Journal of Institute of Control, Robotics and Systems (2016) 22(1):8-16
http://dx.doi.org/10.5302/J.ICROS.2016.15.0160 ISSN:1976-5622 eISSN:2233-4335

빈피킹을 위한 스테레오 비전 기반의 제품 라벨의 3 차원 자세 추정

Stereo Vision-Based 3D Pose Estimation of Product Labels for Bin Picking

Udaya Wijenayake, 최 성 인, 박 순 용*
(Udaya Wijenayake[1], Sung-In Choi[1], and Soon-Yong Park[1,*])
[1]School of Computer Science and Engineering, Kyungpook National University

Abstract: In the field of computer vision and robotics, bin picking is an important application area in which object pose estimation is necessary. Different approaches, such as 2D feature tracking and 3D surface reconstruction, have been introduced to estimate the object pose accurately. We propose a new approach where we can use both 2D image features and 3D surface information to identify the target object and estimate its pose accurately. First, we introduce a label detection technique using Maximally Stable Extremal Regions (MSERs) where the label detection results are used to identify the target objects separately. Then, the 2D image features on the detected label areas are utilized to generate 3D surface information. Finally, we calculate the 3D position and the orientation of the target objects using the information of the 3D surface.

Keywords: bin picking, stereo vision, MSER, pose estimation

I. 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 a robot to be able to detect the shape, size, position, and alignment of the objects in the bin accurately. Most of the previous methods used 2D or 3D features to detect, localize, and reliably estimate the pose of the objects.

Kirkegaard and Moeslund [1] proposed a bin picking system, which recognize and localize multiple complex objects only using simple visual clues (2D featured). The methods solely depend on 2D features have drawbacks of handling object reflection, lighting condition of the environment and overlapping target objects. To avoid these drawback, most of the recent researches focus not only on 2D features but also 3D features of the objects.

Boughorbel et al. [2] used a laser rangefinder to reconstruct the 3D scene of the target objects and find the geometry of the bin contents to perform precise grasping operations. Berger et al. [3] 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. [4] proposed a pose estimation technique using range images that can be used in robotic bin picking. Kirkegaard and Moeslund [5] developed a bin picking system based on Harmonic Shape Contexts (HSC) and graph-based matching. Most of the 3D based bin picking systems use laser range scanners to reconstruct the 3D surface of the target objects accurately [6,7].

In our proposed system, all the target objects are attached with labels. Therefore, we proposed a method that uses 2D features of the labels to detect the object and 3D surface information 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 various image processing algorithms are applied to detect the labels. Then 2D feature points are detected in the label areas, and feature matching is used to identify corresponding labels from the two views. This correspondence information is used to reconstruct the 3D surface of the objects using triangulation technique. Final pose estimation, which includes the 3D position and the orientation, is done based on the reconstructed 3D surface of the label.

A detailed description of the proposed label detection method is given in Section II while the Section III describes how to find the 3D plane of the label. Section IV explains the label pose estimation process, which includes the 3D position and the orientation. Finally, the accuracy evaluation of the proposed system is given in Section V while Section VI concludes the paper.

II. LABEL DETECTION

In this paper, a label detection method is proposed using Maximally Stable Extremal Regions (MSER). In computer vision, MSER is used as a method of blob detection in images. Matas et al. [8] first proposed this technique in 2002 to find

* Corresponding Author
Manuscript received August 24, 2015 / revised November 14, 2015 / accepted December 16, 2015
우다야 위제나야카, 최성인, 박순용: 경북대학교 IT대학 컴퓨터학부
(udaya@vision.knu.ac.kr/ellim5th@naver.com/sypark@knu.ac.kr)
※ 본 연구는 산업통상자원부 및 한국산업기술평가관리원의 산업핵심기술개발사업 (NO. 10038660, 양팔 작업을 위한 센서융합 인지기반 제어기술 개발 및 다중로봇 협업생산공정 적용 기술 개발)과 한국연구재단 일반연구자지원사업 (2014R1A1A2059784)의 지원으로 수행되었음.

