Two-Stage Robotic Bin Picking of Small Metallic Objects

Meike Herbert⬥, Paul Bach, Markus Lieret⬥, Jens Fürst and
Jörg Franke⬥

Abstract

Robotic grasping of small metallic objects such as bolts is a challenging task due to the small dimensions and textureless reflective surfaces. Depth images acquired of such objects are often noisy and error-prone. In addition, overlapping of parts occur as they are provided randomly oriented in a box such as a small load carrier. To overcome the limitations of existing solutions for bolt separation, a flexible and cost-effective system is developed using an industrial robot and a magnetic gripper. In a two-stage procedure, the bolts are first grasped blindly from a box and placed on a flat surface. In the second step, object detection and pose estimation is performed and the individual bolts are grasped and inserted into a fixture, so that finally the bolts are in a defined position. Industrial use cases for this system are the automated preparation of bolts for robotic screwing processes or automated commissioning of small objects for assembly tasks. The methodology, implementation and evaluation of the proposed solution is presented in this paper.

Keywords

Bin picking · Magnetic gripper · Object detection · Convolutional neural network

M. Herbert (✉) · P. Bach · M. Lieret · J. Franke
Institute for Factory Automation and Production Systems (FAPS), Friedrich-Alexander-
Universität Erlangen-Nürnberg, Erlangen, Germany
e-mail: meike.herbert@fau.de

J. Fürst
Siemens Healthcare GmbH, Erlangen, Germany

1 Introduction

The automation of assembly tasks provides benefits such as increasing productivity and quality, relieving employees of monotonous tasks and reducing costs. Due to their flexibility, industrial robots are applied for automation in assembly in different use cases like the automation of handling and screwing processes. The flexibility of the automation system is particularly crucial with low quantities, a high number of variants or short product life cycles.

To enable robotic assembly automation, the individual components usually have to be placed in defined poses, which is also required for bolts used for robotic screwing. Conventional systems such as step feeders or vibratory bowl feeders do not provide the necessary flexibility to handle different types of bolts and require additional space and investment. If, in contrast, the industrial robot designated for the automation task is used for handling of parts, system utilization can be increased and further investment costs can be saved.

One use case considered is the robotic screwing of components, where the industrial robot can be used to grasp and separate the bolts. If the assembly process takes place in a partially automated production line with 1- or 2-shift operation, the preparation can be done by the automated system in the overnight shift. Another use case is the automated commissioning of parts to provide them in a defined number and position e.g. in shadow boards in order to reduce search times in manual assembly.

The aim of this research is to develop a cost-effective solution for the use cases mentioned above. The use of an industrial robot with a suitable end effector enables grasping of different objects and therefore offers high degree of variant flexibility. The metallic bolts used in these applications are characterized by small dimensions and a textureless, reflective surface. Those characteristics hamper the realization of a cost-effective and flexible solution and face existing bin-picking solutions with challenges.

Therefore, a novel two-stage method for robotic bin picking of small magnetic objects is presented in this paper. The proposed system is characterized by its flexibility and the use of edge computing devices for object detection, pose estimation and motion planning. Hereby, the system can be easily integrated into existing applications without the need of major modifications on the overall robotic cell.

In the following, the corresponding state of the art is described in detail and the need of action is identified. The methodology, implementation and evaluation are presented subsequently.

2 State of the Art

In addition to the selection of a suitable gripper, a key requirement in the implementation of bin-picking solutions is the precise and robust estimation of a suitable grasping pose. There are already numerous solutions for determining the object pose or suitable grip-

ping positions, which are summarized, for example, in [1]. However, many of those solutions require high computation power and are often not suitable for small, textureless and symmetric objects as present in the specified use case. Therefore, the detailed state of the art regarding object recognition on edge devices as well as pose estimation and grasping of small, metallic objects is presented in the following.

For object recognition in colour images, the state of the art offers a variety of solutions that can also be executed on low-power hardware. Widely used solutions include YOLO resp. the version optimized for mobile devices, Tiny-YOLO [2] or Pelee [3].

The segmentation of individual objects can also be performed robustly on low-power hardware. Solutions such as FuseNet, SegNet, or YolactEdge can run on edge devices like the NVIDIA Jetson TX2 or NVIDIA Jetson AGX Xavier, enabling semantic segmentation and, in the case of YolactEdge, instance segmentation at 30 FPS and above [4, 5].

