Accurate Label Detection Technique for 3D Vision Based Robotic Bin Picking

Udaya Wijenayake, Sung-In Choi and Soon-Yong Park
School of Computer Science & Engineering, Kyungpook National University
e-mail: udaya@vision.knu.ac.kr, ellimn5th@naver.com, sypark@knu.ac.kr

Abstract

In the field of computer vision and robotics, bin picking is a major application area where the object pose estimation is important. Different approaches such as 2D feature tracking and 3D surface reconstruction have been introduced in order to accurately estimate the object pose. We have proposed a new approach where we can use both 2D image features and 3D surface information to accurately identify the target object and estimate its pose. As an initial step, we proposed a label detection technique using Maximally Stable Extremal Regions (MSER) where the label detection results can be used to identify the target objects separately.

1. Introduction

Bin picking is a task of picking random objects from a container or bin, which is mostly achieved by vision-guided robotic systems. Different computer vision techniques have been used in order for robot to be able to accurately detect the shape, size, position, and alignment of the objects in the bin. Most of the previous methods use 3D features of the objects to automatically detect, localize, and reliably estimate its pose.

Boughorbel et al. [1] used a laser range finder to reconstruct the 3D seen of the target objects and then find the geometry of the bin contents to perform precise grasping operations. Berger et al. [2] proposed another bin picking system consist of a grid pattern projector and a visual camera. They determined the pose of the target object by analyzing the properties of the grid pattern projected on the objects. Park et al. [3] proposed a pose estimation technique using range images that can be used in robotic bin picking.

Few bin picking systems have been proposed based on 2D features of the target objects [4] instead of using 3D features. In our proposed system, all the target objects are attached with labels. Therefore, we proposed a method that use 2D features of the labels to detect the object and 3D features to identify the orientation of the detected objects. Fig. 1 shows the step-by-step pose estimation process of the proposed bin picking system. First, two different views of the target objects are captured using a stereo camera and processed separately to detect labels. Then 2D features are detected on the label areas and feature matching is used to identify corresponding labels from the two different views. These correspondences information are used to reconstruct the 3D surface of the objects. Final object pose estimation and 3D distance measurement is done based on the reconstructed 3D surface. In this paper, we only explain the label detection stage of the proposed system, which is very important to achieve high performance and accuracy.

2. Proposed Label Detection Method

We propose a label detection method using Maximally Stable Extremal Regions (MSER). In computer vision, MSERs are used as a method of blob detection in images. This technique was first proposed by Matas et al. [5] in 2002 to find correspondences between two images with large viewpoint difference. Recently, MSER technique is widely used in wide baseline stereo matching, text detection, and object recognition algorithms. MSERs have all the properties required of a stable local detector such as closed under continuous geometric transformations, invariant to affine intensity changes, and scale invariant.

MSERs denote a set of distinguished regions defined by an extremal property of its intensity function in the region and on its outer boundary. As an example, if we were shown a movie of thresholded images I_g where g represents the corresponding threshold value, first a white image is appeared and subsequently black spots corresponding to local intensity minima will appear and grow. These 'black' spots will eventually merge, until the whole image is black. The set of all connected components in the complete thresholded image sequence is defined as extremal regions. If an extremal region is stable over a large range of threshold values, such regions are defined as MSERs.

In our proposed system, we use a MSER detection algorithm implemented based on Component Tree data structure [6]. Here, the component tree is a rooted connected tree where each node of the component tree represents a connected region R_i (extremal region) within the input image I_{in}.

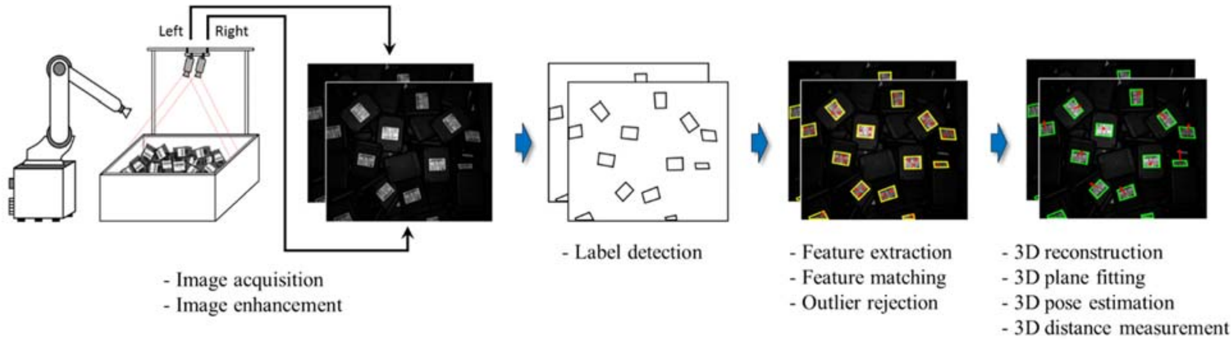

Figure 1. Proposed 3D vision based robotic bin picking system.