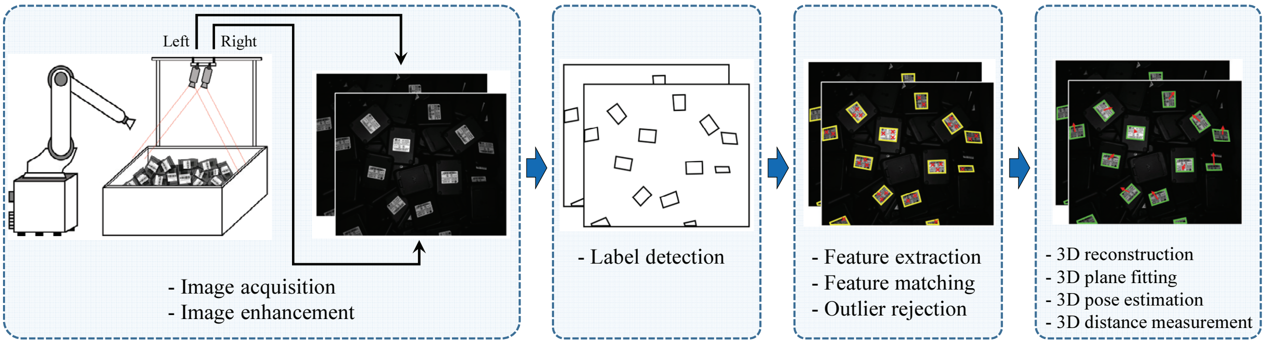
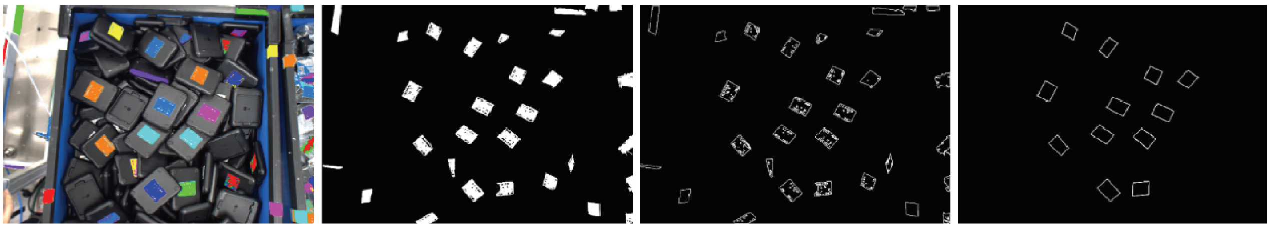


그림 1. 제안된 3차원 비전 기반의 로봇 빈피킹 시스템.
Fig. 1. Proposed 3D vision based robotic bin picking system.

그림 2. 다양한 샘플에 대한 MSER의 결과.
Fig. 2. Results of MSER detection on various samples.



(a) (b) (c) (d)

그림 3. 라벨의 검출 과정 (a) 입력영상에 대한 MSER 검출결과 (b) MSER 검출결과의 이진영상 (c) 검출된 모든 폐곡선 (d) 정제 과정 이후에 남은 폐곡선들.
Fig. 3. Process of label detection. (a) MSER detection results on the input image. (b) MSER detection results as a binary image. (c) All detected contours. (d) Remaining contours after applying refinements.

correspondences between two images with a large viewpoint difference. Recently, MSER technique is extensively used in wide baseline stereo matching, text detection, and object recognition algorithms. Continuous geometric transformations, invariant to affine intensity changes, and scale invariance are some of the properties of MSER, which enable it to be a stable local detector.

In the proposed system, an MSER detection algorithm implemented based on Component Tree data structure [9] is used to identify the label areas approximately. 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}.

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

These extremal regions (nodes of the component tree) are the 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. Thus, a child region R_i of a parent region R_j satisfies the following equation.

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

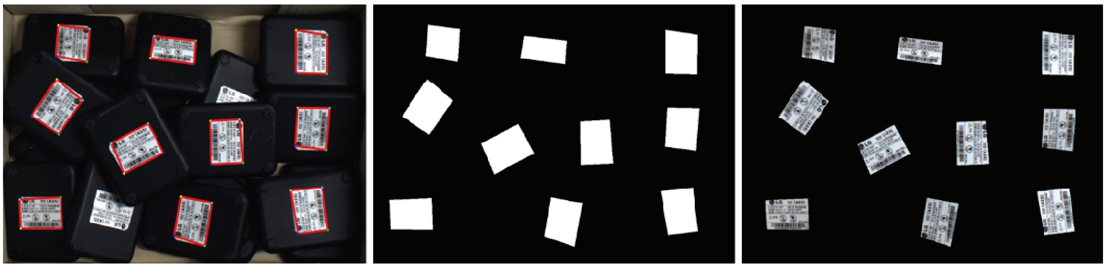
For each connected region R_i within the tree, a stability value ψ is calculated using the following equation where Δ is a user-defined parameter of stability range.

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

The nodes of the tree that have a stability value ψ, which is a local minimum along the path to the root of the tree, are then considered as MSERs.

그림 4. 다양한 샘플들에 대한 라벨 검출 결과.

Fig. 4. Label detection results on various samples.



(a) (b) (c)

그림 5. 배경과 라벨영역의 분리를 위한 이진 마스크 생성 (a) 라벨 검출 결과 (b) 라벨 검출 결과를 이용한 이진 마스크 영상 (c) 배경을 제거한 라벨 영상.

Fig. 5. Creating a binary mask for separating the label areas from the rest of the image. (a) Label detection results. (b) Binary mask image created using the label detection results. (c) Label areas after removing the background.