Due to the dimensions and metallic surface of the bolts as well as their unordered positions e.g. in a load carrier, the automated bin picking of bolts is highly challenging. The textureless, metallic surface of the bolts causes reflections and does not provide many distinct features which leads to significant noise in the data captured with common RGB-D cameras and inadequate point clouds of the objects. Thus, some approaches try to grasp and separate bolts or similar objects without the use of computer vision.

Mathiesen et al. present a solution whereby a robot equipped with a scoop-shaped tool grabs the required parts from a box. Within the tool, an orienting groove is used to ensure that only objects with the desired orientation are kept in the scoop. Afterwards the oriented objects can be grasped from the scoop using a separate tool or gripper [6].

Ishige et al. also avoid the application of computer vision and use a gripper with two individually movable fingers and integrated tactile sensors for object grasping and separation instead. First, multiple objects are grasped from a box at once and the number of bolts between the fingers is counted using the tactile sensors. Then the gripper fingers are moved so that excess bolts fall out and finally only one bolt remains in the gripper [7].

Complementary, von Dirgalski et al. propose the combined use of computer vision and force sensors to determine the pose of an object between the gripper fingers [8].

This contrasts with methods using colour and depth data to determine the pose of individual bolts. Furukawa et al. use RGB-D data combined with a template matching approach to detect M6 bolts and subsequently grasp them using a two-finger gripper [9]. The solution presented by Nakano is based on machine learning instead and uses a single shot 6DoF pose estimator to determine the pose of a bolt before grasping it [10].

To circumvent the effects of erroneous depth information for reflective objects, Sato et al. propose a two-step process. In this process, multiple objects are grasped from a load carrier using a magnetic gripper and are placed on a flat surface. Subsequently the objects are classified using RGB information and are individually grasped with the magnetic gripper. Thereby the objects remain in an unknown pose and can thus only be sorted but not fitted into a fixture or mounted to other components [11].

Another way of coping with noisy depth data is the 6DoF pose estimation pipeline for textureless, metallic objects presented by Blank et al. However, this solution is only suitable to a limited extent for small, bulk components, since the objects are too close to each other and overlap and thus cannot be clearly segmented [12].

While all of the presented solutions enable the separation of small, metallic objects, they still come with drawbacks. The approaches either require the design of an object-specific tool, provide the separated objects with an unknown pose, or are only suitable for objects above a certain size. At the same time, established solutions for separation and orientation such as vibratory bowl feeders or step feeders do not offer the necessary flexibility and are characterized by high space requirements and investment costs. Thus, a novel approach for the separation of small, metallic bolts is presented in the following.

3   Two-Stage Bin Picking: System Design and Methodology

The method enables to provide the bolts in a defined pose after the two-stage grasping process. At the same time, the approach also copes with noisy depth information and is characterized by its low investment costs and flexibility.

3.1   Requirements and System Design

The overall aim is the development of a flexible and cost-effective system that can be easily adapted to different objects or variants. The bolts to be separated have small dimensions with a total length of about 15 mm to 35 mm and diameters of 3 mm to 5 mm at the cylindrical shaft and 10 mm to 15 mm at the head of the bolt. The metallic bolts are magnetic and have textureless, reflective surfaces.

In addition, space requirements and the integration of suitable sensors have to be considered. Small and lightweight sensors for object recognition should be arranged at the robotic end effector, while fixed infrastructural sensor systems should be avoided. A cost-efficient solution is preferably selected.

The proposed setup consists of an industrial robot with an appropriate end effector containing a magnetic gripper and a vision sensor. Thus, there is the restriction that only magnetic objects, such as metallic bolts, can be picked. A box e.g. a small load carrier containing the bolts in random poses is placed in the robot's workspace. In addition, a fixture is used to store the objects in a defined position after grasping.

Grasping the small metallic objects directly from the box is complex and challenging. Due to overlapping of the randomly oriented parts, difficulties arise in finding suitable grasping poses and high accuracy is required when grasping the objects. Furthermore,

reflections occur causing noisy or fautly depth measurements and difficulties in identifying the bolts. Therefore, a two-stage procedure is proposed.

3.2 Two-Stage Procedure for Bin Picking with Magnetic Gripper

Due to the challenges described above a two-stage procedure is presented for the bin picking task consisting of a blind i.e. visionless grasp into the box in the first stage and the grasping of individual bolts from the work surface in the second stage. The procedure is shown in Fig. 1.