$$\forall p \in R_i, \forall q \in \text{boundary}(R_i) \rightarrow I_{in}(p) \geq I_{in}(q) \quad (1)$$

These extremal regions (nodes of the component tree) are identified as connected regions within the binary threshold images I_{bin}^{g} where $g \in [\min(I_{in}) \max(I_{in})]$.

$$I_{bin}^{g} = \begin{cases} 1 & I_{in} \geq g \\ 0 & otherwise \end{cases} \quad (2)$$

The edges of the tree define an inclusion relationship between the connected regions. For a region R_i, which is a child of the region R_j satisfies the following equation.

$$\forall p \in R_i \rightarrow p \in R_j \quad (3)$$

For each connected region $R_i$within the tree, a stability value Ψ is calculated using the following equation where Δ is a user defined stability range parameter.

$$\psi\left(R_i^g\right) = \frac{\left(\left|R_j^{g-\Delta}\right| - \left|R_k^{g+\Delta}\right|\right)}{\left|R_i^g\right|} \quad (4)$$

MSERs correspond to those nodes of the tree that have a stability value Ψ that is a local minimum along the path to the root of the tree.

In the proposed system, the above-explained MSER detection algorithm is used to identify the image regions, which are corresponding to the label areas (see Fig. 2). Minimum and maximum allowable region sizes are defined according to the target samples in order to avoid the detection of unnecessary regions. Δ value is also changed according to the target sample in order to achieve higher detection rate.

After detecting the MSERs, rotated rectangles enclosing the minimum area of the MSERs are found and these rectangles are used to define the candidate label areas. As multiple MSERs can be detected within the same label area, we follow a duplicate removing process, which combines these multiple MSERs in to a single label area. In order to remove duplicates, first the four corner points of each candidate label areas (rotated rectangles) are ordered from top left to bottom left. Then the label areas, which are located very near to each other, are found by measuring the distance between corresponding corner points. These detected nearby label areas are then combined to a single label and the new corner points are defined by finding the average coordinates.

Figure 3. Final label detection results using MSER on different samples.

Figure 2. Results of MSER detection.

Fig. 3 shows some of the final label detection results on few samples we have prepared in the laboratory.

3. Discussion

We proposed a robotic bin picking system that uses stereo vision techniques to estimate the pose of the target objects. Whole system consist of several stages including image acquisition and enhancement, label detection for identify target objects, 2D feature extractions, 3D reconstruction, pose estimation and distance measuring. Among these several stages, only the label detection stage is explained in this paper. Using the MSER techniques, proposed method were able to achieve nearly 100% detection rate for simple targets (Fig. 3a, c) and nearly 80% detection rate for more complex targets (Fig. 3d). Furthermore, the proposed method were able to detect the labels from a 1928×1448 image within 0.5 s.

Acknowledgement

This project was financially supported by the Electronics and Telecommunications Research Institute (ETRI), Republic of Korea, for the research work of ‘Developing a near distance bin picking sensing module’.

References

[1] F. Boughorbel , Y. Zhang , S. Kang , et al., “Laser ranging and video imaging for bin picking,” Assem. Autom., vol. 23, no. 1, pp. 53–59, 2003.

[2] M. Berger, et al., “Vision guided bin picking and mounting in a flexible assembly cell,” 13th Int. Conf. Ind. Eng. Appl. Artif. Intell. Expert Syst., pp. 109–118, 2000.

[3] I. K. Park, M. Germann, M. D. Breitenstein, et al., “Fast and automatic object pose estimation for range images on the GPU,” Mach. Vis. Appl., vol. 21, no. 5, pp. 749–766, 2010.

[4] K. Rahardja and a. Kosaka, “Vision-based bin-picking: recognition and localization of multiple complex objects using simple visual cues,” Proc. IEEE/RSJ Int. Conf. Intell. Robot. Syst. IROS ’96, vol. 3, pp. 1448–1457, 1996.

[5] J. Matas, O. Chum, M. Urban, et al., “Robust wide-baseline stereo from maximally stable extremal regions,” Image Vis. Comput., vol. 22, no. 10, pp. 761–767, Sep. 2004.

[6] M. Donoser and H. Bischof, “Efficient Maximally Stable Extremal Region (MSER) Tracking,” 2006 IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. - Vol. 1, 2006.