The proposed system use the above-explained MSER detection algorithm to identify the image regions, which are possibly corresponding to the label areas (see Fig. 2). We define the minimum and maximum allowable region size according to the target samples to avoid the detection of unnecessary regions. The value of Δ is also changed according to the target sample to achieve higher detection rate.

Then, a binary image as in Fig. 3(b) is generated using the MSER detection results to separate the MSERs from the background. By following this step, we were able to merge multiple MSERs that correspond to the same label area into a single connected component region. After generating the binary image, a contour detection algorithm with several refinement steps is applied for identifying the labels separately from all detected MSERs. First, too large or too small contours that are not corresponding to label areas are removed by examining the contour size. All the remaining contours are approximated to polygons using the Ramer-Douglas-Peucker algorithm. Then we remove the polygons, which have the following properties; not convex, not parallelepiped, have less or more than four vertices, and have a large width to height ratio. Fig. 3(c) shows the results of contour detection and Fig. 3(d) shows the remaining contours after applying the filters mentioned above. Using these refinement steps, we were able to remove the partially visible labels and unnecessary regions detected during the MSER detection step.

Remaining contour polygons are considered as the detected labels while the four vertices of the polygon are considered as the corner points of the label. A corner-finding technique with subpixel level accuracy is used to refine the four corners to achieve higher accuracy and stability in the later stages of the proposed bin picking system. Fig. 4 shows the final label detection results on few samples.

III. 3D PLANE FITTING

Label plane estimation in 3D is required to identify the position and the orientation of the label. After detecting the labels from both left and right camera images, first, a feature matching technique is applied to identify the label correspondences. Then, stereo vision based 3D reconstruction alone with a plane fitting technique is used to calculate the 3D plane of the label.

1. Label matching

Finding the matching labels from the left and right camera images is the first step of the label plane estimation. We match the detected labels by extracting SIFT feature points [10] and then finding the correspondences using feature matching. The feature point extraction is limited only to the local label areas by using a binary mask as in Fig. 5(b) to improve both the accuracy and speed. We limit the number of detected feature points in between 40 and 100 for each label on both left and right camera images.

All the feature points extracted from the left and right images are matched using a 1:1 comparison of each SIFT feature descriptors based on a brute-force algorithm. We remove the feature matching results that do not adhere to the following conditions.

- *Rotation Constraint*: the angular difference between the feature points should be less than 20°.
- *Scale Constraint*: scale difference between matching points should be less than one step, or they should be on the same

scale.

- *Uniqueness Constraint*: matching points that maintain one-to-one relationship are only considered as successful matches while the matching points with one-to-many relationship are excluded.

After finishing the initial matching, epipolar geometry based outlier removing technique is applied to increase the robustness of the feature matching. According to the epipolar geometry, when x and x' are corresponding points of a stereo image pair in homogeneous coordinates, Fx depicts an epipolar line of which the corresponding point x' must lie. Therefore, Equation 5 must hold for all pairs of corresponding points [11].

$$x'^{T}Fx = 0 \;\rightarrow\; [u' \quad v' \quad 1]\begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix}\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = 0 \qquad (5)$$

We calculate the fundamental matrix F using the intrinsic camera parameters K and the essential matrix E as in Equation 6. Camera calibration is used to find the intrinsic (K) and extrinsic parameters ($[R\,|\,t]$) of the stereo camera system. The essential matrix E is calculated by Equation 7.

$$F = (K'^{T})^{-1}EK^{-1} \qquad (6)$$

$$E = [t]_{\times}R = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix}\begin{bmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{bmatrix} \qquad (7)$$

In the proposed system, if the reprojection error between two matching points x and x' is greater than 0.75, they are considered as outliers. We calculate the reprojection error as in Equation 8, where $d(a,b)$ represents the Euclidean distance between point a and b.

$$\varepsilon = d(x', Fx)^2 + d(x, F^{T}x')^2 \qquad (8)$$

After completing the feature matching, all the remaining feature points are grouped into labels using the label polygon information. Then we use the feature matching information to find the matching labels from the left and right images and assigned a label ID. A look-up-table (LUT) and a sorting algorithm are used to speed up the matching process.