An image of the workspace is taken in a defined scan pose using the vision system attached to the industrial robot. The position of the camera is parallel to the work surface at a predefined height. If no bolt is detected, there is a blind grasp whereby the robot moves the magnetic gripper into the box without using visual information. Some bolts are grasped and afterwards placed on the work surface next to the box.

If a bolt is detected in the image, its pose is estimated, the bolt is grasped and placed in the fixture. The process is repeated until the required number of bolts is reached resp. the fixture is completely equipped.
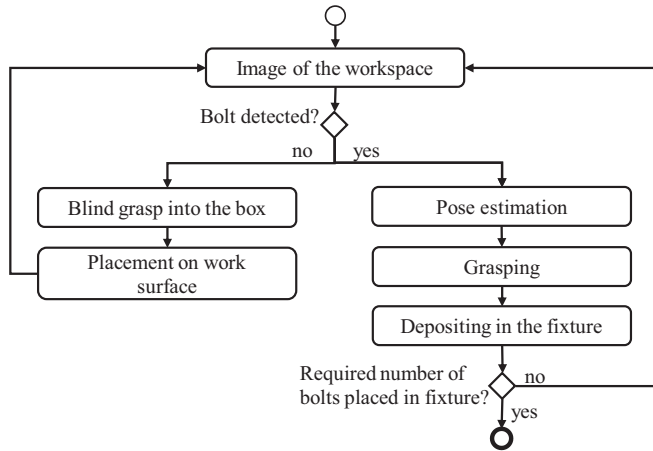
**Fig. 1** Two-stage procedure for bin picking with magnetic gripper

3.3 Grasping Process and Pose Estimation

A custom made magnetic gripper consisting of an electromagnet with a microcontroller is used to grasp the bolts. Process knowledge is used to control the force of the magnetic gripper depending on the type of bolts (especially the weight) and the intended picking

process (first step or second step). When grasping blindly into the box in the first step, the force is strong enough to pick up several bolts and place them on the work surface. Subsequently, in order to grasp one bolt, the tip of the magnetic gripper is placed at the head of the bolt and an appropriate force is set to grasp exactly one bolt.

Therefore, the pose of the bolt is determined using the implemented computer vision system. The bolts lying on the work surface have three degrees of freedom (DOF), two translational and one rotational DOF. Applying a previously trained convolutional neural network (CNN), the positions of the bolt as well as the bolts' head are determined in the image. The center points of the bounding boxes of these two object classes are provided. In addition to the position of the bolt in x- and y-direction, the orientation of the bolt is determined using the two center points and the corresponding angle β, as depicted in Fig. 2.

Fig. 2 Position of bolt lying on the work surfaces a) side view and b) top view
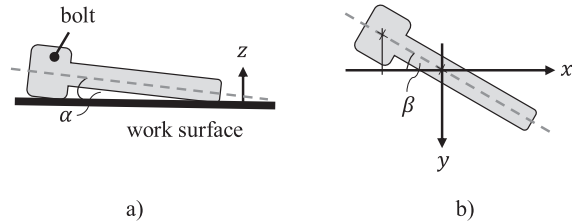
4   Implementation and Evaluation

To evaluate the described system, it is implemented as depicted in Fig. 3. A lightweight robot UR10 from Universal Robots is used to automate the process and an Intel RealSense L515 LIDAR is used to capture the environment (see Fig. 3a). The sensor uses the time-of-flight principle and provides a point cloud of the environment with a maximum resolution of 1024×768 spatial points. In addition, a colour image of the environment with a maximum resolution of 1920×1080 pixels is captured and superimposed with the generated point cloud. This allows the pixel coordinates of the colour image to be converted to the corresponding spatial points.
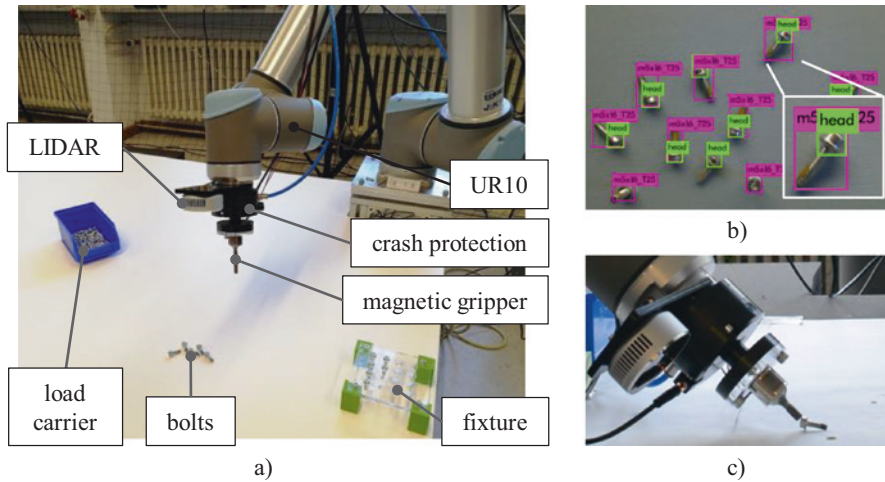
Fig. 3 **a** Implementation of the overall system, **b** identification of bolts using YOLOv3 and **c** grasping position of an individual bolt

