International Journal of Control, Automation, and Systems (2012) 10(2):362-373
DOI 10.1007/s12555-012-0216-9
ISSN:1598-6446 eISSN:2005-4092
http://www.springer.com/12555

Stereo Vision Based Automation for a Bin-Picking Solution

Jong-Kyu Oh, Sukhan Lee*, and Chan-Ho Lee

Abstract: As flexibility becomes an important factor in factory automation, the bin-picking system, where a robot performs pick-and-place tasks for randomly piled parts in a bin through measuring the 3D pose of an object by a 3D vision sensor, has been actively studied. However, conventional bin-picking systems that are employed for particular tasks are limited by such things as the FOV (Field of View), the shape of landmark features, and computation time. This paper proposes a general-purpose stereo vision based bin-picking system. To detect the workpiece to be picked, a geometric pattern matching (GPM) method with respect to the 2D image with a wide FOV is applied. The accurate 3D pose of a selected workpiece among the pick-up candidates is acquired by measuring the 3D positions of three features in the workpiece using the stereo camera. In order to improve the 3D position estimation performance, the GPM method is also used instead of the stereo matching method. The multiple pattern registration and ellipse fitting techniques are additionally applied to increase the reliability. The grasp position of a workpiece without collision is determined using the pose of the object and the bin information. By using these methods a practical bin-picking strategy is established to operate robustly with minimum help from the human workers in the factory. Through experiments on commercial industrial workpieces and industrial robot, we validated that the proposed vision system accurately measures the 3D pose of part and the robot successfully manipulates the workpiece among randomly stacked parts.

Keywords: Bin-picking, industrial robot, robot vision, stereo vision.

1. INTRODUCTION

Recently there have been increasing concerns in regards to a flexible manufacturing system that can adapt automation equipment to varying environments effectively and so improve the productivity of a factory automation system. To establish a flexible manufacturing system in a production site, intelligent robots which can perceive a workpiece within the workspace, recognize a situation, and manipulate themselves autonomously, are essential. An important part of this is accurate visual sensing by the machine vision systems, because these systems enable robots to recognize the environment around the workspace and makes the manufacturing process flexible.

The applications of intelligent industrial robots have broadened from welding to assembly, which requires a more sophisticated control. The material handling process involves an especially high degree of difficulty, and so can have a noticeable impact on the production output rates and costs; fixtures can be removed and thus save the amount of working space required in high-value manufacturing areas. Thus, there are increasing demands from manufacturers for a reliable sensing and an application technology so that robotic systems can be applied to factories [1,2].

In the study of part feeding automation, some research has looked into bin picking systems that perform pick-and-place tasks for randomly oriented parts from bins or boxes [3,4]. Ban proposed a laser vision based bin-picking system [5]. The proposed laser vision sensor consists of a projector which projects a cross pattern and a camera to capture the image. This sensor calculates the X, Y and roll through 2D image processing with respect to the 2D image, and computes the Z, pitch and yaw by analyzing a projected image with the laser pattern. However this system has a narrow sensing range since the 3D pose of an object is calculated only from the surface area where the laser slit is projected.

Bin-picking systems using stereo vision have been widely researched. A study by Rahardja calculates the position and normal vector of randomly stacked parts by means of a stereo camera [6]. However this system needs unique landmark features, which are composed of seeds and supporting features such as holes, to identify the target object and estimate the pose of the object. Berger introduced a stereo system with a grid projector [7]. This system calculates the pose of an object using CAD models, which cannot be provided in real applications.

Commercial real-time stereo cameras which provide a

Manuscript received March 15, 2011; accepted November 21, 2011. Recommended by Editorial Board member Kang-Hyun Jo under the direction of Editor Hyouk Ryeol Choi.

This work was supported in part by the grant from the Industrial Source Technology Development Program of, in part by the Intelligent Robot Program under the Frontier R&D Initiative of, also in part by the ITRC Program NIPA-2012-(C1090-1221-0008) of MKE, Korea, as well as in part by WCU Program (R31-10062-0) and PRCP (2011-0018397) of NRF / MEST, Korea.

Jong-Kyu Oh and Sukhan Lee are with the Intelligent System Research Center of Sungkunkwan University, 300 Cheoncheon-dong, Jangan-gu, Suwon, Gyeonggi-do 440-746, Korea (e-mails: {jkoh, lsh}@ece.skku.ac.kr).

Jong-Kyu Oh and Chan-Ho Lee are with the Electro-Mechanical Research Institute of Hyundai Heavy Industries Co., Ltd., 102-18 Mabuk-dong, Giheung-gu, Yongin-si, Gyeonggi-do 449-716, Korea (e-mails: {jkoh, leechanh}@hhi.co.kr).

* Corresponding author.

dense depth map have been already launched [8,9]. However these systems have a depth estimation error, which is caused by the lack of matching information in homogeneous areas, occluded areas, and noisy regions. Therefore in order to implement a bin-picking system with the use of these stereo cameras, additional range image processing is required to verify and compensate for the false 3D reconstruction.

There are recent studies that use structured-light sensors. Schraft proposed the pose estimation method that compares the 3D models of a workpiece with a range image captured by a laser scanner [10]. To implement this kind of pose estimation method, off-line models encompassing various viewpoints are first registered into a database by rotating a reference object at a constant angle. However this system has disadvantages in that the accuracy and speed of the pose estimation depends on the number of registered models in the database. In addition, active vision sensors cannot guarantee a reliable range image for metallic objects due to reflections; most workpieces in industrial fields are metallic.

In order to tackle the aforementioned problems, in this paper we propose a general-purpose stereo vision based bin-picking system. The GPM with respect to the 2D image with a wide range field of view is applied to detect pick-up candidates without a range image of the bin objects. An accurate 3D pose of the selected workpiece among the pick-up candidates is calculated by measuring the 3D positions of three features with the use of a stereo camera, where they are not collinear on the workpiece. Contrary to the conventional stereo vision system, the GPM method, with respect to the stereo image, is also used instead of the previously mentioned stereo matching method to find the correspondence points in the stereo images. Since this system can calculate the 3D position of the features by just registering the patterns for the features of an object, it can be employed for various general-purpose applications. In addition, a multiple pattern registration and an ellipse fitting techniques are applied in order to increase the reliability against changes in illumination and appearance. The trajectory of the robot to grasp the object without collision is determined by the pose of object and the bin information. Furthermore a bin-picking strategy, in which the robot operates automatically without the intervention of a worker, is established.

The organization of this paper is as follows: Section 2 describes the system overview of the proposed vision guided robotic system. Sections 3 and 4 present the method which detects the pick-up candidates, and calculates the 3D pose of the selected object, respectively. The method which decides the grasp point for an object is illustrated in Section 5. The bin-picking strategy is depicted in Section 6. The experimental results are shown in Section 7 and we conclude the paper in Section 8.

2. SYSTEM OVERVIEW

To establish a reliable bin-picking system, sensors that recognize the working environment and a robot which performs precise manipulations on the workpiece should be incorporated concisely. In this paper, we configured a stereo vision based robotic system incorporating two parts. The first is a vision system which consists of a frame grabber board, two cameras and lenses. The second part is the robot system, which includes the robot controller. The stereo camera can measure the 3D pose of objects within the robot's working area because it is attached to the robot's end-effector instead of to a fixed position in the workspace. The robot controller and the vision system are connected via RS232C or Ethernet, so the position data of the robot and the object can be interactively shared between them.