2. 3D reconstruction

We reconstruct the labels using left/right correspondences of identical label IDs. For every left/right matching points $p_l = [x_l, y_l]^T$ and $p_r = [x_r, y_r]^T$, we calculate a 3D point $P = [x_w, y_w, z_w]^T$ using the triangulation given in Equation 9. Here, $d = x_l - x_r$, $o_l = o_r = [x_c, y_c]^T$, and f represent the disparity, optical center, and focal length respectively. b is the baseline between the cameras after stereo rectification. Fig. 6 depicts the stereo geometry of the system and Fig. 7 shows two different 3D views of labels after reconstruction.

$$x_w = \frac{b(x_l - x_c)}{d}, \quad y_w = \frac{b(y_l - y_c)}{d}, \quad z_w = \frac{bf}{d} \qquad (9)$$

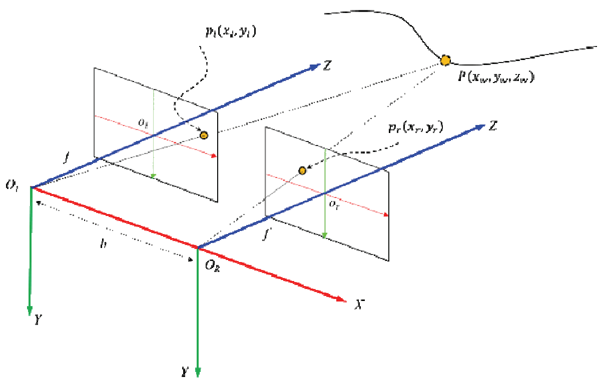


그림 6. 스테레오 삼각법을 이용한 3차원 복원 기하학.

Fig. 6. Stereo triangulation based 3D reconstruction geometry.

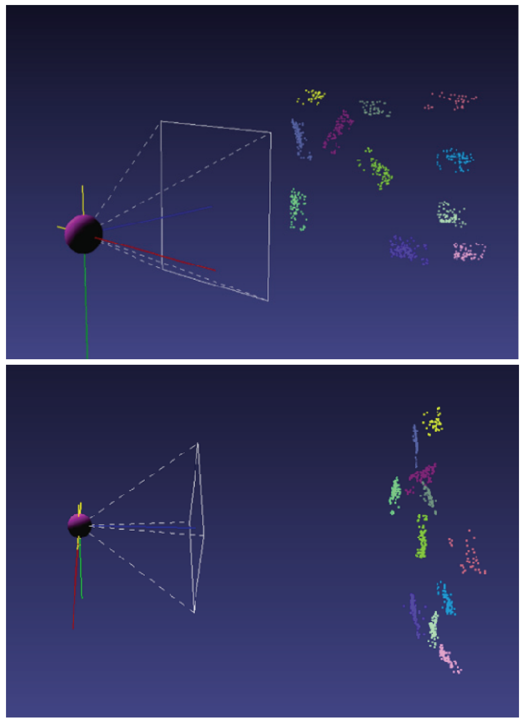


그림 7. 스테레오 기반의 라벨 특징점들의 3차원 복원.

Fig. 7. Stereo vision based 3D reconstruction of label feature points.

3. Plane fitting

We calculate the plane equation of a label using the reconstructed 3D points that belong to the label area. In general, the equation of a 3D plane can be expressed as $ax + by + cz + d = 0$, and a, b, c, d should be known to determine the plane. When considering the plane as a vector $P = [a, b, c, d]^T$, the plane equation can be rewritten as in Equation 10.

$$ax + by + cz + d = 0 \;\rightarrow\; [x \quad y \quad z \quad 1]\begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = 0 \qquad (10)$$

If there are n number of 3D points that belong to the 3D plane, P can be calculated using the linear equation

$$AP = 0 \quad \rightarrow \quad \begin{bmatrix} x_1 & y_1 & z_1 & 1 \\ x_2 & y_2 & z_2 & 1 \\ \vdots & \vdots & \vdots & 1 \\ x_n & y_n & z_n & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix} = 0. \tag{11}$$

Single Value Decomposition (SVD) is used to determine the P from the above linear equation. If the dot product between camera Z axis and the plane normal is a positive value, P is negated to change the surface normal towards the camera plane.

IV. LABEL POSE ESTIMATION

We estimate the label pose, which includes the 3D position and the orientation reference to the left camera origin, using the four corners and the 3D plane of the label.

1. Label position

If any corner of the label in the image plane is $p_i(x_i, y_i)$ and the intrinsic camera matrix is K, vector e_i depicted in Fig. 8 can be calculated using Equation 12.

$$e_i = K^{-1}[p_i, 1]^T \tag{12}$$

When the 3D plane of the label is defined as $P = [a, b, c, d]$ and the normalized vector of e_i is denoted as $\hat{e}_i(i, j, k)$, crossing point (E_i) of vector e_i and plane P can be found using Equation 13 where t_i is given by Equation 14. These crossing points (E_i) are considered as the 3D coordinates of the four label corners. The centroid C of the label, which is calculated using Equation 15, is used as the 3D position of the label reference to the left camera. The normal vector N of the label plane is defined as $N = [a, b, c]$ using the plane equation.