The CNN YOLOv3 is used to identify the bolts. To train the CNN, 350 images were manually annotated. Thereby, the complete bolt as well as the head of the bolt are annotated individually in order to distinguish the respective parts during recognition. For training, the annotated image dataset is divided into the actual training data, the validation data and the test data in a ratio of 80, 10 and 10%. The hyperparameters for the training were chosen as listed in Table 1. After completing the training, the model achieved a mean average precession of 95% on the test set. Figure 3b shows the identification of the bolts using YOLOv3 with bounding boxes of the complete bolt (purple) as well as the heads (light green).

Table 1 Hyperparameters used for the training of YOLOv3

Hyperparameter	Value
Batch size	64
Subdivisons	16
Iterations	6000
Initial learning rate	0.001
Steps and corresponding scales	4800 (scale: 0.1), 5400 (scale: 0.1)

The developed magnetic gripper consists of an electromagnet with a diameter of 25 mm, a length of 20 mm and a maximal retention force of 50 N. The electromagnet is switched via a bridge circuit and the overall magnetic gripper is controlled via an Arduino Uno, whereby the magnetic field strength can be adjusted via a pulse-width modulated signal. An additional crash protection is installed between the gripper and the robot flange to avoid damage to the robot in the case of a faulty gripping attempt.

The software required to achieve an automated grasping and separation process is implemented using the Robot Operation System (ROS). The darknet_ros package can be used for the integration of YOLO. The motion planning for the UR10 is done using the MoveIt framework. The calculation of the gripping pose, the control of the magnetic gripper as well as the sequence control is implemented using the middleware provided by ROS. The whole software including object recognition, gripping pose calculation and motion planning is executed on an NVidia Jetson AGX Xavier.

To evaluate the presented system multiple test runs are performed. During the test runs M5 cylinder head bolts are used as gripping objects. The objective of every test run is to place 16 bolts in the fixture. The process starts by grasping multiple bolts from the box and placing them on the work surface. The bolts are then grasped individually (see Fig. 3c) and inserted into the holes of the fixture. Once no more bolts are detected on the work surface, new bolts are grasped from the box. Each test run continues until the fixture is fully equipped or an error occurs.

No errors or faults occurred during the runs when grasping the bolts out of the box. In every iteration, between two and five bolts were grasped from the box and placed on the work surface. In the subsequent process of grasping and placing the individual bolts, 70% of the bolts were grasped successfully and 60% of the bolts could be deposited successfully in the fixture.

The following issues caused unsuccessful placements in the fixture and failed grasping attempts. An unsuccessful placement of a bolt was always connected to a faulty pose estimation resp. a faulty grasping attempt. When the bolt is not centred beneath the gripper or tilted after grasping it from the work surface, it cannot be placed in the fixture correctly. Since currently no optical verification of the correct grasping pose is integrated, a wrongly oriented or tilted bolt is placed next to a hole when it is inserted into the fixture. This can also result in an unacceptably high compression force and an emergency stop of the robot.

Unsuccessful or faulty grasping attempts are mainly caused by bolts lying close together. In this cases, the bolt and its head cannot be recognized unambiguously and thus, no or an invalid grasping pose is calculated. When no bolt was grasped from the work surface, the process continues, but the space in the holder remains empty after the placement is completed.

Another error was residual magnetization of the bolts after insertion into the fixture, which prevented it from being released from the gripper. However, this error can be reliably prevented by moving the gripper away at an angle after insertion.

5   Summary and Outlook

The system presented in this paper enables the flexible and cost-effective separation of bolts using an industrial robot and a magnetic gripper. After separation, the bolts are in a known pose, allowing them to be inserted directly into a custom fixture. As the depth data of the small and reflective bolts is noisy and error-prone, a two-stage process is used to separate the bolts. First multiple bolts are grasped from the box and are placed on an even surface. Afterwards, the object detection and pose estimation is performed to grasp a single bolt in a defined manner. The presented implementation and evaluation demonstrates the functionality and potential of the system. Further test runs have to be performed with different bolt variants.

