2013 IEEE International Conference on Robotics and Automation (ICRA)
Karlsruhe, Germany, May 6-10, 2013

Efficient Bin-Picking and Grasp Planning Based on Depth Data

Dirk Buchholz[1], Marcus Futterlieb[1], Simon Winkelbach[1] and Friedrich M. Wahl[1]

Abstract— **The problem of object localization is a well-known problem in industrial robotics. Manufactured parts arrive at factories as bulk goods in boxes. Single parts need to be picked out of the boxes and have to be fed to a machine. The task of automatically isolating single objects is known as the bin-picking problem. Even in modern factories the task of bin-picking is not automated widely yet. The automatization of this task is expensive since state-of-the-art solutions require object-class specific algorithms. In this paper we present an applicable solution for the bin-picking problem which is based on a standard 3d-sensor and is able to handle arbitrary objects. Furthermore, it is robust against noise and object occlusions. Additionally, we propose an approach for optimal grasp pose estimation with collision avoidance that effectively reduces system cycle times.**

I. INTRODUCTION AND RELATED WORK

In the context of industrial robotics the task of object localization is a very important one. With the self-imposed goal of creating human friendly and ergonomic factories, manufacturers see themselves faced with several tasks where intelligent autonomous robots are needed. One important example of these tasks is robotic bin-picking. By creating an autonomic, generic and robust system, it would be possible to spare human workers from monotonic, unergonomic and not favorable work. For this reason much research has been done on this field. Many publications deal with this problem and offer a wide range of approaches towards its solution. The proposed approaches use different sensor types and different object detection algorithms but nearly all of them have limitations in some aspects. Furthermore, the important subtask of collision avoidance and grasp pose planning is not or only very briefly mentioned in most of the publications.

Berger et al. show an approach based on plane detection [1]. In a second step they determine the object pose of the isolated object. Ghita and Whelan show a system that locates box-like polyhedral objects [2]. Safranov et al. show an approach that only locates cylindric objects [3]. This is similar to the work of Oh et al. who rely on edges of cylinders [4]. All these works constitute interesting approaches but have in common that they are limited to special geometries of the objects. An object independent solution to the bin-picking problem is shown by Kirkegaard and Moeslund [5]. But this solution is sensitive to noise which makes it difficult to set it up in an industrial environment. Similar problems appear in the work of Hema et al. who need optimal lighting conditions [6]. A very promising approach based on mesh simplifications and the iterative closest point algorithm was proposed by Boehnke et al. [7]. Unfortunately, they do not give any practical results of their system. In [8] the ***Ran****dom* ***Sa****mple* ***M****atching* (RANSAM) algorithm is used to locate single objects in a bin. The RANSAM algorithm was originally developed to solve the 3d puzzle problem [9]. Not focusing on specific object features this approach can deal with arbitrary object shapes. Papazov et al. also use this algorithm to solve the object localization problem but do not aim at industrial applications and instead of a collision avoidance use an impedance controlled robot [10].

Regarding the problem of collision avoidance and grasp planning there are only a few publications that deal with this problem. One example is the work of Schyja et al. [11] in which the use of existing point cloud based algorithms like RAPID [12] is proposed.

In this paper we describe a bin-picking system with its object localization algorithm and collision avoidance mechanism. The localization algorithm is an enhancement of existing surface based localization techniques which overcomes issues which, besides others, occur when dealing with objects mainly consisting of planar faces and using scanners with fixed viewpoints. The proposed collision avoidance mechanism is designed to be easy to use and efficient to calculate and is capable to estimate optimal grasp poses in predefined regions. The described techniques are applicable to arbitrary hardware making the system easy to implement in industrial environments.

II. BIN-PICKING SYSTEM AND OBJECT LOCALIZATION

The proposed bin-picking system is based on 3d scans of the bin and CAD models of the objects scrambled in it.