$$E_i = t_i \hat{e}_i \tag{13}$$

$$t_i = \frac{-d}{ai + bj + ck} \tag{14}$$

$$C = \frac{1}{4} \sum_{i=1}^{4} E_i \tag{15}$$

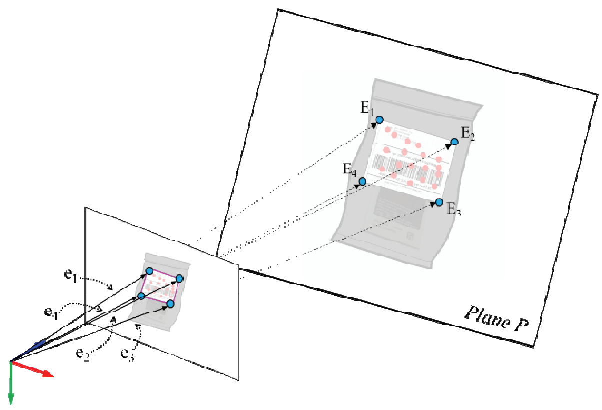


그림 8. 역투영을 이용한 라벨 코너점들의 3차원 좌표 획득.
Fig. 8. Obtaining 3D coordinates of the label corner points using back-projection.

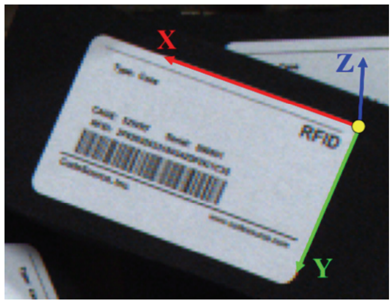


그림 9. 라벨의 3차원 자세 추정을 위한 라벨 좌표계의 정의.
Fig. 9. Defining a label coordinate system to estimate the 3D orientation of the label.

2. Label orientation

To find the 3D orientation of a label with respect to the left camera, first, a label coordinate system is defined as in Fig. 9, where the origin locates in one corner, X-axis goes along the longer side of the label, Y-axis goes along the shorter side of the label, and Z-axis goes along the normal vector of the label plane. The distances between four corner points are used to identify the longer side and the shorter side of the label separately. Here, 3D Euclidean distance $D = \sqrt{E_{i+1} \cdot E_i}$, where $E_i(X_i, Y_i, Z_i)$ is the 3D coordinates of the corner points and "·" is the dot product, is used instead of the 2D distance in the image plane due to the perspective view problem.

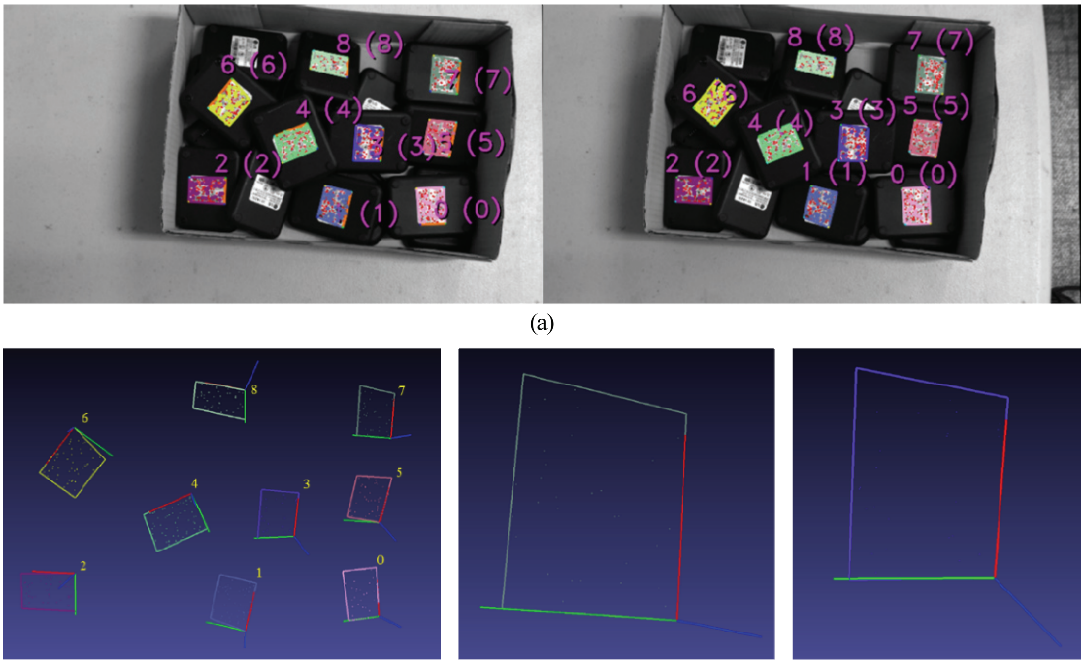
The X, Y, and Z axis vectors of the label coordinate system is used as the label orientation, which will be needed to rotate the robot arm accurately. Fig. 10 shows few examples of label pose estimation results on a 3D viewer.