Fig. 1 describes the architecture of the proposed bin-picking system. It consists of the preparatory stage, which includes the initial setup of the vision system and the definition of the robot's tasks, and the operation stage, in which the system automatically performs the pick-and-place task in the workspace by means of the measured pose information. In the preparatory stage, the operator has to register each pattern of the workpiece, calibrate the sensor, register the 3D pose of the reference object, and assign the trajectory of the robot for the task. In the operation stage, the pick-up candidates among the randomly scattered objects are first selected using the GPM method, which compares the geometric features between the current input image and the pre-registered pattern in the database. The left camera of the stereo pair is the only one used for the detection of the candidates; the robot moves far away from the bin to get an image with a wide range FOV.

According to the position of pick-up candidates, the robot moves into the fine search position. Then an accurate 3D pose of the object is calculated through the determination of a correspondence point in the stereo image, triangulation, and creates an object coordinate. After that the grasp point is determined with the 3D pose of the object and the bin information. Finally the robot modifies its trajectory.

The main technical functions of the proposed bin-picking are the detection of the pick-able candidates, the 3D pose estimation of the part, the determination of the

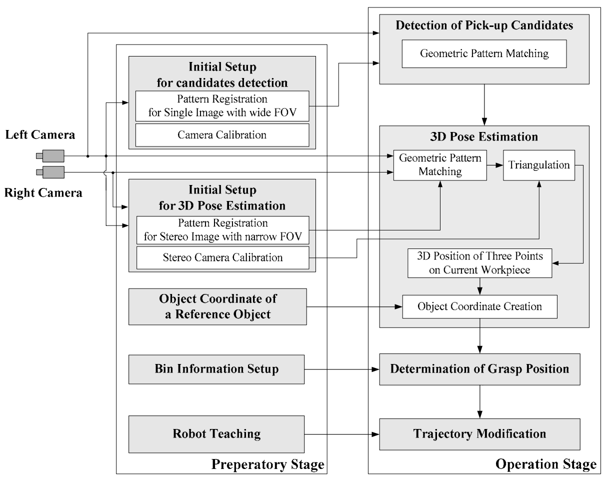


Fig. 1. The diagram of the proposed bin-picking system.

grasp point without collision, and bin-picking strategy; these functions are explained in detail in the following sections.

3. THE DETECTION OF THE PICK-UP CANDIDATES

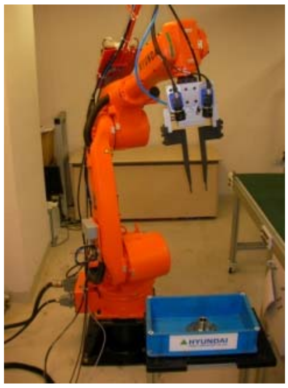
To grasp a part among the stacked objects in a bin, the potential pick-able candidates are preferentially identified. There are some studies that detect the pick-up candidates based on the depth information acquired from a 3D vision sensor [7,11,12]. These studies assumed that the topmost object will be the desired object to be picked from a bin with piled objects. However, it is certain that this method cannot provide reliable results when the 3D point clouds from the sensor are unclear. Besides it requires a large amount of computation time because this method has to reconstruct the whole sensing range.

Considering the computation time and the reliability of the sensing data, in this paper, the potential pick-able parts are found by using the GPM [13]. Contrary to the traditional pattern matching method commonly known as "normalized correlation", the GPM makes use of a set of geometric features instead of the intensity, so it robustly finds the target object despite variations in size, angle, and illumination.

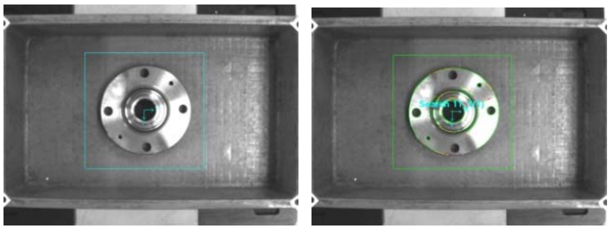
To implement the GPM as a part locator, an input image with wide range FOV, enough to encompass the whole bin, is needed. Generally an additional camera attached on top of the robotic cell is in charge of detecting the potential pick-up candidates [5]. However, this causes an increase in the cost of installing a robotic cell. In this paper, we utilize the left camera of the stereo camera pair attached to the robot, eliminating the additional camera. As focal length of the lenses should not change after the camera calibration process, an image with a broad range FOV can be obtained by increasing the stand-off distance, which means the distance between the stereo camera and objects.

Fig. 2(a) represents the robot's position when the left camera of the stereo pair captures the image used for the detection of the pick-up candidates. The focal length of the lens is 12 mm, and stand-off distance is 890 cm. The commercial automotive front wheel hub is used as a reference object for bin-picking. Fig. 2(b) shows the pattern for matching, and the origin of the pattern and the registered geometric features is described in Fig. 2(c).

Fig. 3 shows a GPM result for the pick-up candidates. Their pick-up priority is assigned by the pattern matching ratio. According to the pick-up priority, an object to be measured in the 3D pose estimation module is selected, and the robot moves to the fine search position, which provides a close-up viewpoint on the selected object, as shown in Figs. 4 and 5. Compared to the robot position for the detection of pick-up candidates, the stand-off distance for 3D pose estimation is short so that the vision system can effectively see the region of interest without obstacles. Therefore the vision system can find key features robustly in the stereo images and the accuracy of the depth estimation is increased. A detailed explanation of the 3D pose estimation is done in the next section.



(a) The robot position.



(b) Pattern registration. (c) Pattern recognition.

Fig. 2. The robot position and pattern recognition for the detection of the pick-up candidates.

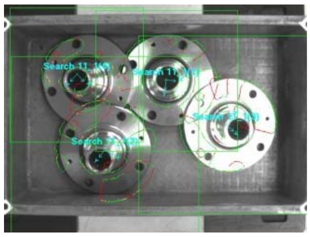


Fig. 3. The detected pick-up candidates among the scattered objects.

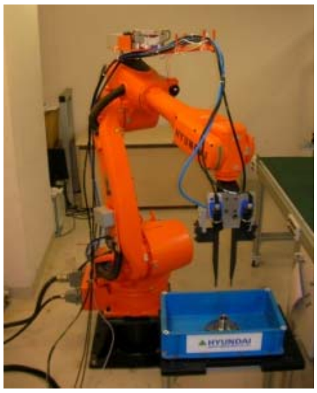


Fig. 4. The robot position for the 3D pose estimation.

4. THE 3D POSE ESTIMATION

In order to handle a workpiece which is distributed in a disordered environment by means of an industrial robot, the 3D position and orientation of the workpiece needs to be accurately computed.

In this paper, the accurate 3D position and the orientation of the object is computed using a stereo camera. To get the 3D information from the stereo images, the camera calibration, the determination of the correspondence points and the creation of object coordinates should be performed in advance.

In the preparatory stage, the camera calibration and the pattern registration are performed. In the camera calibration module, the camera projection matrices of the stereo camera are obtained. In the pattern registration and training module, the patterns which are used as the correspondence points of the stereo images are registered and trained. When in the operation stage, the stereo camera snaps stereo images and finds the patterns by means of the GPM. Then the 3D positions of the workpiece are estimated by using the corresponding points of the stereo images and the projection matrices of the stereo camera. Finally the 3D pose of the object is calculated by creating the object coordinates, which is generated by the 3D positions of three features on the object.

4.1. The pattern registration and training

The 3D reconstruction of an object or an environment is a major research in the field of computer vision; several approaches, such as stereo vision, laser vision, and structured-light techniques, have been developed [14-16].

Stereo vision is well-known method that gets its 3D information through disparity, which is defined as the difference between the correspondence points. The most difficult problem of stereo vision is to find the correspondence points from the stereo images, which is called “stereo matching”. Traditional stereo vision systems, using stereo matching algorithms, have matching error due to the difference in the intensity of the stereo images and lack of matching information. Especially when the part tilts, the intensity and the appearance of the features in the stereo images change significantly, causing false depth estimation. This sensitivity has severely limited the use of stereo vision systems in high accuracy industrial applications.

