clustering approach. Finally, a template-based visual tracking method is used to locate and track center position of the top OOI in the box. When implemented on an Intel Core i5-4440 3.1GHz platform, the proposed algorithm achieves real-time performance about 30 frames per second at 640 × 480 image resolution in the experiments.*

Keywords: Keypoint matching, Object detection, Mean shift clustering, Visual tracking, Random bin picking

1. **Introduction.** Among a variety of robotic random bin picking approaches, computer vision techniques have been widely used in development of vision-based object detection and tracking systems to help robots be more effective in processing target selection and manipulation tasks. Many vision-based object detection and recognition algorithms are presented in the literature; however, only little attention is paid on development of a visual sensing system for detecting an object-of-interest (OOI) from multiple stacking objects randomly placed in a box.

To detect an OOI appearing in the image, various keypoint-based object detection approaches have been proposed to detect the OOI in a query image. Keypoints are image features that can be repeatedly detected in images robust to translation, scaling, and rotation transformations. Several keypoint description methods have been reported in the literature, such as Scale-Invariant Feature Transform (SIFT) [1], Principal Components Analysis-based SIFT (PCA-SIFT) [2] and Speeded-Up Robust Features (SURF) [3] algorithms. Using one of the existing keypoint description methods, the keypoint-based object detection method first extracts a reference keypoint set from a given reference image of the OOI. Next, an object detection module extracts a query keypoint set from incoming images captured from the visual sensing system. The OOI then can be found according to the number of keypoint matches between the query and reference keypoint sets. Finally, a RANdom Sample Consensus (RANSAC) algorithm [4] is applied to removing outliers in the keypoint matches for improving the accuracy rate of object detection.

Traditional keypoint-based object detection methods are useful to detect a single OOI in a complex environment; however, they may fail to detect multiple objects randomly placed in a box, as shown in Figure 1. This problem motivates us to design an automatic

Figure 1. Target selection problem considered in this study: multiple stacking objects are randomly placed in a box, and a vision system has to decide which one should be taken from the box first

target selection and tracking system based on keypoint matches to efficiently detect and track the top OOI from the multiple stacking objects randomly placed in the box. This paper, therefore, presents our proposed design to deal with tasks of multiple object separation, target selection, and target tracking efficiently. Moreover, the proposed method is computationally efficient. When implemented on an Intel Core i5-4440 3.1GHz platform, the frame rate of the proposed method achieves about 30 frames per second (fps) for a video stream with size 640×480 pixels. Thus, the applicability of the proposed target selection and tracking system is greatly improved in automatic robotic manipulation. Experiment results validate the detection robustness and tracking performance of the proposed method applied in the scenario of random bin picking.

In the remainder of the paper, Section 2 introduces the proposed target selection and tracking method. Section 3 presents the experimental results to evaluate the performance of the proposed method. Section 4 concludes the contributions of this work.

2. **Proposed Design.** Figure 2 shows the flowchart of the proposed automatic target selection and tracking system, which consists of three stages: object detection, object clustering, and object tracking. In object detection stage, the keypoint-based object detection method is used to detect all possible objects in the image based on keypoint matches between a given object reference image and an input query image. In our implementation,

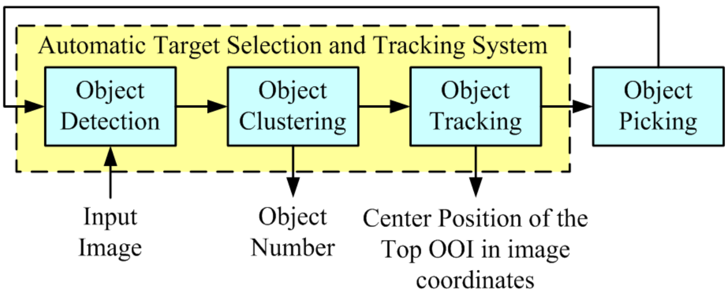

Figure 2. Flowchart of the proposed automatic target selection and tracking system to deal with the random bin picking problem shown in Figure 1