In order to address the errors encountered during the evaluation, the following improvements and enhancements will be made in the next development step. An additional colour camera will be integrated to check if a bolt has been successfully grasped from the surface and whether it is in the required pose. If not, the bolt is put down again and the grasping process is repeated. Furthermore, after placing the bolt in the fixture, the colour camera of the LIDAR should be used to check that the bolt has been inserted correctly.

Moreover, a combined force and position control will be used while inserting a bolt in the holes of the fixture to compensate for small deviations in the placement position. As shown by Metzner et al., the application of a suitable compensation strategy can significantly increase the success rate when inserting objects into holes [13].

Finally, after integrating the improvements mentioned above, an evaluation of the overall process with different bolt types will be carried out and the success rate when loading different fixtures will be evaluated.

References

1. Du, G., Wang, K., Lian, S., Zhao, K.: Vision-based robotic grasping from object localization, object pose estimation to grasp estimation for parallel grippers: a review. Artif. Intell. Rev. **54**(3), 1677–1734 (2021)
2. Bochkovskiy, A., Wang, C., Liao, H.M.: YOLOv4: optimal speed and accuracy of object detection. ArXiv abs/2004.10934, pp. 1–17 (2020)
3. Wang, R.J., Li, X., Ling, C.X.: Pelee: a real-time object detection system on mobile devices. In: Proceedings of the 32nd International Conference on Neural Information Processing Systems (NIPS'18), pp. 1967–1976. Curran Associates Inc., Red Hook (2018)
4. Liu, H., Rivera Soto, R.A., Xiao, F., Jae Lee, Y.: YolactEdge: real-time instance segmentation on the edge. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 9579–9585. IEEE, Xi'an (2021)
5. Sun, Y., Zuo, W., Liu, M.: RTFNet: RGB-thermal fusion network for semantic segmentation of urban scenes. IEEE Robot. Autom. Lett **4**(3), 2576–2583 (2019)

6. Mathiesen, S., Iturrate, I., Kramberger, A.: Vision-less bin-picking for small parts feeding. In: IEEE 15th International Conference on Automation Science and Engineering (CASE), pp. 1657–1663. IEEE, Vancouver (2019)

7. Ishige, M., Umedachi, T., Ijiri, Y., Taniguchi, T., Kawahara, Y.: Blind bin picking of small screws through in-finger manipulation with compliant robotic fingers. In: 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 9337–9344. IEEE, Las Vegas (2021)

8. von Drigalski, F., Hayashi, K., Huang, Y., Yonetani, R., Hamaya, M., Tanaka, K., Ijiri, Y.: Precise multi-modal in-hand pose estimation using low-precision sensors for robotic Assembly. In: 2021 IEEE International Conference on Robotics and Automation (ICRA), pp. 968–974. IEEE, Xi'an (2021)

9. Furukawa, H., Sato, J., Yamada, T., Ito, K., Ito, S.: Grasping position detection using template matching and differential evolution for bulk bolts. In: IECON 2019 - 45th Annual Conference of the IEEE Industrial Electronics Society, pp. 5622–5627. IEEE, Lisbon (2019)

10. Nakano, Y.: Stereo Vision Based Single-Shot 6D Object Pose Estimation for Bin-Picking by a Robot Manipulator. ArXiv abs/2005.13759, pp. 1–7 (2020).

11. Sato, Y., Harada, K., Sakata, N., Wan, W., Ramirez-Alpizar, I.G.: Two-stage picking method for piled shiny objects. In: Uhl, T. (ed.) Advances in Mechanism and Machine Science, vol. 73, pp. 2049–2058. Springer International Publishing, Cham (2019)

12. Blank, A., Hiller, M., Zhang, S., Leser, A., Metzner, M., Lieret, M., Thielecke, J., Franke, J.: 6DoF pose-estimation pipeline for texture-less industrial components in bin picking applications. In: 2019 European Conference on Mobile Robots (ECMR), pp. 1–7. IEEE, Prague (2019)

13. Metzner, M., Leurer, S., Handwerker, A., Karlidag, E., Blank, A., Hefner, F., Franke, J.: High-precision assembly of electronic devices with lightweight robots through sensor-guided insertion. Procedia CIRP **97**(6), 337–341 (2021)