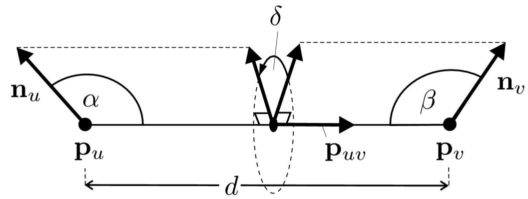
A. RANSAM for Industrial Bin-Picking

The basis for the RANSAM algorithm is a set of two vertices with their appropriate normals called an oriented point pair or ”dipole”. It serves as a generic surface feature that can be computed in every 3d mesh. Therefore, no specific surface features like planes or circles have to be present in the objects that need to be located. The dipole has the property to build 4 rotation and translation invariant features as can be seen in Fig. 1. Using these 4 invariant features as coordinate axes, two 4d relation tables are built. One table for the scan data and one table for the CAD model [8].

To locate an object, dipoles are generated alternatingly by picking two arbitrary vertices of the scan and the model. To estimate surface normals in the scan, a triangulation of the points is performed. After a dipole is built and stored in

[1]All Authors are with Institut für Robotik und Prozessinformatik, Technische Universität Braunschweig, 38106 Braunschweig, Germany {d.buchholz, s.winkelbach, f.wahl}@tu-bs.de, m.futterlieb@gmail.com

978-1-4673-5643-5/13/$31.00 ©2013 IEEE



d	distance between point $\mathbf{p}_u$ and point $\mathbf{p}_v$
α	angle between the normal $\mathbf{n}_u$ and vector $\mathbf{p}_{uv}$ ($\mathbf{p}_{uv}$ is the unit vector pointing from $\mathbf{p}_u$ to $\mathbf{p}_v$)
β	angle between the normal $\mathbf{n}_v$ and vector $\mathbf{p}_{uv}$
δ	signed angle between $\mathbf{n}_u$ and $\mathbf{n}_v$ around $\mathbf{p}_{uv}$ (i.e. dihedral angle between the plane with normal $\mathbf{n}_u \times \mathbf{p}_{uv}$ and the plane with normal $\mathbf{n}_v \times \mathbf{p}_{uv}$)

Fig. 1. Rotation and translation invariant features of a dipole.

its relation table the other table is checked for a collision, meaning that a similar dipole was already found in the other data set.

If such a collision occurs, a pose hypothesis is found by computing a unique frame that transforms the model dipole onto the scan dipole. Due to the fact that the relation tables are filled continuously the probability of a collision rises over time making the algorithm faster for each new entry in the tables. The CAD model stays constant all the time leading to a well filled table and fast localization times during runtime.

To evaluate the hypotheses, the number of points in contact is estimated. Owing to the fact that similar dipoles can be found at various locations on the objects many false hypotheses are found. To deal with this, a very fast evaluation of each hypothesis is needed. This is done by applying a *kd-tree* [13] for nearest neighbor search and using a quality forecast by an efficient Monte-Carlo strategy.

After a valid hypothesis has been found, the well known *ICP* algorithm [14] is applied to refine the pose estimate.

To complete the task of autonomously grasping parts and moving them to a defined deposit, some kind of collision avoidance has to be ensured. This problem will be dealt with in Section IV.

III. Localization and Evaluation Improvement via Redundancy Elimination

For the proposed approach it is necessary that the scan data that is used for localization exclusively consists of the bin content. The reason for this is the basis of the localization algorithm. The RANSAM algorithm is based on surface data. When applying this algorithm not to two complete models, but to one CAD model and one *scanned* surface this has to be considered. Scanning reflecting metal parts with an optical scanner mostly leads to low point density on the reflecting surfaces. In contrast to the objects the used bin in the experiments was made of plastic which had much better reflectance properties in sense of scanability leading to much higher point densities. In this context this means that false matchings (an object is matched into the bin) are likely to get high qualities because the quality estimate depends on the amount of touching points.

