Vision Guided Bin Picking and Mounting in a Flexible Assembly Cell

Martin Berger, Gernot Bachler, and Stefan Scherer

Computer Graphics and Vision, Graz University of Technology
{berger,bachler,scherer}@icg.tu-graz.ac.at

Abstract. In this contribution a vision system for the flexible assembling of industrial parts is presented. A new three step approach is described. It consists of three independent vision guided modules. The picking module allows to pick objects from an unorganized heap or out of a bin, the pose determination module delivers the exact position of the isolated object and the surveillance module allows to verify the success of mounting the parts. This allows all the system stages to consist of standard components, while ensuring a high degree of flexibility, adaptability and robustness. Successfull results achieved with a prototype system implemented at our industrial cooperating partner are presented.

Keywords: Bin Picking, Industrial Assembling, Vision, CAD Model Fitting

1 Introduction

The task of picking objects from an unorganized bin for industrial production purposes has been a great challenge in vision and automation research over the past ten years [16,17,1,12]. The goal was to enable a robot arm to pick some parts out of a bin without any knowledge of their shape or position. As in many cases the pieces to be assembled do not come in a bin or are well organized in appropriate containers, this topic became less significant because of the missing prospective to come into operation. Nevertheless for rather small series (typically thousands of pieces) it becomes very expensive to adapt a production line to new parts. Thus, a device being able to manage the stated task, would help to save money and time. Simplifying the hardware, using standard components instead of special ones, enclosing modules of intelligent vision software is what we propose in the present contribution. Previous approaches tried to solve the task of picking and mounting in one step, yielding complex systems and very sophisticated algorithms, while mainly suffering from flexibility. We introduce a novel three-step concept consisting of bin picking, pose determination and mounting surveillance. In the first step no knowledge about the parts is needed, step two and three will suppose given simplified CAD models. The work was carried out in cooperation with local industrial partners in the field of automation, therefore another goal is to ensure the real time capability of the developed algorithms.

R. Loganantharaj et al. (Eds.): IEA/AIE 2000, LNAI 1821, pp. 109–117, 2000.

The paper is organized as follows. Sect. 2 describes previous and related research, Sect. 3 introduces the new approach in detail. The system modules are described in Sect. 4 and Sect. 5. A description of the working environment is given in Sect. 6, together with some practical application remarks. Sect. 7 concludes the paper and gives a brief outlook to future work.

2 Related Work

The problem of picking parts from a heap was addressed by various authors. Rahardja and Kosaka [16] extract simple features such as circular and polygonal parts from stereo images representing complex objects. Their system incorporates both object identification and pose determination in one single step.
The work of Ikeuchi [12] is based on a CAD representation of an object which allows to build an interpretation tree offline and to select the optimal features at each determining process. The algorithms do not emphasize any type of backtracking.
Trobina and Leonardis [17] present an object-grasping system based on range images. They detect antipodal planar patches where a robot arm can apply with a gripper. The information is extracted piecewise by a Recover-and-Select paradigm. After the selection of the best grasping hypothesis, an object is removed from the pile. No investigations about the precision of grasping or the assembling capability of the system are presented.
Al-Hujazi and Sood[1] propose to determine grip points for a vacuum gripper from dense range images. They detect edges by the residual analysis, segment the (single) object in an appropriate number of surface patches and then calculate a grip point.
Pose determination of objects in 3D was considered as a mandatory byproduct of object recognition systems. The task of determining the pose of objects in 3D by fitting some (even parametric) CAD models to images was first addressed by Lowe [14], based on some work on perceptual organization and grouping [13]. The algorithm refines the 3D pose parameters according to the observed errors in a 2D image. Araujo et al. [2] proposed extensions to Lowe's algorithm. Other authors addressed the mathematical analysis of the 'point cloud to point cloud' matching [9] and mapping in presence of noise given the correspondences [11,10]. Other known recognition (and therefore pose determination) systems in a calibrated environment are ACRONYM (Brooks [7]), HYPER (Ayache and Faugeras [3]) and SCERPO (Lowe [13]). In the paper which introduced the RANSAC[1] paradigm, Fischler and Bolles [8] also presented an early recognition system. The establishment of distance metrics for the alignment of models to images was addressed by Wineshall and Basri [19].