2.1 Special case

In the proposed bin picking system, one set of target objects contains labels with cut out corner edges as shown in Fig. 11(a). Due to the cut out edge, the label contains five corners, and the proposed label detection method detects only four out of it. For an example, if the four corners detected as in Fig. 11(b) and the origin of the label coordinate system selected as in Fig. 11(c), determination of the Y-axis will be incorrect.

To solve this problem, first, four 3D vectors along each side of the label are defined as shown if Fig. 11(b). Then the angle between two vectors (Equation 16) is used to identify which two facing sides are more parallel compared to other two. If the longer facing sides are more parallel compared to shorter sides, either one of the two vectors go along the longer sides of the label is used to define the X-axis. Then the Y-axis is determined by calculating the cross product between Z-axis (surface normal) and X-axis. If the shorter facing sides are more parallel, first Y-axis is defined using one of the two vectors go along the shorter sides, and the cross product determines the X-axis.

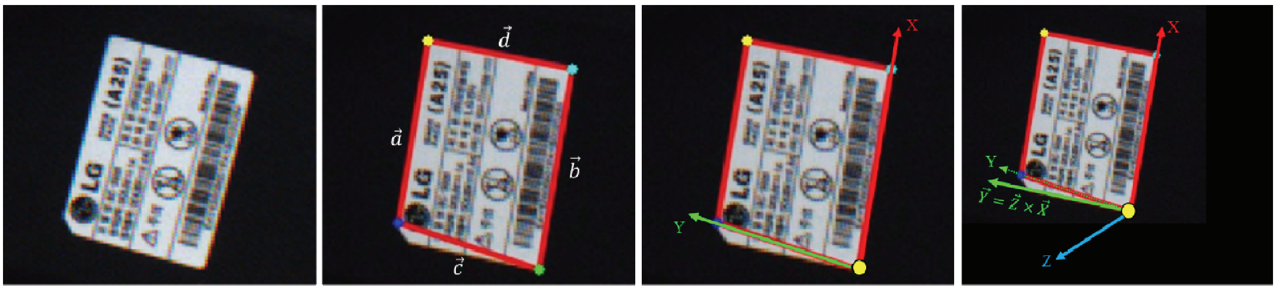
In the example image, longer sides ($\vec{a}$ and $\vec{b}$) are more parallel compared to the shorter sides ($\vec{c}$ and $\vec{d}$). Therefore, first, vector $\vec{b}$ is defined as the X-axis and the Y-axis is corrected as in Fig. 11(d) by finding the cross product between



(a) (b) (c) (d)

그림 10. 라벨의 자세추정 결과 (a) 좌측 및 우측 영상과 정합된 라벨 정보 (b) 라벨들의 방향 정보를 보여주는 3차원 모습 (c) 라벨 7번의 확대 모습 (d) 라벨 3번의 확대 모습.

Fig. 10. Label pose estimation results. (a) Left and right input images with the label matching information. (b) 3D view of the labels with the orientation information. (c) Enlarge view of label 7. (d) Enlarge view of label 3.



(a) (b) (c) (d)

그림 11. (a) 잘려진 모서리를 가진 라벨 (b) 라벨의 네 모서리의 검출 결과 (c) 잘못된 Y축 인식 결과 (d) Z축과 X축을 이용한 Y축의 수정 결과.

Fig. 11. (a) A label with cut out corner edge. (b) Four corners detection result. (c) Incorrect identification of Y-axis. (d) Correction of Y-axis using Z-axis and X-axis.

X and Z-axises.

$$\cos \quad \theta = \frac{\vec{a} \cdot \vec{b}}{\|a\| \|b\|} \tag{16}$$

V. EXPERIMENTAL RESULTS

We implemented the proposed method in a stereo vision system comprising two Point Grey GS2-FW-14S5C-C CCD cameras with 6 mm lenses, which provides images of dimensions 1290×960. The baseline between the two cameras was about 5 cm, and the distance between the target objects and the stereo-camera system was about 1.5 m. We calculated the intrinsic and extrinsic parameters of the stereo camera by calibrating the system using Zhang's camera calibration method [12].

An experiment was conducted to evaluate the distance measurement accuracy of the proposed stereo vision system. A laser rangefinder (Bosch GLM250 VF Professional), which has the measurement accuracy of ±1 mm within the range of 0.05 - 250 m, was used to compare the distance measurement results. As depicted in Fig. 12, the stereo camera, and laser rangefinder was attached to a sliding bar and move it horizontally to measure the distance to a selected object using the both methods. A flat calibration board was used to calibrate the vertical error between the stereo camera and the laser rangefinder as shown in Fig. 13. We placed the calibration board in different locations and measured the distance to a particular point on the pattern board using the laser rangefinder and the stereo system. Average distance difference between the two devices was used as the vertical offset.