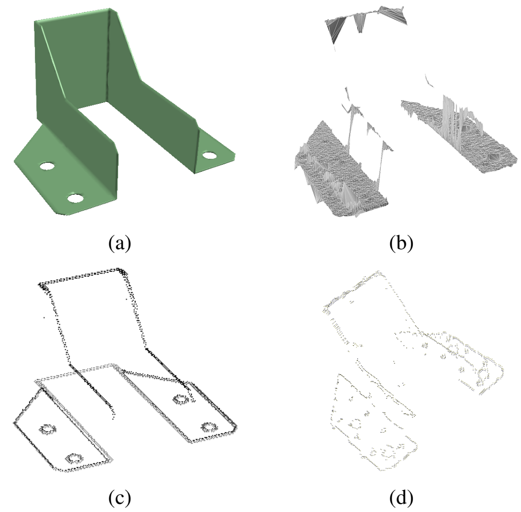


Fig. 2. Joist hanger. (a) CAD model, (b) scan from above, (c) CAD edge model, (d) scan edges.

To overcome this issue one can simply delete the bin points of the scan for the matching procedure and re-include them for collision avoidance algorithms, i.e., using a calibration of the bin pose. This is done because it also reduces the amount of points leading to faster localization times.

Considering the goal of a generic bin-picking solution this leaves an issue that will be topic of the next section because a similar problem can occur for a specific class of objects.

A. Problem Description

The problem with the bin mentioned above is a conceptual one occurring under different circumstances. The reason is the basis for the localization algorithm being oriented surface patches. If the objects that are scattered in the box mainly consist of planar faces like the joist hanger shown in Fig. 2, the issues are explainable by looking at state of the art scanning systems that can be found in modern industrial applications. Most of these scanning systems only determine the distance values of 3d points in respect to some optical center. This means the scan data is generated using only one point of view. When scanning an object like a joist hanger from only one side, e.g., from above, the scan data is extremely incomplete and the point density is not distributed uniformly. This means that faces with a normal pointing towards the optical center will be well scanned whereas points with normals at an angle near $\pi/2$ relative to the viewing rays will not be visible. This can be seen in Fig. 2(b) where only two faces are completely visible.

Furthermore, the object itself contains plenty of ambiguities meaning that on the planar surfaces there will be lots of similar dipoles that actually lead to false matchings. Moreover, the identification of false matchings will not work properly because of the incomplete scan data, which results in false positive matchings that lead to false robot movements. Choi et al. [15] address this problem by using an extended set of dipoles for the search algorithm. For this extended set further computations have to be performed to extract the needed parts for the dipole, namely line segments or boundary segments.

B. Using 2d Image Analysis to Enhance the Localization Performance

To overcome the problems of the RANSAM approach when dealing with not uniformly distributed surface data, a modification of the previously presented localization algorithm is needed. This enhancement deals with the special issues that occur when a scene is scanned from one viewing direction only (the scanner is fixed and only has one viewing direction). Regarding this, common 3d scanners not only produce point clouds, but this point cloud can also be understood as a depth image. Also, not only the depth values of the scanners can be used but also the intensity of the scanned points in the image. For a laser line scanner like the one used in the experiments these intensities are the reflectance values of the laser line giving hints to the surface orientations. Both these images can be used to extract edges. This can be done by simply applying the well-known *Sobel* operator. Thus, planar surfaces are discarded and only edges remain resulting in a much more uniform 3d point density. Obviously, the same procedure has then to be applied to the CAD model as well, either by using a 3d edge extraction algorithm or by projecting the model onto the image plane and using the same functionalities as mentioned above (see Fig. 2 (c) and (d)).