[1] Random Sample Consensus

3 A Three-Step Concept

In contrast to the stated approaches, we split the complex task of flexible automated assembling into three consecutive subtasks. They are almost indepedent among one another, to guarantee highest flexibility and robustness together with easy adaptability. The three stages of our approach are:

– **Bin Picking.** To enable highest flexibilty in picking a wide variety of objects, we try to identify planes on a heap of unorganized parts, where a *vacuum gripper* can apply. A stereo setup together with a grid projector is placed over the working region. The robot (or an appropriate manipulator) picks an object identified to lie on top and drops it on a separate workplace. The algorithm is based on a structured light approach combining shape analysis and stereo matching. Note that no further assumptions on the shape or on the pose of the objects are made.
– **Pose Determination.** Once an object is separated from an unorganized heap, its exact pose is determined, now including some knowledge about it. The first submodule calculates a pose guess, the second fits a simple CAD model iteratively to image features. It is also possible to reject an object if it does not belong to the a given class or if it exceeds some tolerances in shape. A manipulator equipped with a grasp tool can pick up the piece suitably and mount it.
– **Assembly Inspection.** The third stage of our approach ensures the correct mounting of the part and detects failures. This step is done in traditional way for now and is subject to future research.

The following sections will describe the bin picking and the pose determination module in detail.

4 Plane Detection, Grip Point Selection and Picking

This section will outline the image processing steps to robustly detect planar surfaces of parts on a heap and calculate their location in 3D in order to perform the picking. Fig. 1 shows a grayscale image pair of a heap of chain parts and a regular, high contrast grid projected on it.

The grid is segmented with a local threshold method and morphologically thinned to obtain a grid skeleton, which is analyzed and stored in a locally perspective projection distortion invariant representation, as can be seen from Fig. 2. Adjacent intersections are represented by accordant adjacent points in the plane representation (see [4] for details). Connected intersections are called *plane candidates*.

Next the plane candidates from the left and the right image are matched, missing intersections are detected and false matches are rejected. Once correspondences have been established, the subpixel positions of the intersections are calculated from the grayscale images, starting from the position achieved from the segmented and skeletonized grid. Then the gridpoints are reconstructed

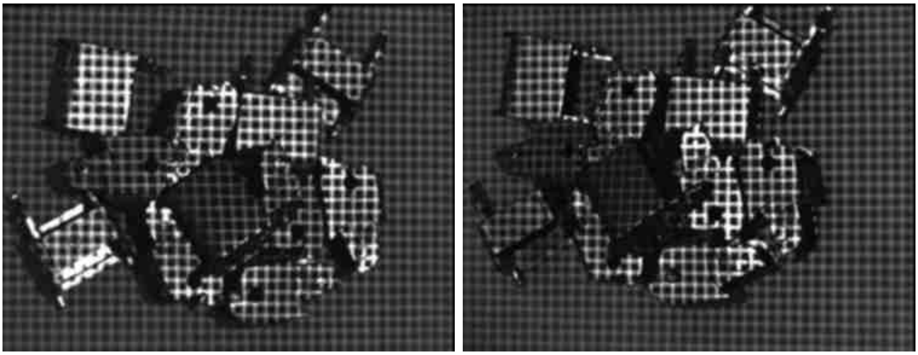


Fig. 1. An image pair showing a heap of unorganized parts with a high contrast grid projected on it.

in 3D, a grippoint is selected on the plane and a normal vector is calculated. A selection procedure is applied to establish the topmost largest plane. Their coordinates and normal vector are passed to the control of the robot arm [5]. Fig. 3 (b) shows a robot picking one of the chain parts from Fig. 1.