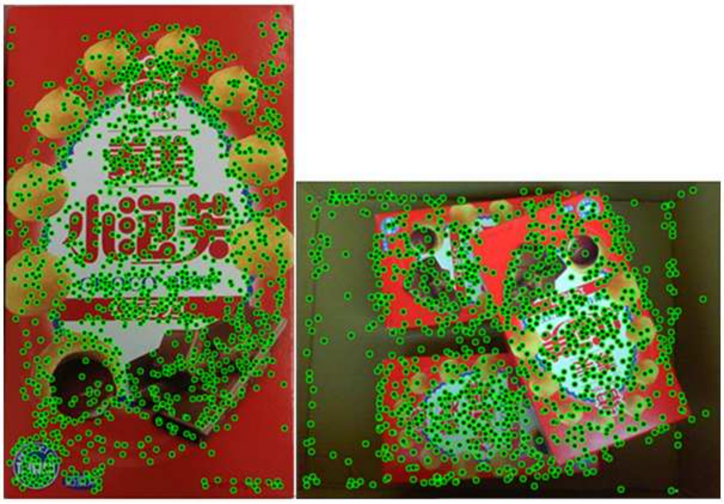

Figure 3. Results of keypoint extraction from an object reference image (left) and an input query image (right)

the SURF algorithm was employed as the keypoint description method to extract keypoint descriptors of the input image. Figure 3 shows the SURF keypoints extracted from the reference and query images. Next, a fast keypoint descriptor matching algorithm [5] is applied on the SURF keypoints to search keypoint matches between the reference and query keypoint sets.

In object clustering stage, a center-point voting algorithm [6] is firstly used to estimate the position of center point of each object in the image. Then, the number of objects is determined using a data clustering algorithm, such as K-means [7] or mean shift [8] techniques, to classify all detected objects into different data clusters. In this study, we used mean shift algorithm for data clustering as it is suitable to classify multiple classes without an initial guess of the number of classes. Let m_q and m_r respectively denote the matched SURF keypoints in the query and reference images. Then, the center point c_q of a corresponding object in the query image is estimated by

$$c_q = m_q + r \cdot \mathbf{M}_{ori}(\phi) \cdot v_r, \tag{1}$$

where r is a scale ratio defined as

$$r = \frac{\sigma_q}{\sigma_r}, \tag{2}$$

where σ_q and σ_r denote the scale of the keypoints m_q and m_r, respectively. The symbol v_r denotes a reference central vector from the center point of the reference object to the matched reference keypoint m_r such that

$$v_r = m_r - c_r, \tag{3}$$

where c_r is the center point of the reference object. The symbol $\mathbf{M}_{ori}(\phi)$ denotes an orientation rotation matrix associated with the difference between the orientation angle of the two matched keypoints m_q and m_r such that

$$\mathbf{M}_{ori}(\phi) = \begin{bmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{bmatrix}, \tag{4}$$

$$\phi = \theta_q - \theta_r, \tag{5}$$

where θ_q and θ_r are the orientation angles of the keypoints m_q and m_r, respectively. Figure 4(a) shows the center-point voting result obtained from Equation (1). It is clear from Figure 4(a) that Equation (1) efficiently maps each matched query keypoint to the location near the center point of a corresponding OOI. Next, the mean shift clustering algorithm is used to classify the center-point voting result into multiple object clusters by

(a) (b)

FIGURE 4. Results of object clustering process: (a) center-point voting result obtained using Equation (1); (b) data clustering result obtained using mean shift algorithm, in which the black points indicate the estimation of center point of each OOI in the image

FIGURE 5. Template patch used in the test of object tracking stage

iteratively searching the densest places of the voted center points. By doing so, the number of objects and the center point position of each object can be simultaneously estimated. Figure 4(b) shows the object clustering result obtained from the mean shift clustering algorithm. It is clear from Figure 4(b) that the voted center points are classified into three data classes, which is the correct number of objects appearing in the query image. Moreover, the black points indicate the estimated center point of each detected object, which is close to the real center point of each object in the image.

In object tracking stage, a template-based visual tracking method [9] is employed to track the top OOI in real-time. Object tracking is an important function in vision systems to handle object picking tasks as it provides an efficient way to continuously locate the center point of the OOI in the image with real-time performance. Thus, a visual tracking algorithm is adopted in the proposed system to track the top OOI after clustering and locating each OOI in the query image. Here, the object cluster having maximum voted center points is determined as the top OOI. Moreover, a template-based visual tracking algorithm is employed in our design because a part of the reference image around the center point of the OOI can be used as the template patch for template matching. Figure 5 shows the template patch used in the experiments. The detection robustness and tracking performance of the proposed automatic target selection and tracking algorithm is validated in the next section.