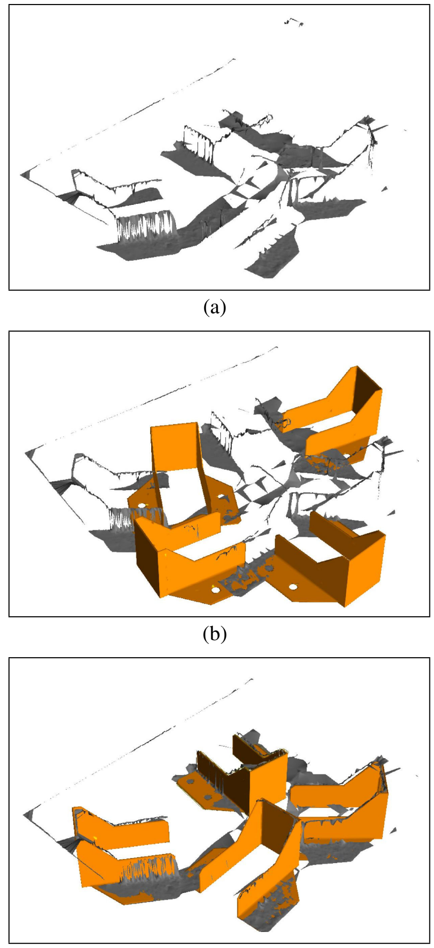
C. Modification of the RANSAM algorithm

The RANSAM algorithm uses oriented point pairs for localization. At this point of the work all planar faces have been deleted. Estimating surface normals at edge points is very unstable. To overcome this the RANSAM search method is adapted to the new edge data sets. When no normals are available, the used dipole would only contain one invariant which is not enough. By adding a third point and building a random triangle (tripole) using three random vertices of the point cloud, three invariant features can be computed using the three distances of the vertices . Using this, two 3d relation tables can be built up and the algorithm can now deal with point sets without normals. This method was also proposed in [16].

The described pose evaluation remains the same, but only uses the edge data which reduces the computation time and enhances the robustness in the described scenario.

D. Advantages and Applications of Edge Variation

With the new version of the algorithm several parts of the system can be adapted to deal with unsuitable data sets (cf. Fig. 3). On the one hand, like mentioned above, it is possible to only use edges for localization. But if the objects that need to be localized are not causing the problems described above (like the piston rods shown in Section V), the original algorithm leads to more robust results because the dipole contains more significant information as the tripole. Therefore, on the other hand, it is possible to use the full data at first and enhance the accuracy of the localization by applying the edge version afterwards. This has the advantage that the estimated pose of the first step can be used to generate the edge data of the CAD model. The edge model that results using the known viewing direction only consists of the edges that are visible to the scanner. Furthermore, by checking if edges in the scan data are adjacent to the edges of a located object, false positive matchings can be dismissed. This step can be done in 2d as well as 3d.



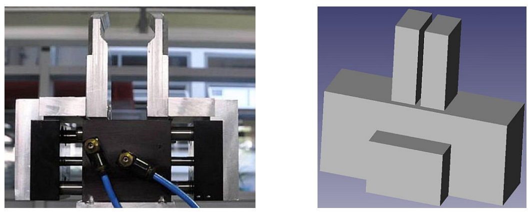
(a) (b) (c)

Fig. 3. Application of the RANSAM modification. Located Objects are displayed in orange, scan data in gray. (a) Scan of a bin filled with four joist hangers. (b) Erroneous localization result using the RANSAM with dipoles. (c) Correct localization result using the tripole variation.

IV. Grasp Planning

As already mentioned in the introduction, in industry raw parts are often sensitive to damages and are therefore only allowed to be grasped at a very limited amount of areas. Additionally, the grippers available in the portfolio of the manufacturers are mostly parallel jaw grippers. Nevertheless, an intelligent system for grasp planning is needed. If only a small set of predefined grasp poses (defined relative to the object coordinate system) is used, parts scattered in boxes will in many cases not be accessible for the gripper even though they could be grasped easily, hence leading to long cycle times when a large amount of objects needs to be localized until one is found that is accessible. In this section we introduce a ”semi automatic” grasp pose estimation that overcomes the mentioned issues.