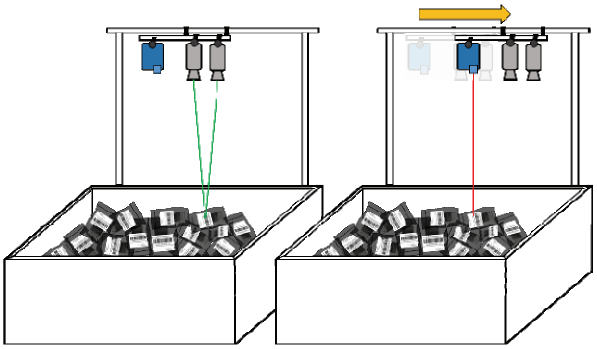


그림 12. 레이저 거리계를 이용한 3D 자세 추정 결과의 정밀도 분석.

Fig. 12. Accuracy analysis of 3D pose estimation results using a laser rangefinder.

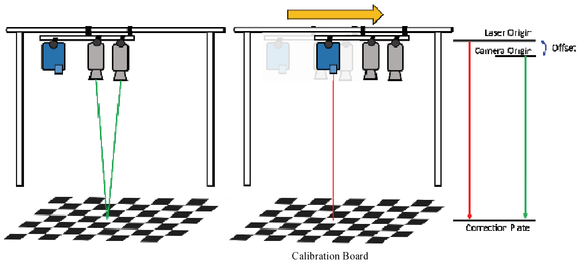


그림 13. 레이저 거리계와 스테레오 카메라 사이의 거리 오프셋 보정.

Fig. 13. Correction of the distance offset between the laser rangefinder and the stereo camera using a calibration board.

The distance measurement accuracy of the proposed system was evaluated with four different object samples, which are shown in Fig. 14. Three rounds of distance measurements tests were conducted for each sample, and 10-15 labels were used in each round. The average measurement errors are given in the Table 1 along with the number of labels used for each test. The graph in Fig. 15 depicts these results along with the standard deviation of the error. For all samples, the proposed system could achieve an average distance error of 0.97 mm, which is good enough for industrial bin picking systems.

Pose estimation accuracy of the proposed system could not be evaluated as no ground truth data is available describing the object pose. Due to the reference bin-picking systems are based on very different approaches and the implementation details are not available, we could not directly compare the proposed method with them.

Finally, we evaluated the processing time of the proposed method by measuring the time used for the two stages; label detection and 3D pose estimation separately. All the computations were performed on a common PC with a 3.9 GHz Core i7 CPU and 8 GB RAM. The OpenCV library was used for image processing and 3D vision tasks. For each sample, we got the average value after measuring the processing time of 20 consecutive frames. The results are given in Table 2. The processing time mainly depends on the number of labels visible on one image frame.

VI. CONCLUSIONS

We proposed an object pose estimation method, which can be used for robotic bin picking systems, employing stereo vision techniques. First, MSER based label detection technique was proposed to identify target objects on the images captured from a stereo camera. A 2D feature matching

(a) Sample 1. (b) Sample 2. (c) Sample 3. (d) Sample 4.

그림 14. 정밀도 분석에 사용된 네 종류의 서로 다른 샘플들.

Fig. 14. Four different types of samples used for accuracy analysis.

표 1. 각 샘플에서 세 번의 실험에 대한 평균 거리 오차. 각 샘플 및 각 실험에 사용된 라벨의 수도 포함.

Table 1. Average distance error for each sample during the three rounds of measurement. Number of labels used for each sample and each round is also given here.

		Sample 1	Sample 2	Sample 3	Sample 4
Test 1	# Labels	14	15	14	12
	Avg. Error (mm)	0.762	1.137	0.570	1.100
Test 2	# Labels	12	15	14	10
	Avg. Error (mm)	0.813	1.737	0.377	1.209
Test 3	# Labels	15	13	12	11
	Avg. Error (mm)	0.668	0.971	0.634	1.707
Average Distance Error		0.748	1.282	0.527	1.339

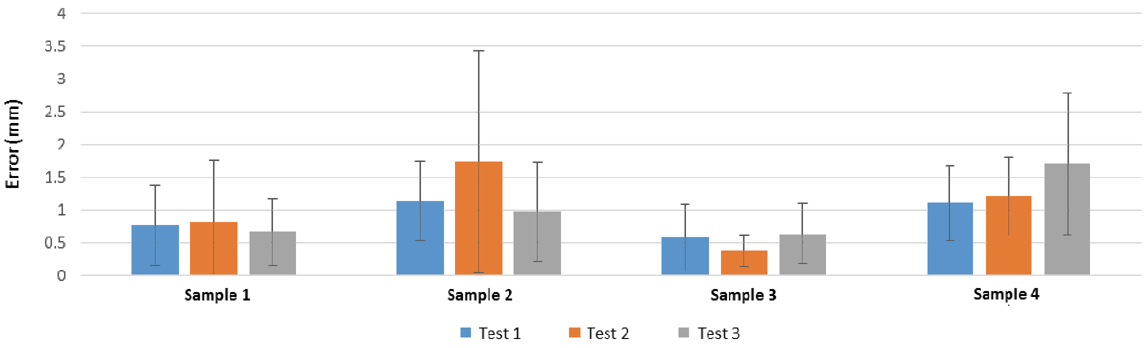


그림 15. 거리측정 오차에 대한 평균값과 표준편차.
Fig. 15. Illustration of the average and the standard deviation of the distance measurement error.