In this paper, to overcome the traditional problems of stereo matching, correspondence points are determined using the GPM [13] instead of finding the correspondence points by using a stereo matching algorithm. An operator registers the patterns of the left and right images including the key features of the workpiece and trains the registered patterns in the preparatory stage. The patterns are not in collinear because they are used for creating the object coordinates later. In addition, the search area for each pattern is restricted by regions of interest with consideration of the computation time and the recognition performance. Fig. 5 shows the patterns of the left and right image. They consist of one key feature that involves the whole object, and four sub-features, which are circles distributed uniformly on the object. Fig. 6 is the result of the pattern recognition for the stereo images. The image coordinates of the detected patterns in the pattern recognition process are used as the correspondence points of the stereo image.

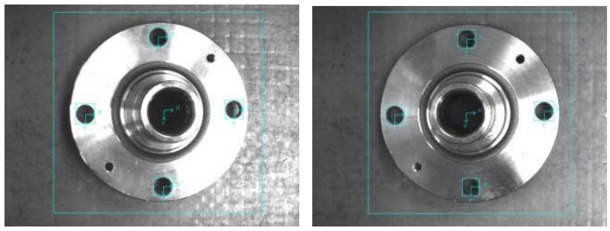


Fig. 5. The patterns of the left and right image.

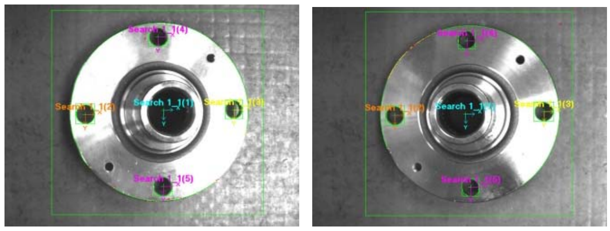


Fig. 6. The results of the pattern recognition for the left and right image.

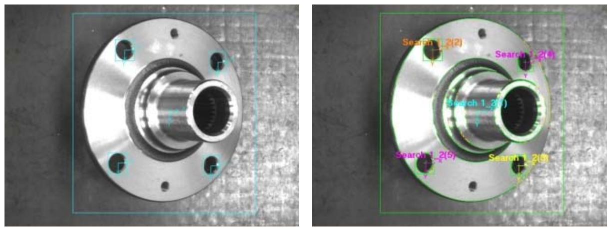


Fig. 7. A sample of the multiple pattern registration, and the result of the pattern recognition for the left image.

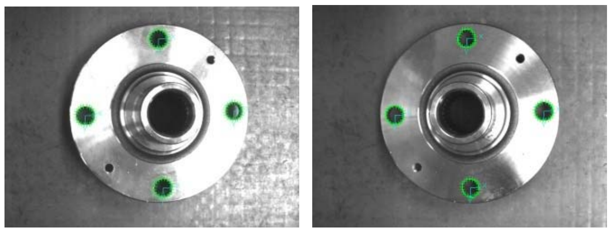


Fig. 8. The result of the ellipse fitting which is applied after the pattern recognition process.

In order to increase the performance of the pattern recognition against an illumination change or a degradation of the appearance, a method that registers multiple patterns for various environments can be applied, as shown Fig. 7. Besides, if the workpiece has circle-type features such as holes, we can additionally employ an ellipse fitting algorithm so that we can accurately calculate the correspondence points of the stereo images. Fig. 8 is the results of the ellipse fitting for the stereo image after the pattern recognition process.

Since we apply robust image processing algorithms, such as the GPM, registration of multiple patterns and ellipse fitting to increase the performance of 3D position estimation, the proposed vision system becomes more reliable than the conventional stereo vision systems.

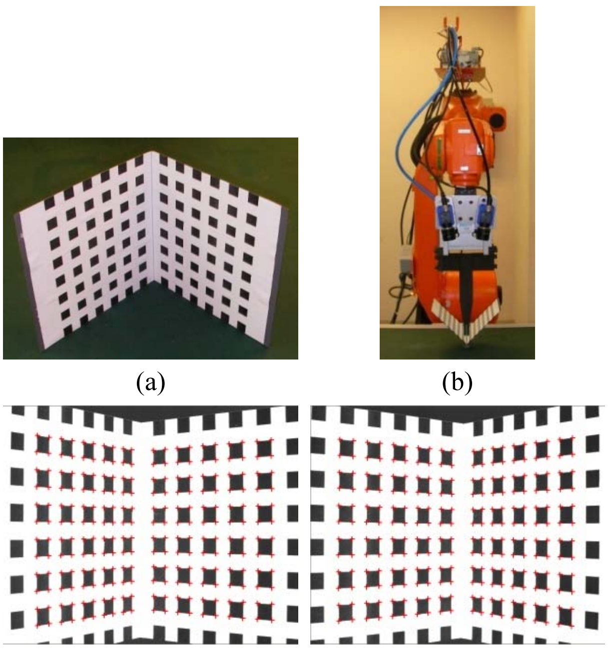
4.2. The camera calibration

A camera calibration is needed to establish the relationship between the 3D coordinates and their corresponding 2D image coordinates. Under perspective projection, the 3D point $\mathbf{X} = [\,X\ Y\ Z\ 1\,]^{\mathrm{T}}$ in space is projected to an image point $\mathbf{x} = [\,u\ v\ 1\,]^{\mathrm{T}}$ in the image plane via a projection matrix $\mathbf{P}$ as:

$$\mathbf{x} = \mathbf{K}[\mathbf{R} \mid \mathbf{T}]\mathbf{X} = \mathbf{P}\mathbf{X}, \tag{1}$$

where $\mathbf{P}$ consists of the intrinsic parameter $\mathbf{K}$ and extrinsic parameter which includes the rotation matrix $\mathbf{R}$ and the translation vector $\mathbf{T}$ [17]. In order to obtain the $\mathbf{P}$ matrix, a calibration jig which can provide 3D position information in unknown environment is required. In this paper we use calibration board composed of square patterns as shown Fig. 9(a).

The procedure of camera calibration is as follows: First, we put the jig within the common field of view in the stereo camera and capture stereo images. Then, the 3D positions of the corner points based on the robot coordinates are assigned by teaching the pre-defined points on the calibration board, as shown in Fig. 9(b). The corner points of the stereo images, which are depicted as the red cross-type points in Fig. 9(c) and (d), are extracted by the Harris corner detector [18]. These are used to make the world coordinates. Finally the camera calibration is performed through the mapping from the known corner points in the 3D space to the corner points in the image plane. The camera calibration is performed just once in the preparatory stage. After the camera calibration process, we acquire the $\mathbf{P}_{\mathrm{L}}$ and $\mathbf{P}_{\mathrm{R}}$, which is the left and right camera projection matrix, respectively.



(a) (b)

(c) (d)

Fig. 9. The stereo camera calibration. (a) The calibration board for the stereo camera, (b) The robot-based 3D position assignment by teaching the robot, (c), (d) Conner points in stereo image.

4.3. The 3D coordinate reconstruction

If the projection matrices of the left and right camera are given, and if the pair of corresponding points in the stereo image through the pattern recognition is given, we can determine the coordinates of the scene point which corresponds to the points in the two image planes.

The relationship between point $\mathbf{X}$ in the scene and its corresponding point $\mathbf{x}_{\mathrm{L}}$ and $\mathbf{x}_{\mathrm{R}}$ in the left and the right image coordinates can be written as follows:

$$\mathbf{x}_{\mathrm{L}} \cong \mathbf{P}_{\mathrm{L}}\mathbf{X}, \tag{2}$$

$$\mathbf{x}_{\mathrm{R}} \cong \mathbf{P}_{\mathrm{R}}\mathbf{X}, \tag{3}$$

where $\mathbf{P}_{\mathrm{L}}$ and $\mathbf{P}_{\mathrm{R}}$ denote the projection matrix of the left and right camera, respectively.