All collision handling mechanisms discussed in this section use a CAD model of the end-effector as basis. In the experiments a standard parallel jaw gripper was used which was modeled using four cuboids consisting of 32 vertices overall (see Fig. 4).



(a) (b)

Fig. 4. The gripper used in the experiments. (a) Gripper. (b) CAD model.

A. Key Grasp Frame and Collision Volume

To achieve both goals – high variety of grasp poses at limited object regions and easy definition of these poses – the ***Key Grasp Frame*** concept (KGF) is introduced. A KGF consists of a starting pose for the end-effector defined in the CAD model coordinate system, a set of degrees of freedom (DOF) and an associated range for each DOF in which this starting pose can be varied. This concept will still allow a wide variety of grasping positions (due to a quasi continuous variation of poses), while maintaining the demands of the industry to be in control of which region of the object classifies as a picking position.

Besides automatic grasp pose variation, a concept for pose evaluation is needed. State of the art collision avoidance mechanisms often use point cloud based implementations like RAPID [12]. Looking at sensors widely used in industry this concept does not optimally use the available data. Fixed viewpoint scanners can be seen as generating depth images only and no real 3d data. Interpreting the scan data in this way, collision analysis methods can be computed completely in 2d as depth images can be treated as simple gray valued images.

To use the depth images as basis for collision calculations, a depth image of the gripper model is needed, too. For this the gripper is separated into convex subparts which is a simple task for common grippers. Then two images for each part are generated using the simulated grip pose by combining the located object frame in sensor coordinates and the KGF basis frame. One image contains the upper part and one the lower part of the gripper.

To evaluate a given pose, simple difference images are computed using both gripper images and the depth image of the scan. Whenever a pixel of the upper face of the end-effector is above the scan, but the same pixel of the lower face is located below, a collision in that pixel of the image is located. With the value of the differences of the pixels

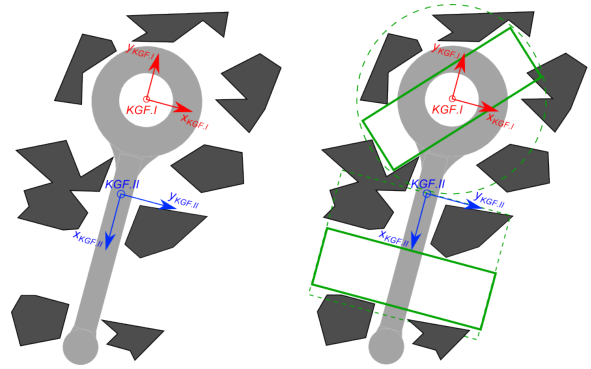


Fig. 5. Visualization of the KGF Concept

and the size of the pixels an exact collision volume can be calculated for that pixel.

Whenever both surfaces are located below the scan, no clear evaluation can be done for that pixel because the gripper could be in occluded free space or in occluded collisions. This state is called threat volume.

If both faces of the gripper are above the scan, no collision volume exists.

To evaluate the whole pose, all single pixel collision volumes for all single convex subparts of the gripper are summed up. The threat volume can be added to the collision volume with a predefined scaling factor. This scaling factor depends on the situation and the degree of safety that is needed for the specific task.

The overall collision (and threat) volumes build a penalty value for each pose. To avoid collisions this penalty must be under a defined threshold. This threshold cannot be zero because due to the real life situation, the range data is subject to a certain level of noise. In addition, the fingers of the end-effector are chamfered, which makes minor collisions harmless to the hardware.

B. Optimal Grasp Pose Estimation

With the previously defined KGFs it is possible to evaluate each pose of the end-effector using the defined set of DOFs and their ranges. To minimize computational costs a new range image of both, the scan and the gripper is generated using the gripper coordinate system transformed to the base frame of each KGF. The z-axis of the gripper being the same as the approach vector serves as depth axis. With this new coordinate system the scan as well as the gripper is rendered using orthographic projection. These rendered images can be thought of taken ”looking through the gripper” using a virtual orthographic camera.