Figure 6. Experimental results of the proposed automatic target selection and tracking algorithm to consecutively detect and track one of multiple targets randomly placed in a box

3. **Experimental Results.** The proposed algorithm was implemented in C++ with OpenCV 2.4.3 library running on a Windows 7 platform with 3.1GHz Intel Core i5-4440 processor. Three objects randomly placed in a box were used as the targets to test detection robustness and tracking performance of the proposed system in the application of real-time target selection and tracking. Figure 6 presents the experimental results, in which the proposed algorithm aims to consecutively detect and track one of multiple targets placed in the box. In Figure 6(a), the three objects in the box are detected successfully. In Figure 6(b), the top object is selected as the OOI to be tracked and picked up from the box, as shown in Figure 6(c). In Figures 6(d), 6(e) and 6(f), the second largest object cluster is then chosen as the OOI to be taken out from the box. Finally, the last object is detected as the OOI to be picked up, as shown in Figures 6(g)-6(i). Therefore, these experimental results validate the detection robustness and tracking performance of the proposed algorithm. Note that, the proposed algorithm is also computationally efficient because the frame rate of the entire system is about 30 fps when processing a standard VGA (640×480 image resolution) video stream. This advantage improves the usability of the proposed algorithm in practical applications.

4. **Conclusions.** In this paper, a novel automatic target selection and tracking algorithm is proposed based on keypoint matches. The proposed algorithm consists of object detection, object clustering, and object tracking processes. Any SIFT-like keypoint detector can be used in the object detection process of the proposed algorithm. Next, the object clustering process estimates the number of objects and the center point position of each object using a center-point voting process combined with the mean shift clustering

algorithm. Moreover, the top object also can be determined according to the number of voted center points in the classified object clusters. Finally, a template-based visual tracking algorithm is used to track the top OOI in real-time. Experimental results validate the performance of the proposed algorithm, in terms of detection accuracy, usability, and computational efficiency. In future work, the proposed algorithm will be integrated into a robot manipulator system to show the performance of the proposed method in applications of random bin picking.

Acknowledgment. This work was supported by the Ministry of Science and Technology (MOST) of Taiwan under grant MOST 103-2221-E-032-068 and 103-2632-E-032-001-MY3. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] D. G. Lowe, Object recognition from local scale-invariant features, *IEEE International Conference on Computer Vision*, Kerkyra, Corfu, Greece, pp.1150-1157, 1999.
[2] Y. Ke and R. Sukthankar, PCA-SIFT: A more distinctive representation for local image descriptors, *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington DC, USA, pp.506-513, 2004.
[3] H. Bay, A. Ess, T. Tuytelaars and L. V. Gool, SURF: Speeded up robust features, *Computer Vision and Image Understanding*, vol.110, no.3, pp.346-359, 2008.
[4] M. A. Fischler and R. C. Bolles, Random sample consensus: A paradigm for model fitting with applications to image analysis and automated catography, *Communications of the ACM*, vol.24, no.6, pp.381-395, 1981.
[5] C.-Y. Tsai, A.-H. Tsao and C.-W. Wang, Real-time feature descriptor matching via a multi-resolution exhaustive search method, *Journal of Software*, vol.8, no.9, pp.2197-2201, 2013.
[6] Z. Haladova and E. Šikudo, Multiple instances detection in rgbd images, *Lecture Notes in Computer Science*, vol.8671, pp.246-253, 2014.
[7] J. MacQueen, Some methods for classification and analysis of multivariate observations, *Proc. of the 5th Berkeley Symposium on Mathematicial Statistics and Probability*, vol.1, pp.281-297, 1967.
[8] Y. Cheng, Mean shift, mode seeking, and clustering, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol.17, no.8, pp.790-799, 1995.
[9] R. Brunelli, *Template Matching Techniques in Computer Vision: Theory and Practice*, John Wiley & Sons, 2009.