5 Pose Determination of Isolated Objects

This section presents the second module of the assembly cell. After picking an object, the robot places it on a separate workplace and a high resolution camera grabs an image. The pose determination module calculates a pose and passes those points to the robot where the object can be grasped for mounting. The pose is determined by fitting a CAD model to image edge features using an iterative refinement procedure. To ensure fast convergence, a rather precise initial guess of the pose must be given. A very powerful approach for obtaining a coarse pose (we will call it pose *estimation*) is the Parametric Eigenspace [15].

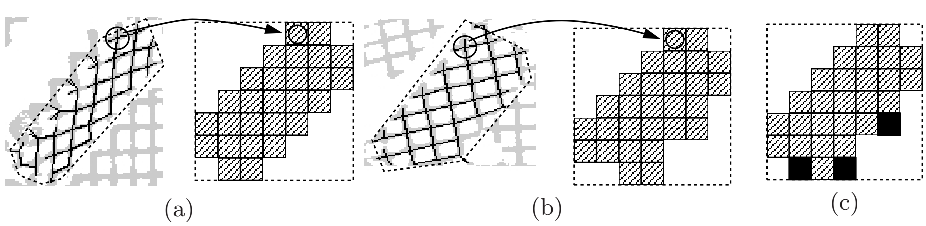


Fig. 2. Each intersection of a plane is represented as a pixel in the perspective invariant plane representation. Segmented plane from **(a)** left view, **(b)** right view. **(c)** Matching results for **(a)**-**(b)**, intersection points with no correspondence are marked black.

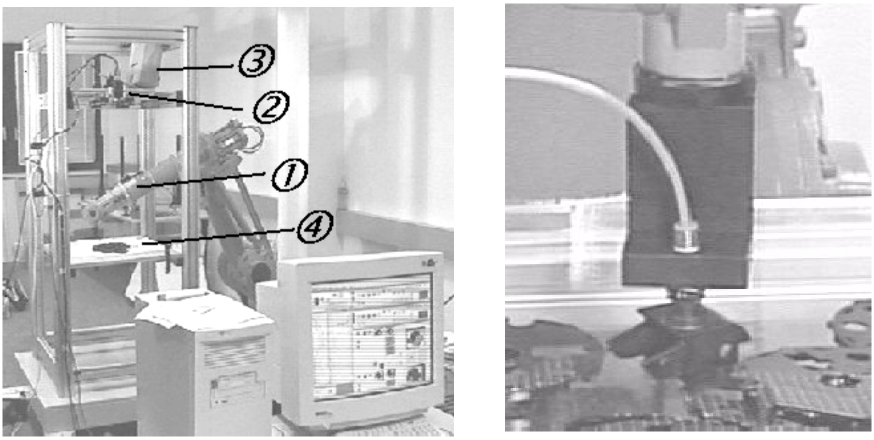
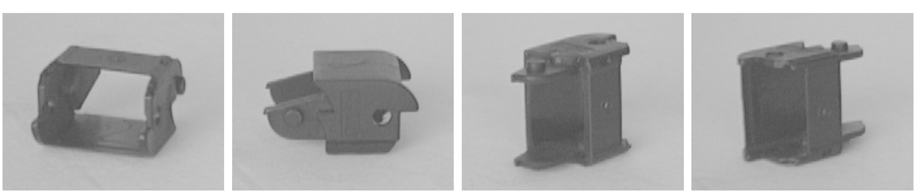


Fig. 3. (Left) The experimental setup consisting of a robot arm (1), a stereo rig (2), a grid projector (3) and a workplace (4). (Right) The robot picking a part from a heap of unorganized objects.

This PCA[2]-based method deals with the appearance of objects depending on the viewpoint parameters. In the present case a one-parametric description was chosen to obtain a representation of each stable pose (see Fig. 4) sampled at a 10-degree resolution. The eigenspace representation is generated from these images in an offline learning phase. See [6] for details.