The area of the rendered scene is directly given by the type of the DOF (rotational or translational) and the range of the free parameter. The concept is shown in Fig. 5. Here the area for rendering is shown in green scattered lines, the base frame for two exemplary KGF as coordinate system is shown in red and blue.

Now, for the translational DOFs a simple correlation-like procedure of both images is performed using only the free

DOF. The gripper image is shifted pixel-wise along its free axis over the scan image. The collision volume is stored for each step resulting in a collision function dependent of the variable parameter. Using this function, all parameter values below the collision threshold are located and used to build a distance map for all possible collisions. The distance maps of all subparts of the gripper are then combined resulting in an overall optimal value for the variable parameter. The described procedure can be efficiently implemented using integral images introduced in [17].

In the case of a rotational DOF a preprocessing step has to be performed. To convert the rotational problem into a translational one, the rendered images are transformed into polar coordinates. The following steps are then the same as before.

After the parameters for all defined KGF are computed, the one with the lowest penalty volume is chosen.

The collision volume that comes from this calculation is not as accurate as possible. Due to the (possibly) tilted point of view of the gripper relative to the scan direction threat and collision volumes may have changed. Therefore, a second collision volume estimation as described above is performed using the original depth data for only the optimal frame.

To visualize the concept, an example of the algorithm estimating an optimal pose using a rotational DOF KGF is shown in Fig. 6.