Equations (2) and (3) are expressed in terms of homogeneous coordinates:

$$\mathbf{M}\mathbf{X} = 0, \tag{4}$$

where

$$M = \begin{bmatrix} p_{11}^{L} - u_L p_{31}^{L} & p_{12}^{L} - u_L p_{32}^{L} & p_{13}^{L} - u_L p_{33}^{L} & p_{14}^{L} - u_L p_{34}^{L} \\ p_{21}^{L} - v_L p_{31}^{L} & p_{22}^{L} - v_L p_{32}^{L} & p_{23}^{L} - v_L p_{33}^{L} & p_{24}^{L} - v_L p_{34}^{L} \\ p_{11}^{R} - u_R p_{31}^{R} & p_{12}^{R} - u_R p_{32}^{R} & p_{13}^{R} - u_R p_{33}^{R} & p_{14}^{R} - u_R p_{34}^{R} \\ p_{21}^{R} - v_R p_{31}^{R} & p_{22}^{R} - v_R p_{32}^{R} & p_{23}^{R} - v_R p_{33}^{R} & p_{24}^{R} - v_R p_{34}^{R} \end{bmatrix}, \tag{5}$$

$$\mathbf{X} = [X \quad Y \quad Z \quad 1]^{\mathrm{T}}. \tag{6}$$

Thus, we can estimate the scene point through singular value decomposition (SVD) related techniques. The four elements of the last column of $\mathbf{V}$, obtained by the SVD of $\mathbf{M}$ (i.e., $\mathbf{M}$=$\mathbf{UDV}^{\mathrm{T}}$) are the homogeneous coordinates of $\mathbf{X}$.

4.4. The hand-eye calibration

In vision guided robotic applications, the initial vision set-up such as focal length of the lens and the position of the camera, must be the same for both the preparatory stage and the operation stage. If any of these items are changed, an operator must re-calibrate the vision system. However in bin-picking application, it is natural that the position of camera attached to the robot is always changed according to the distribution of the workpieces. This problem can be solved if the relationship of the camera with respect to the robot end-effector is given, which is referred to as the hand-eye calibration [19].

Fig. 10 represents the coordinate frames used to perform the hand eye calibration, where {B}, {E}, {C}, and {W}are the coordinates of the robot base, the robot end-effector, the camera, and the world, respectively. The relationship between each coordinate can be described by a homogenous transformation matrix.

The relationship between the robot end-effector coordinates and the camera coordinates, which we want to get, can be found in (7). ${}^{\mathbf{E}}\mathbf{H}_{\mathbf{C}}$ can be calculated because ${}^{\mathbf{B}}\mathbf{H}_{\mathbf{E}}$ is provided by the robot controller, ${}^{\mathbf{B}}\mathbf{H}_{\mathbf{W}}$ is acquired by assigning the object's positions by means of the robot, and ${}^{\mathbf{C}}\mathbf{H}_{\mathbf{W}}$ is calculated by the data measured from the

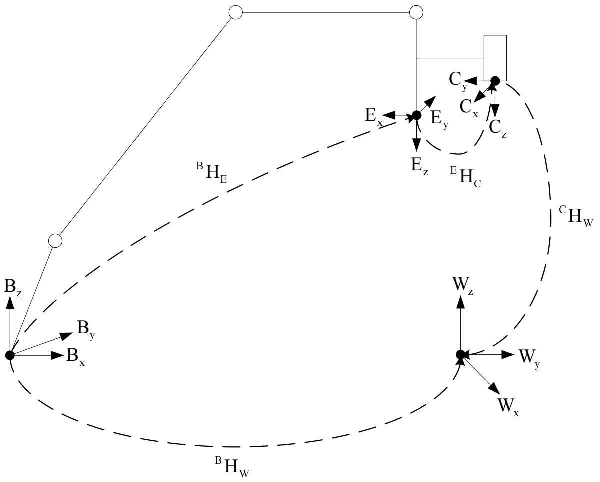


Fig. 10. The coordinate frames for hand-eye calibration.

camera. If $^{\mathbf{E}}\mathbf{H}_{\mathbf{C}}$ is acquired using (7), (8) transforms the 3D position of the object with respect to the camera coordinates into the 3D position of the object with respect to the robot base coordinates wherever the stereo camera mounted on the end-effector of the robot moves. The index *i* in (8) represents any point on the object.

$$^{\mathbf{E}}\mathbf{H}_{\mathbf{C}} = {^{\mathbf{B}}\mathbf{H}_{\mathbf{E}}^{-1}}\, {^{\mathbf{B}}\mathbf{H}_{\mathbf{W}}}\, {^{\mathbf{C}}\mathbf{H}_{\mathbf{W}}^{-1}}, \tag{7}$$

$$^{\mathbf{B}}\mathbf{P}_{\mathbf{i}} = {^{\mathbf{B}}\mathbf{H}_{\mathbf{E}}}\, {^{\mathbf{E}}\mathbf{H}_{\mathbf{C}}}\, {^{\mathbf{C}}\mathbf{P}_{\mathbf{i}}}. \tag{8}$$

4.5. The object coordinate creation

The 3D pose of the object can be estimated by calculating the transformation matrix between the robot base coordinates and the object coordinates. In this paper we define the object coordinates as seen in Fig. 11. F_1, F_2, and F_3 in Fig. 11 are the pre-defined features in the pattern registration module. Their 3D positions are calculated through the previous process. The X axis of the object coordinate is along vector $\overrightarrow{F_1F_2}$. The Y axis lies on the plane of the three features, which is perpendicular to F_1 on the X axis. The Z axis is along the vector defined by the cross product of vectors X and Y. During the preparatory stage, the homogeneous transformation matrix between the robot base coordinates and the reference object coordinates is registered in the database according to (9). The Euclidian distance of each feature is also saved in the database for the validation of the object coordinates in the operation stage.

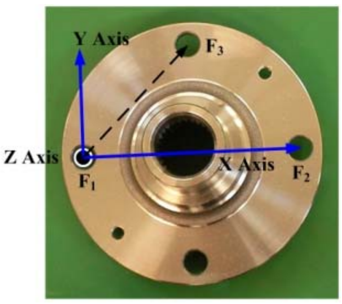


Fig. 11. The workpiece for bin-picking and user coordinate creation.

In the operation stage, the homogeneous transformation matrix between the robot base coordinates and the current object coordinates is computed by using (10), and then the homogeneous transformation matrix between the object coordinates of the current workpiece and that of the reference workpiece is calculated as denoted in (11), which is decomposed into the 6-DOF variance (i.e., t_x, t_y, t_z, r_x, r_y, r_z). It is transmitted via RS-232 or Ethernet to the robot controller. The robot completes the task by modifying the pre-defined trajectory by the amount of the 6-DOF variance.

$$^{\mathbf{B}}\mathbf{H}_{\mathbf{O_{ref}}} = {^{\mathbf{B}}\mathbf{H}_{\mathbf{E}}}\, {^{\mathbf{E}}\mathbf{H}_{\mathbf{C}}}\, {^{\mathbf{C}}\mathbf{H}_{\mathbf{O_{ref}}}}, \tag{9}$$

$$^{\mathbf{B}}\mathbf{H}_{\mathbf{O_{cur}}} = {^{\mathbf{B}}\mathbf{H}_{\mathbf{E}}}\, {^{\mathbf{E}}\mathbf{H}_{\mathbf{C}}}\, {^{\mathbf{C}}\mathbf{H}_{\mathbf{O_{cur}}}}, \tag{10}$$

$$^{\mathbf{O_{ref}}}\mathbf{H}_{\mathbf{O_{cur}}} = {^{\mathbf{B}}\mathbf{H}_{\mathbf{O_{ref}}}^{-1}}\, {^{\mathbf{B}}\mathbf{H}_{\mathbf{O_{cur}}}}. \tag{11}$$