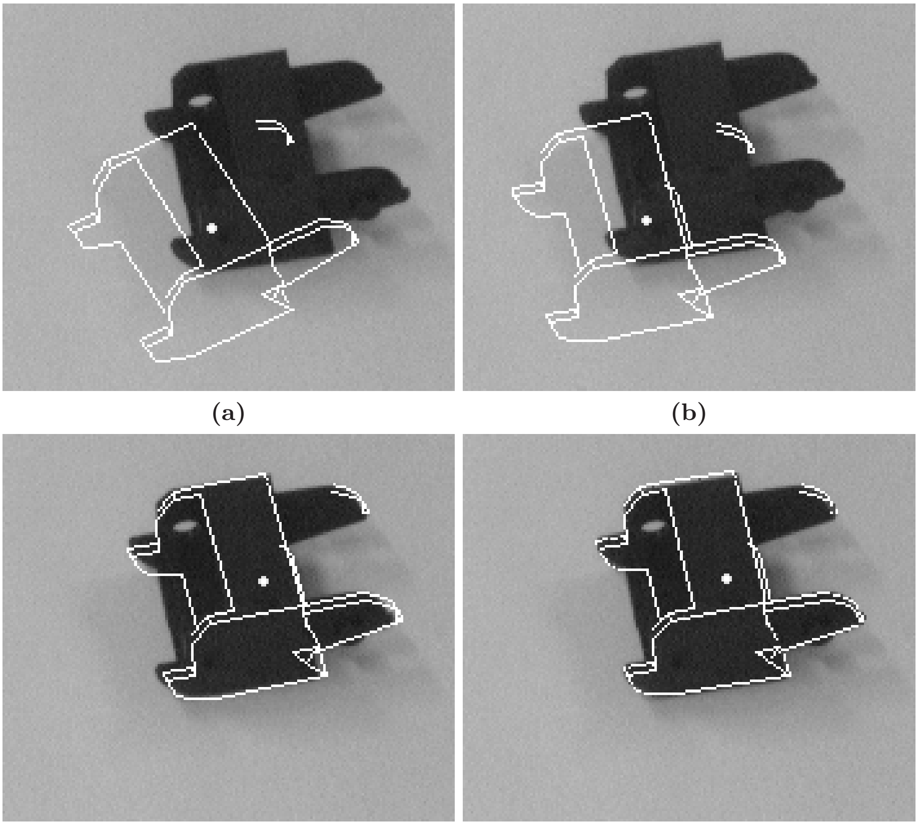


Stable Pose 1 Stable Pose 2 Stable Pose 3 Stable Pose 4

Fig. 4. Four stable poses of an industrial part.

In the online pose estimation, this method provides the stable pose and the approximate rotation angle ϕ of the inspected object around the vertical axis.
Since the parametric eigenspace is an object centered method, there is no a-priori information available on the translation on the workplace. This drawback can be eliminated by virtually placing the CAD model on the workplace (see Fig. 5 (a)), rotating it by the angle ϕ from the eigenspace around its vertical axis (Fig. 5 (b)) and translating it on the image of the object (Fig. 5 (c)). From the translation in the image, which can be calculated by fitting the bounding rectangles of the

[2] Principal Component Analysis



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

Fig. 5. The steps performed by the pose determination module visualized by superimposing the model projection on the image. The white dot indicates the center of gravity in 3D. **(a)** The CAD model placed on the center of the workplace, **(b)** rotation corrected, **(c)** rotation and translation corrected and **(d)** fitted. Note the difference on the left side of the object before **(c)** and after **(d)** the model fitting process.