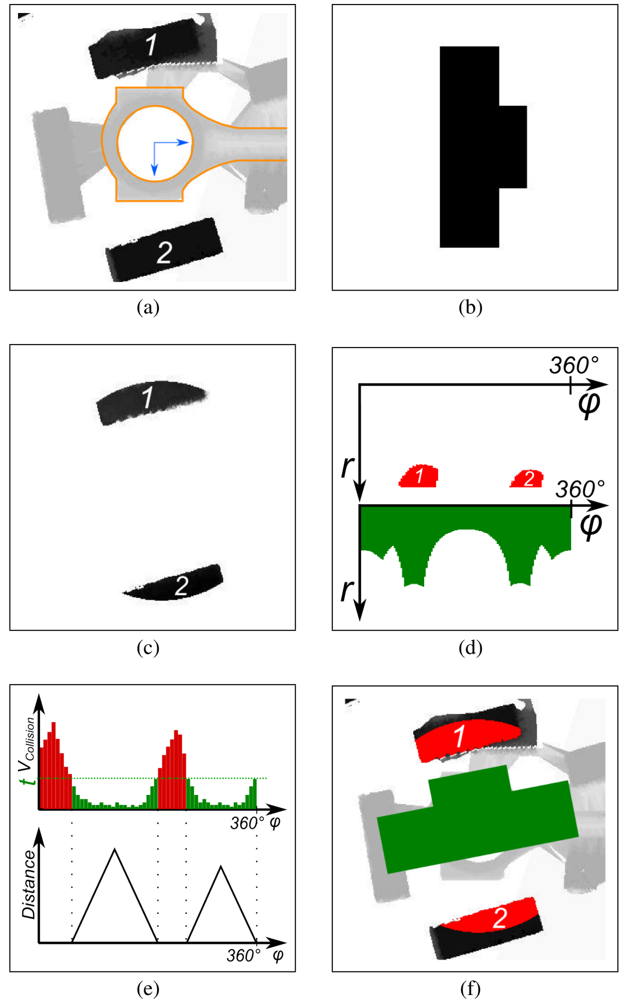


Fig. 6. Example for the optimal grasp pose estimation algorithm only using the palm of the gripper. The KGF is centered in the hole of the object. The part will be grasped by opening the gripper. (a) Depth image of the scan rendered in gripper coordinates, centered and oriented using the base frame of the KGF (blue) at a located object (orange) surrounded by obstacles that would lead to a collision (1 & 2). Dark values are near, bright values are far. (b) Depth image of the gripper palm model. (c) Section of the depth image, reduced by all pixels that do not collide with the gripper due to their height or distance to the KGF base frame. (d) Polar coordinate representation of the gripper palm (green, below) and the interesting parts of the scan (red, above) respectively. This data is used to solve the best pose problem. (e) Collision function dependent on rotational DOF (above) and distance function to all occuring collisions (below) with defined collision threshold t. (f) Superimposed result of the optimal pose.

V. EXPERIMENTAL RESULTS

A. Prototype Hardware

To evaluate the usage of the proposed approach, a prototype system was built up including a Stäubli RX-60 industrial manipulator equipped with a parallel jaw gripper and a SICK IVP Ruler E1200 laser line scanner mounted on a linear axis (cf. Fig. 7). All computations were executed on a PC with a 3.6 GHz CPU and 8 GB of RAM.

B. Experiments

To give an experimental evaluation of the proposed system a series of 156 pick attempts was executed. Using our small box, this means that it was filled 12 times using 13 objects which were randomly dropped. Besides the already shown joist hanger (as example for a planar object) a piston rod (see Fig. 8) (as example for a free form surface object) was used for the experiments. Unfortunately, there was no ground truth available describing the object poses. Nevertheless, to give a measure of the accuracy of the approach we measured the distance of the real end-effector after it arrived at the grasp poses in the scan.

This measure is not very accurate but sufficient to show the applicability of the localization approach. Furthermore, this error includes calibration errors of the scanner and the robot as well as errors in the kinematic parameters of the robot. The average error of this measure was 1.1 mm with a standard deviatation of 0.2 mm leading to a pick success rate of 100% which means that all grasp attempts of the robot were successfull and no collisions occured during the test series.

To generate valid pick poses we used 3 different KGFs with both, rotational and translational DOFs. The KGFs were defined manually. After an object has been detected, one KGF after another is evaluated. As soon as a possible grasp frame is computed this frame is executed. The case that an object that lies under other objects is located occurs and is purposely not avoided. If this object can be picked, the resulting movements in the bin often changes unpickable objects poses and enables the robot to grasp these parts.

The average time in which an object was localized and an optimal pick pose generated measured from the moment on the scan data was available was 5.5 seconds with a standard

(a)

Fig. 7. Prototype bin-picking system.

(a) (b)

Fig. 8. Piston rod. (a) CAD model, (b) scan from above.

deviation of 2.6 seconds. This measure includes situations in which the first located object was not pickable and a second one was located. Regarding only the attempts in which the first localization led to valid grasp poses (which was the case in 85.9% of the test cases) the average time reduces to 4.5 seconds with a standard deviation of 0.3 seconds. The time to scan the bin as well as the robot movement was not measured because these times are very dependent on the used scanner, robot, workspace, etc.

The implementation of the system is not parallel so that all KGFs are processed in a row each taking approx. 0.1-0.3 seconds dependent on the dof type. The computation time can therefore be reduced by parallelizing the single KGFs. Furthermore, the search algorithm can be parallelized, as demonstrated in [16] which is not done in the experiments.

VI. Conclusion and Outlook

This document presents an approach for a generic bin-picking system. It can be built up using commercially available hardware only and is applicable to many kinds of objects. A prototype using the described approach shows its great potential. By introducing an easy to implement variation of the localization algorithm the system even works for objects that are hard to locate using a surface based localization algorithm and a sensor that has a fixed viewpoint (see Fig. 3).

To optimally use the available data, a collision avoidance mechanism was introduced that can also be used to choose optimal grasp poses for the picking procedure. This algorithm can effectively reduce cycle times of the system because the automatically computed parameters for predefined degrees of freedom mostly lead to valid grasp poses for the first localized object which often could not have been grasped using predefined static grasp poses.