표 2. 한 장의 영상에 대한 평균 연산시간. 첫째 및 둘째 행은 라벨 검출과 자세 추정 과정에 걸린 평균 시간. 모든 결과의 단위는 초(s)임.
Table 2. Average processing time for one image frame. The average time used for the label detection process and the pose estimation process are given in the first and second rows of the table respectively. All the data are given in seconds (s).

	Sample 1	Sample 2	Sample 3	Sample 4
Label Detection	0.36	0.34	0.17	0.21
Pose Estimation	1.12	1.09	0.42	0.43
Total	1.48	1.43	0.59	0.64

technique was performed on these detected labels to determine the label correspondences between left and right camera images. We used a triangulation-based 3D reconstruction method and a plane fitting technique to estimate the 3D plane of the label, which can be used to calculate the position and the orientation of the target object.

We evaluated the distance measurement accuracy of the proposed system using a laser rangefinder, and the results prove that the proposed method could achieve industry level standards with less than 1 mm accuracy. We could achieve nearly one second of average processing time, which is comparable with most of the existing methods, for an image frame with around 10-20 completely visible target objects.

REFERENCES

[1] K. Rahardja and A. Kosaka, "Vision-based bin-picking: recognition and localization of multiple complex objects using simple visual cues," *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, vol. 3, pp. 1448-1457, 1996.
[2] F. Boughorbel, Y. Zhang, S. Kang, U. Chidambaram, B. Abidi, A. Koschan, and M. Abidi, "Laser ranging and video imaging for bin picking," *Assembly Automation*, vol. 23, no. 1, pp. 53-59, 2003.
[3] M. Berger, G. Bachler, and S. Scherer, "Vision guided bin picking and mounting in a flexible assembly cell," *Intelligent Problem Solving. Methodologies and Approaches*, vol. 1821, R. Logananthara, G. Palm, and M. Ali, Eds. Springer Berlin Heidelberg, pp. 109-117, 2000.
[4] I. K. Park, M. Germann, M. D. Breitenstein, and H. Pfister, "Fast and automatic object pose estimation for range images on the GPU," *Machine Vision and Applications*, vol. 21, no. 5, pp. 749-766, 2010.
[5] J. Kirkegaard and T. B. Moeslund, "Bin-picking based on harmonic shape contexts and graph-based matching," *Proc. of International Conference on Pattern Recognition*, vol. 2, pp. 581-584, 2006.
[6] W. Y. Chiu, "Dual laser 3D scanner for random bin picking system," *Proc. of International Conference on Advanced Robotics and Intelligent Systems (ARIS)*, pp. 1-3, 2015.
[7] D. Buchholz, M. Futterlieb, S. Winkelbach, and F. M. Wahl, "Efficient bin-picking and grasp planning based on depth data," *Proc. of IEEE International Conference on Robotics and Automation*, pp. 3245-3250, 2013.
[8] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761-767, Sep. 2004.
[9] M. Donoser and H. Bischof, "Efficient maximally stable extremal region (MSER) tracking," *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, 2006.
[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.
[11] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Univ Press, 2003.
[12] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analisys and Machine Intelligence*, vol. 22, no. 11, pp. 1330-1334, 2000.

Udaya Wijenayake

2010년 스리랑카 Colombo 대학교 컴퓨팅학부 학사졸업. 2013년 경북대학교 대학원 컴퓨터학부 석사졸업. 2013년~현재 경북대학교 대학원 컴퓨터학부 박사과정. 관심분야는 3D복원, 3D의료영상, 3D 모션추적.

최 성 인

2007년 2월 금오공과대학교 컴퓨터공학부 졸업. 2009년 8월 경북대학교 전자전기컴퓨터학부 석사졸업. 2015년 8월 경북대학교 대학원 전자전기컴퓨터학부 박사졸업. 2015년 9월~현재 삼성중공업 연구소 책임연구원. 관심분야는 3차원 정합 및 복원, 로봇비전.

박 순 용

1991년, 1993년 경북대학교 전자공학과 학사, 석사졸업. 2003년 미국 뉴욕주립대학교 스토니브룩 박사졸업. 1993년~1999년 한국원자력연구소 로봇기반연구실 선임연구원. 2004년~2005년 한국전자통신연구원 디지털콘텐츠연구단 선임연구원. 2005년~현재 경북대학교 컴퓨터학부 부교수. 관심분야는 3차원스캐닝, 3차원정합, 로봇비전.