projected model and the image of the object, the translation **V** on the workplace is estimated. Integrating the rotation angle ϕ, the translation vector **V** and the calibration data results in a pose estimate in 3D. The subsequent fitting step (we call it *pose determination*) starts from this initial guess. An adapted hidden line algorithm allows to calculate the line clippings in 3D world coordinates, which is essential for the 2D-3D feature correspondence and the 3D parameter refinement. Note that we are given only three degrees of freedom instead of six. Because of the "Ground Plane Constraint" (see for example [2]) they are limited to the rotation angle ϕ and the two dimensional shift vector **V**. Some error measures are necessary for the pose parameter assessment. We define them as the perpendicular distances between model edges and image gradients, as suggested by Lowe [14]. From this error vectors in 2D we can iteratively refine the pose parameters by solving the equation

$$\begin{pmatrix} \Delta\phi \\ \Delta V_x \\ \Delta V_y \end{pmatrix} = J \cdot \begin{pmatrix} e_1 \\ e_2 \\ e_3 \\ \vdots \end{pmatrix} \tag{1}$$

where J is the Jacobian of the 3D-2D mapping function and e_i are the observed errors in the image.

6 Practical Results

In this section experimental results are presented. Fig. 3 (a) shows the industrial setup - a prototype of the assembly cell. There are three standard digital cameras, two of them for the bin picking and one for the pose determination. A common slide projector is used to texture the observed scene with the high contrast grid. The whole system is calibrated with the well-known Tsai method [18]. For learning the stable pose views in their various angular positions we used a turn table. In the current implementation the robot picks an object from a heap and places it on a separate workplace. After the execution of the pose determination module, the robot picks the object again and due to a now known orientation, pokes it in a fit.

3D Plane Reconstruction	#	%
Detected	2190	100,00
Correct	2134	97,44
Incorrect	56	2,56
Not detected	190	8,68
Reconstructed	2134	97,44
Picked	2026	92,51
Missed	108	4,93

Table 1. Summary of the Bin Picking module; results from various test series.

The picking module has more than 92% success rate, which is promising for industrial applicability. Table 1 summarizes the results of several test series. The pose *estimation* (i.e. the geometry integrated eigenspace method) delivers pose parameters corrupted by an error of approximately ±3 degrees in rotation and less than 2 mm in both translation directions. This error is decreased significantly by the subsequent fitting, typically under half a degree in rotation and

Pose Determination Error	$\|\Delta\phi\|$	$\|\Delta V_x\|$	$\|\Delta V_y\|$
Average	0.30^o	0.41 mm	0.37 mm
Standard Deviation	0.21^o	0.19 mm	0.15 mm
Maximum	4.23^o	1.43 mm	1.18 mm
Minimum	0.01^o	0.02 mm	0.01 mm

Table 2. Evaluation of the pose determination module.

half a mm in translation. Table 2 lists the errors for each calculated pose parameter.

The presented system runs completely on standard hardware components, which is a direct consequence of the three-step concept. All of the presented algorithms perform in real time. Despite this, there are still some possibilities left for increasing the speed.

7 Conclusion and Future Work

A novel approach to the problem of picking objects from a bin and mounting them was presented. The task is splitted in three independent subtasks. Plane detection is used to identify grip points on the parts of the unorganized heap. The matching and the stereo correspondence retrieval are computed by a novel structured light method, which combines them in one single processing step. After the isolation of a single part, the second module identifies its pose and enables a manipulator to grasp it precisely. The system uses an geometry integrated eigenspace approach to estimate the pose of the object. This initial guess is refined by an image feature based model fitting. Test series proved the applicability of the stated three-step concept and the described algorithms for industrial environments.

Future work will include grid analysis algorithms such as identification of singularities, correspondence refinement and working height estimation. The pose estimation module will be completed by some fitting strategies. All parts of the system implemented until now are under heavy testing in an industrial environment.