To enhance the applicability of the proposed system for industrial applications, the computation times should be optimized. This can be done by parallelizing the individual algorithms and using multi core systems.

Acknowledgment

We would like to thank the German Academic Exchange Service (DAAD, Vigoni programme) for supporting and partially funding this project.

References

[1] M. Berger, G. Bachler, and S. Scherer, "Vision guided bin picking and mounting in a flexible assembly cell," in *Intelligent Problem Solving. Methodologies and Approaches*, ser. Lecture Notes in Computer Science, R. Logananthara, G. Palm, and M. Ali, Eds., vol. 1821. Springer Berlin / Heidelberg, 2000, pp. 255–321.

[2] O. Ghita and P. F. Whelan, "A bin picking system based on depth from defocus," *Machine Vision and Applications*, pp. 234–244, 2003.

[3] K. Safronov, I. Tchouchenkov, and H. Wörn, "Hierarchical iterative pattern recognition method for solving bin picking problem," in *Proc. of Robotik 2008*, 2008.

[4] J.-K. Oh, K. Baek, D. Kim, and S. Lee, "Development of structured light based bin picking system using primitive models," in *Assembly and Manufacturing, 2009. ISAM 2009. IEEE International Symposium on*, nov. 2009, pp. 46–52.

[5] J. Kirkegaard and T. Moeslund, "Bin-picking based on harmonic shape contexts and graph-based matching," in *Proc. Of International Conference on Pattern Recognition*, 2006, pp. 18–21.

[6] C. Hema, M. Paulraj, R. Nagarajan, and S. Yaacob, "Object localization using stereo sensors for adept scara robot," in *Robotics, Automation and Mechatronics, 2006 IEEE Conference on*, June 2006, pp. 1–5.

[7] K. Boehnke and M. Otesteanu, "Progressive mesh object registration," in *System Integration, 2008 IEEE/SICE International Symposium on*, Dec. 2008, pp. 6–11.

[8] D. Buchholz, S. Winkelbach, and F. M. Wahl, "Ransam for industrial bin-picking," in *Proc. of 41st Int. Symposium on Robotics (ISR)*, 2010, pp. 1–6.

[9] S. Winkelbach, S. Molkenstruck, and F. M. Wahl, "Low-cost laser range scanner and fast surface registration approach," in *DAGM*, 2006, pp. 718–728.

[10] C. Papazov, S. Haddadin, S. Parusel, K. Krieger, and D. Burschka, "Rigid 3d geometry matching for grasping of known objects in cluttered scenes," *The International Journal of Robotics Research*, vol. 31, no. 4, pp. 538–553, 2012.

[11] A. Schyja, A. Hypki, and B. Kuhlenkotter, "A modular and extensible framework for real and virtual bin-picking environments," in *Proc. of International Conference on Robotics and Automation*, May 2012, pp. 5246–5251.

[12] S. Gottschalk, M. C. Lin, and D. Manocha, "Robust and accurate polygon interference detection," http://gamma.cs.unc.edu/OBB/, Version 2.00 [released June 10, 1997].

[13] J. H. Friedman, J. L. Bentley, and R. A. Finkel, "An algorithm for finding best matches in logarithmic expected time," in *ACM Transactions on Mathematical Software*, vol. 3, 1977, pp. 209–226.

[14] P. J. Besl and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[15] C. Choi, Y. Taguchi, O. Tuzel, M.-Y. Liu, and S. Ramalingam, "Voting-based pose estimation for robotic assembly using a 3D sensor," *2012 IEEE International Conference on Robotics and Automation*, pp. 1724–1731, May 2012.

[16] R. Iser, D. Kubus, and F. M. Wahl, "An efficient parallel approach to Random Sample Matching (pRANSAM)," in *IEEE International Conference on Robotics and Automation*, 2009, pp. 1199–1206.

[17] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2001, pp. 511–518.