5. THE DETERMINATION OF THE GRASP PPOINT

If the 3D pose of the reference object is registered in the database, we then teach the trajectory of the robot for the bin-picking task with respect to the reference object in the preparatory stage as shown in Fig. 12. In this case, the robot moves in the normal direction of the reference object and clamps the gripper. In the operation stage, we can modify the grasp point for the object to be picked by using the 6-DOF variance between the reference object and the current object. However if an object with a significantly slanted pose is located in the border region of the bin, as shown in Fig. 13(a), it is clear that the robot collides with the bin. In order to prevent the collision with the bin, the robot should modify the grasping pose with the same grasping position. Thus we adjust the trajectory of the robot so as to approach vertically with respect to the plane of the bin in the case that the pick-up object with a significantly slanted pose is located in the border region of the bin as shown in Fig. 13(b). The border region of in the bin is established with the consideration of the part's size and the position of the bin.

In addition, even though we restrict the pose of the grasp point, it is possible for the gripper to collide with the bin or objects. When any collision occurs, collision detection is important to prevent damage to the object and the robot components. In this paper, not only the collision detection function of the robot controller but also that of an on-line monitoring system is applied [20].

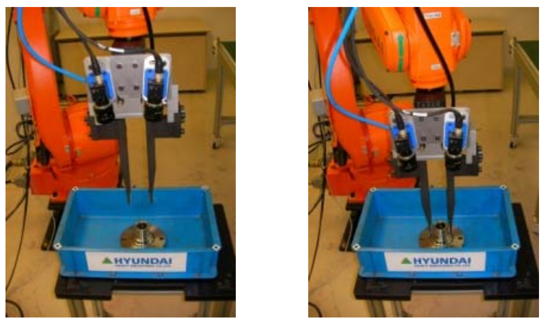
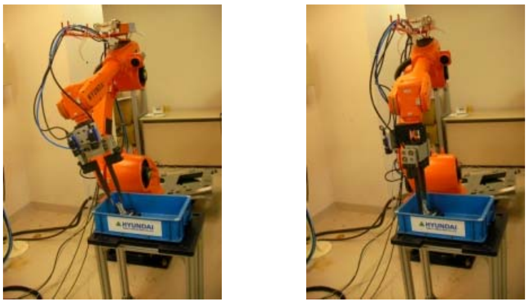


Fig. 12. The trajectory of the robot is taught.



(a) A collision between the robot components and the bin.
(b) The modification of the grasping pose.

Fig. 13. A collision can be avoided by modifying the grasp pose with the use of the poses of the object and bin information.

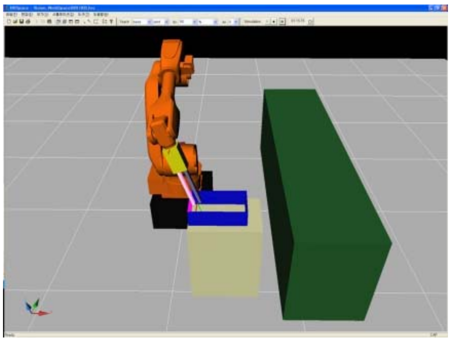


Fig. 14. The on-line monitoring program can observe the collision between the robot component and the bin.

An operator makes the real workspace into the virtual environment off-line, and then the trajectory of robot in the virtual realistic environment can be observed while on-line. Consequently it can inform the operator whether the robot components and bin collide with each other or not. Fig. 14 shows the virtual workspace which is created through the on-line monitoring program, and the highlighted regions of the gripper in Fig. 14 describe the components when collisions occur.

6. THE BIN-PICKING STRATEGY

The robot programming which defines a robot's tasks and its operation procedure is important to determine flexibility in a robot system. To increase the flexibility and productivity of robotic applications, a strategy which takes into account the workflow of the robotic system needs to be defined well.

For a bin-picking system, it is important for the robot to manipulate parts automatically, without stopping the process. In typical vision-guided robotic application, whenever the vision system cannot measure the part's location, the operator has to stop the production line and deal with the problem manually. It reduces productivity of the production line. Thus, in this paper, we designed the strategy to grasp the workpieces one by one automatically from randomly stacked objects, as shown Fig. 15. Even though the stereo camera cannot successfully calculate the 3D pose of a pick-up candidate part, the robot can change the viewpoint of the stereo camera by rotating the end-effector with respect to the X or Y axis. Then it retries to compute the pose of the object. Nevertheless, when it cannot estimate the pose of an object, the robot moves to the next image acquisition position needed for other pick-up candidates. It then tries to calculate the 3D pose for it. The procedure is performed automatically without any help from a human operator excluding a collision or having no more workpieces to be grasped.

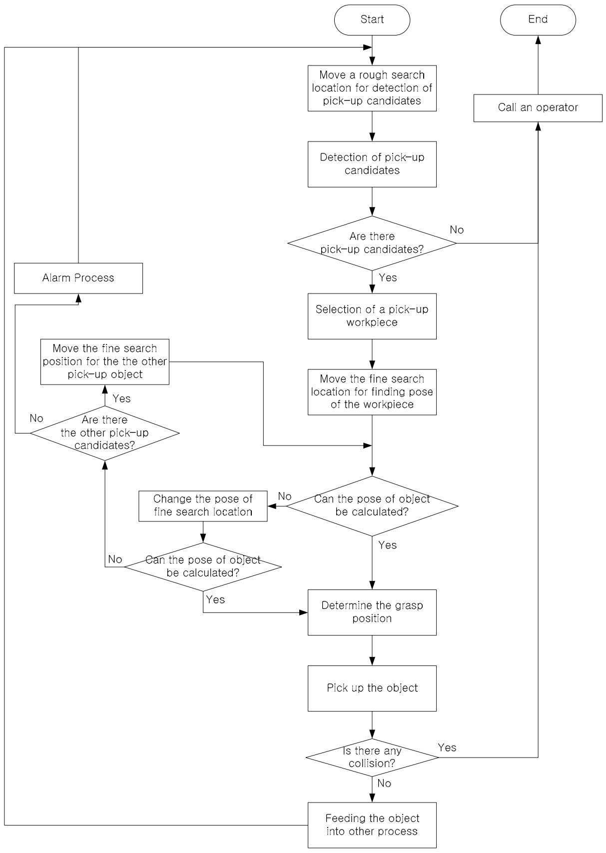


Fig. 15. The flow chart of the proposed bin-picking strategy.

7. THE EXPERIMENTAL RESULTS

For our experiments, two SONY analog cameras, two lenses with focal lengths of 12 mm, and a COGNEX frame grabber board were used for the stereo vision system. The base line of the stereo camera was about 10 cm, and the resolution of the stereo images was 640 × 480. The stand-off distance for the detection of the pick-up candidates and the 3D pose estimation was about 0.5 m and 0.9 m, respectively. A Hyundai HA020 robot was used for the bin-picking task. It picked commercial workpieces such as front wheel hubs, brake disks, and automotive connecting rods from randomly stacked environments.

As we explained in the previous sections, the initial setup of the vision system and the trajectory assignment for the robot's tasks were done in advance. We conducted the quantitative and qualitative experiments as follows.