References

1. Al-Hujazi, E., and Sood, A. Range Image Segmentation with Applications to Robot Bin-Picking Using a Vacuum Gripper. *IEEE Transactions on Systems, Man and Cybernetics 20*, 6 (November/December 1990), 1313–1324. 109, 110
2. Araújo, H., Carceroni, R. L., and Brown, C. M. A Fully Perspective Formulation to Improve the Accuracy of Lowe's Pose-Estimation Algorithm. *Computer Vision and Image Understanding 70*, 2 (1998), 227–238. 110, 114

3. Ayache, N., and Faugeras, O. D. HYPER: A new approach for the recognition and positioning of two-dimensional objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence 8*, 1 (1986), 44–54. 110
4. Bachler, G., Berger, M., Röhrer, R., Scherer, S., and Pinz, A. A Vision Driven Automatic Assembly Unit: Robust Bin Picking. In *Proceedings of the 23rd Workshop of the AAPR* (Steyr, Austria, May 1999), pp. 205–213. 111
5. Bachler, G., Berger, M., Röhrer, R., Scherer, S., and Pinz, A. A Vision Driven Automatic Assembly Unit. In *Proceedings of the 8th Conference on Computer Analysis of Images and Patterns* (Ljubljana, September 1999), pp. 375–382. 112
6. Berger, M., Bachler, G., Scherer, S., and Pinz, A. A Vision Driven Automatic Assembly Unit: Pose Determination from a Single Image. In *Proceedings of the 23rd Workshop of the AAPR* (Steyr, Austria, May 1999), pp. 215–224. 113
7. Brooks, R. A. Symbolic Reasoning around 3-D Models and 2D-Images. *Artificial Intelligence 17* (1981), 285–348. 110
8. Fischler, M. A., and Bolles, R. C. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM 24*, 6 (June 1981), 381–395. 110
9. Gold, S., Rangarajan, A., Lu, C.-P., Pappu, S., and Mjolsness, E. New Algorithms for 2D and 3D Point Matching: Pose Estimation and Correspondence. *Pattern Recognition 31*, 8 (1998), 1019–1031. 110
10. Haralick, R. M., Joo, H., Lee, C., Zhuang, X., Vaidya, V. G., and Kim, M. B. Pose estimation from corresponding point data. *IEEE Transactions on Systems, Man and Cybernetics 19*, 6 (Nov./Dec. 1989), 1426–1446. 110
11. Haralick, R. M., nan Lee, C., Ottenberg, K., and Nölle, M. Analysis and Solution of The Three Point Perspective Pose Estimation Problem. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition* (1991), pp. 592–598. 110
12. Ikeuchi, K. Generating an Interpretation Tree from a CAD Model for 3D-Object Recognition in Bin-Picking Tasks. *International Journal of Computer Vision* (1987), 145–165. 109, 110
13. Lowe, D. G. Three-Dimensional Object Recognition from Single Two-Dimensional Images. *Artificial Intelligence 31*, 3 (March 1987), 355–395. 110
14. Lowe, D. G. Fitting parameterized three-dimensional models to images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 13*, 5 (May 1991), 441–450. 110, 115
15. Murase, H., and Nayar, S. K. Visual Learning and Recognition of 3-D Objects from Appearance. *International Journal of Computer Vision 14* (1995), 5–24. 112
16. Rahardja, K., and Kosaka, A. Vision-Based Bin-Picking: Recognition and Localization of Multiple Complex Objects Using Simple Visual Cues. In *Proceedings of the International Conference on Intelligent Robotics and Systems* (Osaka, Japan, November 1996), IEEE / RSJ. 109, 110
17. Trobina, M., Leonardis, A., and Ade, F. Grasping Arbitrarily Shaped Objects. In *Mustererkennung* (Wien, 1994), W. G. Kropatsch and H. Bischof, Eds., vol. 1, Technische Universität Wien, Xpress, pp. 126–134. 109, 110
18. Tsai, R. Y. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology Using Off-the-Shelf TV Cameras and Lenses. *IEEE Journal of Robotics and Automation RA-3*, 4 (August 1987), 323–344. 115
19. Weinshall, D., and Basri, R. Distance metric between 3d models and 2d images for recognition and classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence 18*, 4 (April 1996), 465–470. 110