7.1. The quantitative experimental results

To evaluate the performance of the proposed system, we compare the robot's amount of movement between the actual robot moves, which is the ground-truth, and the estimate from the stereo camera attached on the robot arm as illustrated in Fig. 16 [21]. The real robot movement amount is easily calculated with the use of (12) because the current and the reference robot position can be obtained through the robot controller.

$$\left({}^{\mathbf{E_{ref}}}\mathbf{H_{E_i}}\right)_{\text{ground}} = {}^{\mathbf{B}}\mathbf{H}^{-1}_{\mathbf{E_{ref}}}\ {}^{\mathbf{B}}\mathbf{H_{E_i}}. \quad (12)$$

The robot movement amount estimated by the stereo camera is calculated by (13), where the pose of the object with respect to the end-effector ${}^{\mathbf{E}}\mathbf{H_O}$ can be computed by using ${}^{\mathbf{E}}\mathbf{H_C}$ and ${}^{\mathbf{C}}\mathbf{H_O}$.

$$\left({}^{\mathbf{E_{ref}}}\mathbf{H_{E_i}}\right)_{\text{est}} = {}^{\mathbf{E_{ref}}}\mathbf{H_O}\ {}^{\mathbf{E_i}}\mathbf{H}^{-1}_{\mathbf{O}}. \quad (13)$$

Whenever we move the robot to a different position and rotation, the stereo vision system estimates the 3D pose of the workpiece. We can then evaluate the accuracy of the proposed system by comparing $({}^{\mathbf{E_{ref}}}\mathbf{H_{E_i}})_{\text{ground}}$ and $({}^{\mathbf{E_{ref}}}\mathbf{H_{E_i}})_{\text{est}}$.

Accuracy tests on the translation and rotation with respect to the X, Y and Z axis were performed. The translation error was estimated every time the robot moved five millimeter with respect to the reference position from -50 mm to 50 mm along each axis. The rotation error was calculated every time the robot rotated one degree with respect to reference position from -20° to 20° along each axis. The statistical results on the translation and rotation are analyzed in Tables 1 and 2.

Through the experimental results, we confirmed that the proposed system has an average translation and rotation error below 1 mm and 0.3°, respectively. The maximum error of 2.5 mm and 1.5° was generated when the object was located outside of the image due to the lack of compensation caused by lens distortion. However the maximum error of this system in real applications can be reduced, because the target object is always located in central region of the stereo images during the 3D pose estimation process. The results are sufficient for these kinds of bin-picking applications. In addition, when the calibration is compensated for the lens distortion more accurate results will be acquired [22].

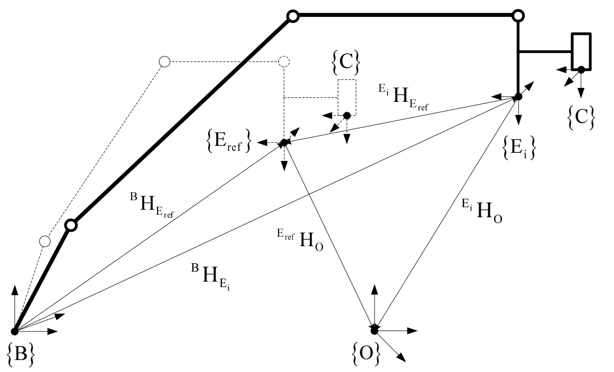


Fig. 16. The coordinate frames for the quantitative experiment.

Table 1. The pose estimation error with respect to the translation (unit: mm, degree).

Statics	X	Y	Z	RX	RY	RZ
Average	0.4958	0.5081	0.4446	0.2944	0.1248	0.0364
Standard Deviation	0.4454	0.5918	0.5186	0.3690	0.0781	0.0326
Max	2.2278	2.3663	2.2479	1.4844	0.3792	0.1178

Table 2. The pose estimation error with respect to the rotation (unit: mm, degree).

Statics	X	Y	Z	RX	RY	RZ
Average	0.4101	0.4555	0.9756	0.2953	0.1035	0.0571
Standard Deviation	0.3384	0.2909	0.4361	0.2178	0.0917	0.0430
Max	1.3968	1.161	1.9171	0.8106	0.4755	0.1686

7.2. The qualitative experimental results

To verify the reliability of the proposed bin-picking system, we confirmed whether it can perform pick-and-place tasks for three commercial workpieces under different environments. Since the front wheel hub and brake disk have circular features, as shown Fig. 17 and Fig. 23, respectively, the ellipse fitting process was performed. Alternatively, the connecting rod does not contain circles, as shown in Fig. 24, so only the GPM was performed to

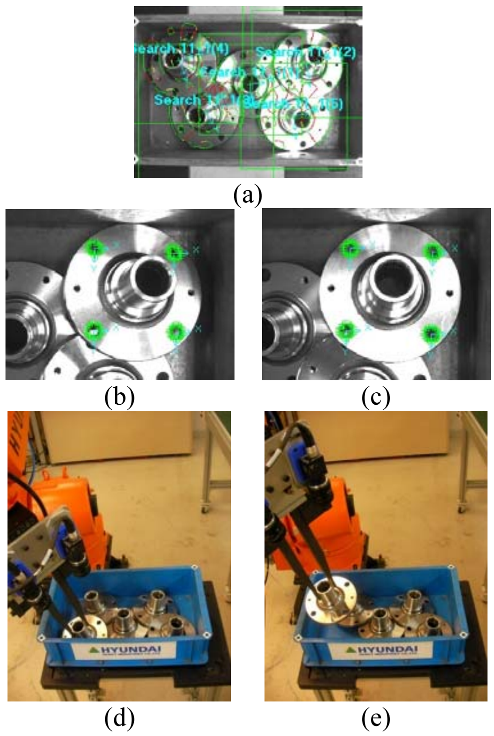


Fig. 17. The bin-picking results of the first trial for the front wheel hub. (a) The detected pick-up candidates of the first trial, (b), (c) The results of the ellipse fitting in the left and right images of the first trial, (d), (e) The grasping of the part in the first trial.

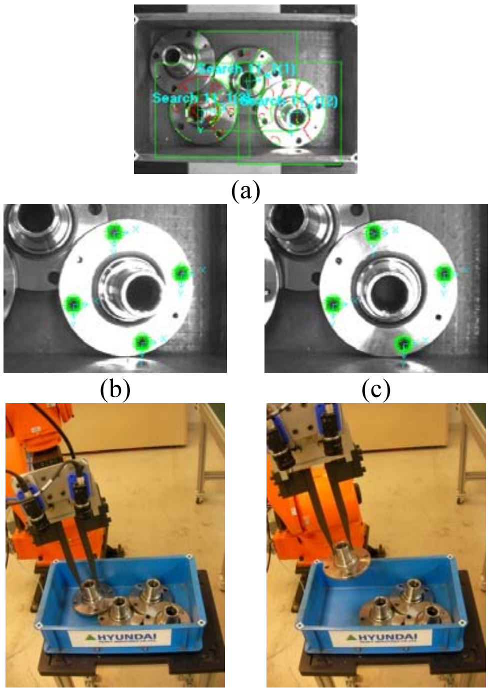


Fig. 18. The bin-picking results of the second trial for the front wheel hub. (a) The detected pick-up candidates of the second trial, (b), (c) The result of the ellipse fitting in the left and right images of the second trial, (d), (e) The grasping of the part in the second trial.

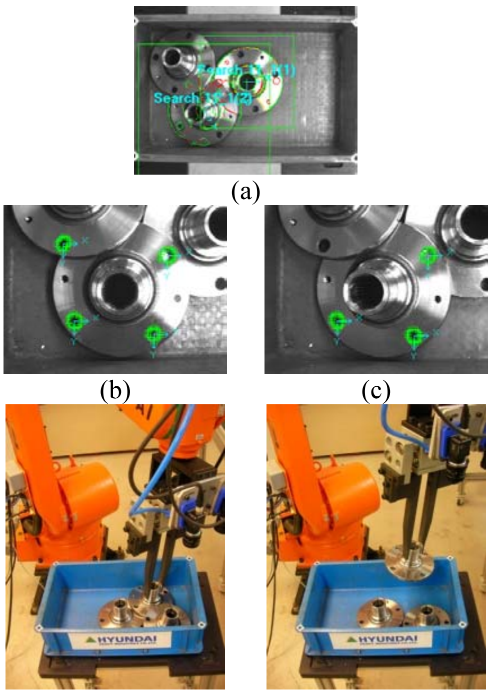


Fig. 19. The bin-picking results of the third trial for the front wheel hub. (a) The detected pick-up candidates of the third trial, (b), (c) The result of the ellipse fitting in the left and right images of the third trial, (d), (e) The grasping of the part in the third trial.

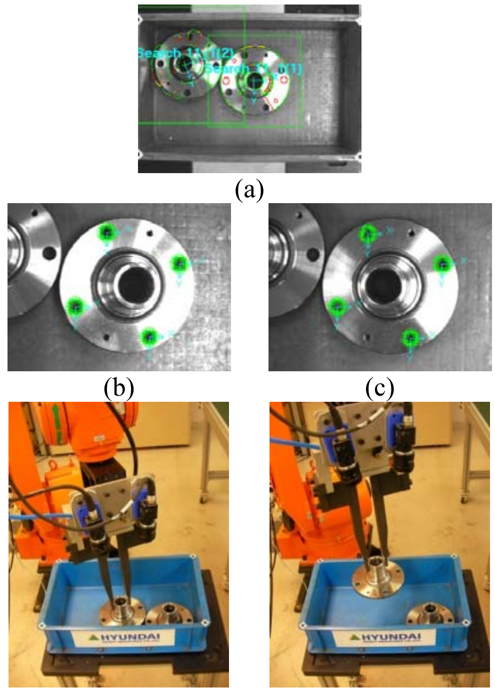


Fig. 20. The bin-picking results of the fourth trial for the front wheel hub. (a) The detected pick-up candidates of the fourth trial, (b), (c) The results of the ellipse fitting in the left and right images of the fourth trial, (d), (e) The grasping of the part in the fourth trial.

find the correspondence points in the tereo images. The results of the bin-picking application for the front wheel hub are shown in Fig. 17 to Fig. 21 for every trial. The (a) figure in every trial describes the detected pick-up candidates, and the (b) and (c) figures in every trial are the results of the ellipse fitting in the stereo images after the pattern recognition. The scene that grasps the measured part is shown in the (d) and (e) figures in every trial. From the results, the vision system accurately detected the center point of the ellipse in the stereo images. Therefore the pose of object could be estimated precisely. Even though the target object is partly occluded by the other objects, as shown in Fig. 19, that is, if only some of the features for obtaining the object coordinates could be detected, this system could calculate the pose of object. In case that objects are stacked on each other, as shown in Fig. 22, it is possible to detect features of the other object. It should be rejected through the verification of the object coordinates using the Euclidian distance among features in the database. The experiment on a large object, such as brake disk, is illustrated in Fig. 23. The results show that the size of the object does not have an effect on the performance of the system. The experiment on the connecting rod is depicted in Fig. 24. Even though the object does not have circular features, the proposed bin-picking system could calculate the 3D pose of the object so that the robot grasped the part successfully. Through these experiments using an industrial robot, we confirmed that the proposed system can be employed for bin-picking applications without a collision.

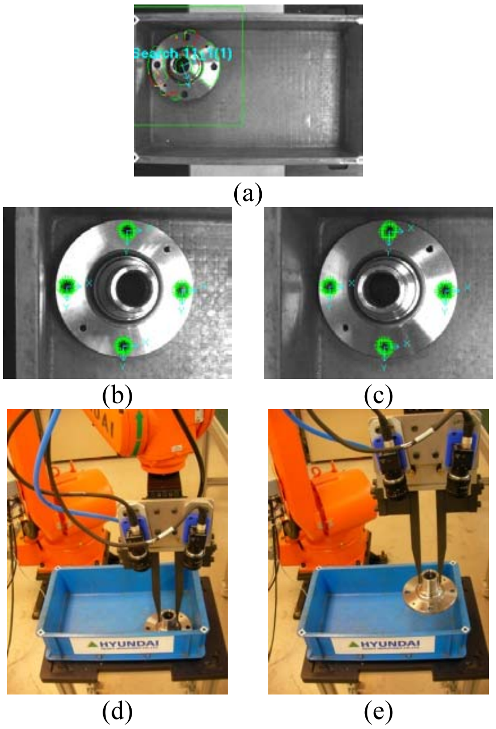


Fig. 21. The bin-picking results of the fifth trial for the front wheel hub. (a) The detected pick-up candidates of the fifth trial, (b), (c) The result of the ellipse fitting in the left and right images of the fifth trial, (d), (e) The grasping of the part in the fifth trial.

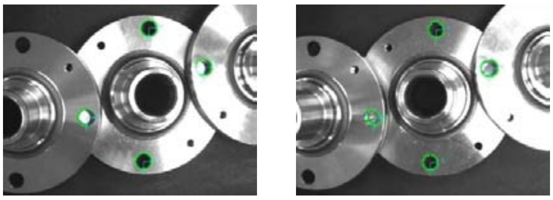


Fig. 22. An example of the verification of the object coordinates using the Euclidian distance feature.

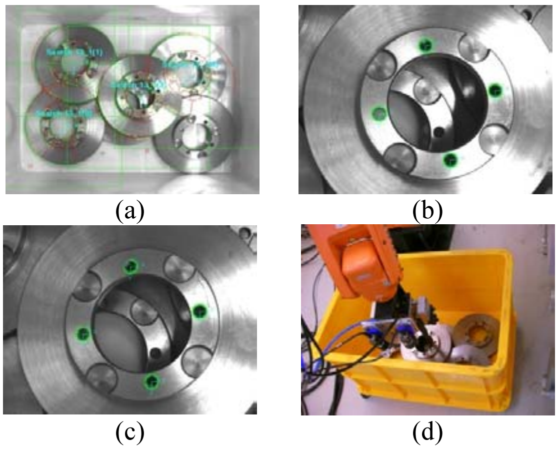


Fig. 23. The results of the bin-picking application for the brake disk. (a) The detected pick-up candidates of the first trial, (b), (c) The results of the ellipse fitting in the left and right images of the first trial, (d) The grasping of the part in the first trial.

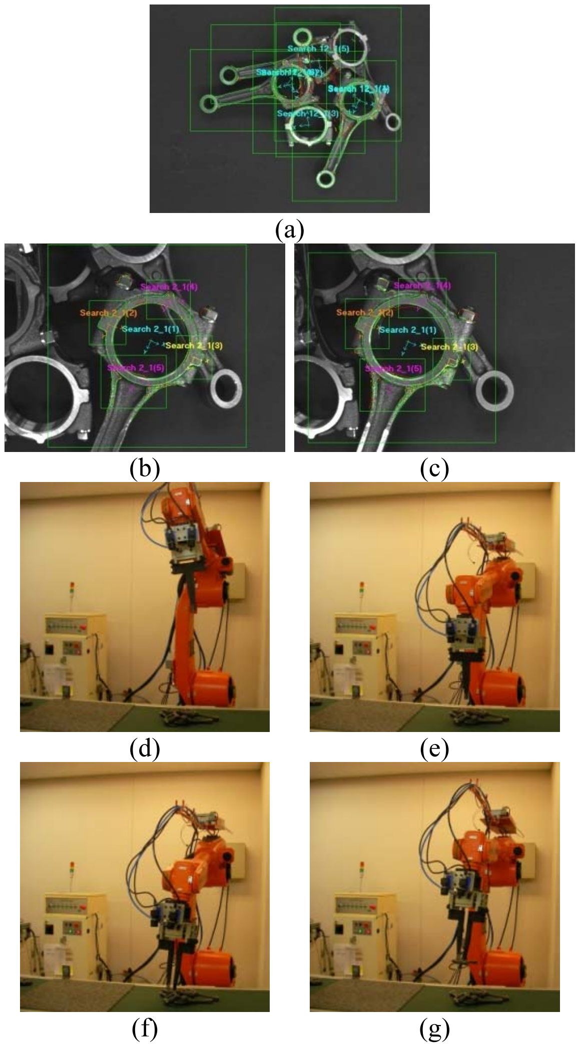


Fig. 24. The results of the bin-picking application for the connecting rod. (a) The detected pick-up candidates of the first trial, (b), (c) The result of the pattern recognition in the left and right images of the first trial, (d), (e) The robot positions for the detection of the pick-up candidates and the 3D pose estimation, (f), (g) The grasping of the part in the first trial.

Fig. 25. The exhibition of the proposed system at ROBOT WORLD 2010, and 2011.

Besides, the proposed system was exhibited at ROBOT WORLD 2010, and 2011 as shown in Fig. 25, which had been held in Korea [23]. During the exhibition, the system was operated successfully without a trivial problem. Through the exhibition, we could confirm the reliability of the proposed system.

8. CONCLUSIONS

This paper proposed a general-purpose stereo vision based bin-picking system. The GPM with respect to the 2D image with a wide range FOV was applied to detect the pick-up candidates. An accurate 3D pose of the selected workpiece among the pick-up candidates was calculated by measuring the 3D positions of three features in an workpiece with the use of a stereo camera.

In order to improve the 3D position estimation performance, the GPM method was also used instead of the stereo matching method to find the correspondence points in the stereo images. Additional functions, such as the multiple pattern registration and ellipse fitting, were also applied to increase the reliability. In contrast with the conventional bin-picking system, it could easily calculate the 3D pose of the target object by just registering the patterns with respect to the features in the object. Therefore it could be employed in various general-purpose applications without the limitations of the shape and the material of the object. The collision-free grasp position of the workpiece was determined using the pose of the object and the bin information. In addition, a bin-picking strategy was established to enable automatic operation with the minimum help of a human worker in a factory.

Through experiments on various commercial industrial workpieces, we validated that the proposed bin-picking system accurately measured the 3D pose of a part and could successfully manipulate the workpiece among the randomly stacked objects.

In the near future, we will do further research on the camera calibration with the goal of compensating for lens distortion, and a structured-light technique for dealing with a textureless object.

REFERENCES

[1] B. Habibi, "Single camera 3D™(SC3D™) a novel technology for guidance of industrial robots in three-dimensional space," *Braintech*, 2002.

[2] J.-K. Oh, C.-H. Lee, S.-H. Lee, J.-S. Hur, and S. Lee "Development of a 3D vision guided robotic system using stereo vision," *Proc. of the 39th International Symposium on Robotics*, pp. 2-7, October 2008.

[3] W. Hardin, "Vision enables freestyle bin-picking," *Vision System Design*, Jun. 2007.

[4] W. Iversen, "Vision-guided robotics: in search of the holy grail," *Automation World*, February 2006.

[5] K. Ban, F. Warashina, I. Kanno, and H. Kumiya, "Industrial intelligent robot," *FANUC Tech. Rev.*, vol. 16, no. 2, pp. 29-34, 2003.

[6] 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*, pp. 1448-1457, Nov. 1996.

[7] M. Berger, G. Bachler, and S. Scherer, "Vision guided bin picking and mounting in a flexible assembly cell," *Proc. of the 13th International Conference on Industrial & Engineering Applications of Artificial Intelligence and Expert Systems*, pp. 109-118, 2000.

[8] Point Grey Research Inc., http://www.ptgrey.com

[9] Videre Design LLC, http://www.videredesign.com

[10] R. D. Schraft and T. Ledermann, "Intelligent picking of chaotically stored objects," *Assembly Automation*, vol. 23, no. 1, pp. 38-42, 2003.

[11] J.-K. Oh and C.-H. Lee, "Development of a stereo vision based position measurement algorithm for the randomly piled objects," *Proc. of International Conference on Advanced Robotics*, pp. 64-69, August 2007.

[12] C. R. Hema, M. P. Paulraj, R. Nagarajan, and Y. Sazali, "Segmentation and location computation of bin objects," *International Journal of Advanced Robotic Systems*, vol. 4, no. 1, pp. 57-62, 2007.

[13] Cognex Corporation, "PatMax and PatQuick," *Cognex*, 2005.

[14] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International Journal of Computer Vision*, vol. 47, no. 1/2/3, pp. 7-42, 2002.

[15] 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.

[16] F. Blais, "Review of 20 years of range sensor development," *Journal of Electronic Imaging*, vol. 13, no. 1, pp. 231-243, 2004.

[17] R. Hartley and A. Zisserman, "Multiple view geometry in computer vision," *Cambridge Press*, 2000.

[18] C. J. Harris and M. Stephens, "A combined corner and edge detector," *Proc. 4th Alvey Vision Conf., Manchester*, pp. 147-151, 1988.

[19] R. Horaud and F. Dornaika," Hand-eye calibration," *Int. Journal of Robotics Research*, vol. 14, no. 3, pp. 195-210, 1995.

[20] Hyundai Heavy Industry Co., Ltd., *Hi5 Robot Controller Manual*, Hyundai Heavy Industry, 2008.

[21] P. Mittrapiyanuruk, G. N. DeSouza, and A. C. Kak, "Calculating the 3D-pose of rigid objects using active appearance models," *Proc. of the Int. Conference in Robotics and Automation*, pp. 5147-5152, 2004.

[22] D.-J. Kang, J.-E. Ha, and M.-H. Jeong, "Detection of calibration patterns for camera calibration with irregular lighting and complicated backgrounds," *International Journal of Control, Automation*, and Systems, vol. 6, no. 5, pp. 746-754, 2008.

[23] http://www.robotworld.or.kr

Jong-Kyu Oh received his B.S. and M.S. degrees in Electronic Engineering from Pusan National University in 1998 and 2000, respectively. He received his Ph.D. degree in Electrical and Computer Engineering from Sungkyunkwan University in 2012. Since 2000, he has been worked with Hyundai Heavy Industries Co., Ltd. His research interests include 3D vision, robot vision, and medical robot.

Sukhan Lee received his B.S. and M.S. degrees in Electrical Engineering from Seoul National University in 1972 and 1974, respectively. He received his Ph.D. degree in Electrical Engineering from Purdue University, West Lafayette in 1982. From 1983 to 1997, he was with the Departments of Electrical Engineering and of Computer Science at the University of Southern California and, from 1990 to 1997, with the Jet Propulsion Laboratory, California Institute of Technology, as a Senior Member of Technical Staff. From 1998 to 2003, he was an Executive Vice President and a Chief Research Officer at the Samsung Advanced Institute of Technology. Since 2003, he has been a Professor of Information and Communication Engineering and WCU professor of Interaction Science at the Sungkyunkwan University, while serving as the Director of the Intelligent Systems Research Institute. Since 2011, he assumed the Dean of the Graduate School of Sungkyunkwan University. Prof. Sukhan Lee has his research interest in the areas of Cognitive Robotics, Intelligent Systems, and Micro/Nano Electro-Mechanical systems. He is currently a fellow of IEEE and a fellow of Korean National Academy of Science and Technology.

Chan-Ho Lee received his B.S. and M.S. degrees in Electronic Engineering from Hanyang University in 1986 and 1991, respectively. Since 1991, he has been worked with Hyundai Heavy Industries Co., Ltd. His research interests include 3D vision, robot vision, and medical robot.