

Cooperative Bin-Picking with Time-of-Flight Camera and Impedance Controlled DLR Lightweight Robot III

Stefan Fuchs, Sami Haddadin, Maik Keller, Sven Parusel, Andreas Kolb and Michael Suppa

Abstract— Because bin-picking effectively mirrors great challenges in robotics, it has been a relevant robotic showpiece application for several decades. In this paper we describe the computer vision algorithms in combination with the sophisticated control schemes of the robot and demonstrate a reliable and robust solution to the chosen problem. This paper approaches the bin-picking issue by applying the latest state-of-the-art hardware components, namely an impedance controlled lightweight robot and a Time-of-Flight camera. Lightweight robots have gained new capabilities in both sensing and actuation without suffering a decrease in speed and payload. Time-of-Flight cameras are superior to common proximity sensors in the sense that they provide depth and intensity images in video frame rate independent of textures. The bin-picking solution presented in this paper aims at extending the classical bin-picking problem by incorporating an environment model and allowing for the physical human-robot interaction during the entire process. Existing imprecisions in Time-of-Flight camera measurements and environment uncertainties are compensated by the compliant behavior of the robot. The overall process is implemented in a generic state machine that also monitors the entire bin-picking process.

I. INTRODUCTION

Bin-picking has been an important showcase application in robotics for several decades. On the one hand, it is of great economic interest as it is one of the last puzzle pieces for achieving fully automated industrial processes. These are basically confronted with the problem of designing an interface between machine and human environment.

Although state-of-the-art manufacturing processes are mostly fully automated, human operators still perform monotonous and partially dangerous tasks especially at the beginning and at the end of the process by either feeding the machine with parts or packing the output. Indeed mostly human operators are replaced by mechanical vibration bowls, structured bins or fixations. Therefore, these solutions are specific, inflexible and expensive in maintenance. On the other hand, bin-picking is of scientific interest because it merges several specific robotic issues such as object recognition and localization, grasp planning, path planning, and collision avoidance, into one common goal. For that reason bin-picking is consistently chosen to demonstrate the effectiveness of the aforementioned algorithms.

However, the transfer of many approaches into an industrial product has not yet fully succeeded. This is to a large extent due to the mere focus on fast recognition algorithms and/or sensors. Such approaches do not take

into account any unpredictable environmental conditions, which severely complicate the task. Usually, the developed solutions are based on simplified assumptions in order to increase recognition speed. Up to now robots are not yet able to compete with the aforementioned mechanical solutions, mainly because recent robotic bin-picking applications do not yet match the industrial standards in accuracy, robustness, speed, and safety.

Besides these issues the presented work is motivated by an additional thought. The first robotic generation is characterized by all-purpose industrial robots replacing manpower for mass production. These robots perform stultifying tasks in structured environments, caged and separated from humans for safety reasons. The second industrial robotic generation will be characterized by highly sensorized lightweight robots with sophisticated soft-robotics features. These robots will be able to serve as an uncaged co-worker for humans significantly enriching possible application scenarios. In 2009, we presented an approach for achieving robust and safe co-worker capabilities [9] with a *DLR light-weight robot III* (LWR-III), whose technology was recently transferred to the robot manufacturer KUKA.

In this paper we present a robust bin-picking scenario, which focuses on the concept of a robotic co-worker including the following contributions:

- The bin-picking task is performed by combining a Time-of-Flight (ToF) camera, state-of-the-art computer vision, and a LWR-III.
- The impedance controlled robot compensates inaccuracies in the localization and ensures a successful item grasping.
- An environment model, which is based on a fast adaptive volumetric data structure, is used for the observation of the dynamic workspace.
- During the entire process the human is able to physically interact with the robot. The user can e.g. initiate interaction schemes at any time during the process.

II. RELATED WORK

The essential sub tasks for the isolated bin-picking problem are object recognition and localization, path planning, and grasp planning. They already represent wide fields of research in the robotics and computer vision community. Usually, they are considered separately from concrete applications as e.g. bin-picking.

In 1982 Horn and Ikeuchi [11] made the first attempt on integrated bin-picking. Torus objects were localized with a shape from shading approach. The robot moved towards the estimated grasping point until smart LEDs inside the gripper indicated the grasp. In 1998 Ghita and Whelan [5]

S. Fuchs, S. Haddadin, S. Parusel and Michael Suppa are with the German Aerospace Center, Inst. of Robotics and Mechatronics. stefan.fuchs@dlr.de

M. Keller and A. Kolb are with the Computer Graphics and Multimedia Systems Group, University of Siegen, Germany maik.keller@uni-siegen.de

presented a bin-picking approach of unpiled objects using an eigenimage analysis for object recognition and a depth from defocusing technique for object localization. In 2001 Kristensen et al. [13] outlined a bin-picking system for polyhedral objects in an industrial robotic work cell. A solid state range camera provided depth and intensity images with a resolution of 640×480 pixels in 6.7 Hz. The grasping was performed in 3 degrees of freedom (DoF) with an accuracy of at least 0.01 m. In 2003 Boughorbel et al. [3] also combined video and range images. They generated an accurately textured 3D model of the scene. The manipulator was tracked in order to validate the object pose estimates. Thus, the robustness against localization errors was increased by adjusting the grasp path. Moreover, in 2007 Leonard et al. [15] addressed possible collisions by introducing oriented bounding boxes (OBB). While a position-based visual servoing approach kept the target object within the camera field of view, the OBB allowed for online collision detection.

So far, the bin-picking was primarily investigated from the computer vision point of view whereas the robot was mainly used as an accurate 6 DoF grasping device. Completely reliable object localization was considered to be the key issue and in a look-and-move manner dynamic environments were neglected. In the course of time a mighty toolbox of object recognition and localization methods has been developed. These approaches count on the complete knowledge of process and environment, and are thus prone to errors coming from model inaccuracies and dynamic environments. Failure is often not detectable or even considered at all. Bin-picking concepts tackling pose estimation errors, e.g. by visual servoing, or fusing information from different sensors were only developed to a minor degree.

This paper accounts for these drawbacks and holistically approaches the bin-picking issue. The ToF sensor and vision algorithms allow for dynamic environments and arbitrary objects. The robot is able to compensate the uncertainty of the computer vision or recognize failure in grasping. Along with a disturbance observer for detecting and estimating external forces (see Haddadin et al. [6]), a state machine monitors the entire application, enables seamless human-robot interaction, and ensures the safety of the human operator.

The remainder of this paper is structured as follows: In Sec. III we introduce the hardware elements and the underlying communication structure. Section IV surveys the computer vision and control methods utilized in this work. In Sec. V we describe the implemented process. Finally, Sec. VI discusses the experimental results, and Sec. VII contains the conclusion.

III. HARDWARE COMPONENTS AND COMMUNICATION

Figure 1 shows the design of the bin-picking demonstration model with a LWR-III and a ToF camera. Figure 2 reveals the concealed methods and the underlying communication structure. In the following paragraph, we introduce the hardware components and the communication structure.

A. The DLR Lightweight Robot III

The LWR-III realizes various features that are crucial for both the bin-picking task and direct interaction with humans.

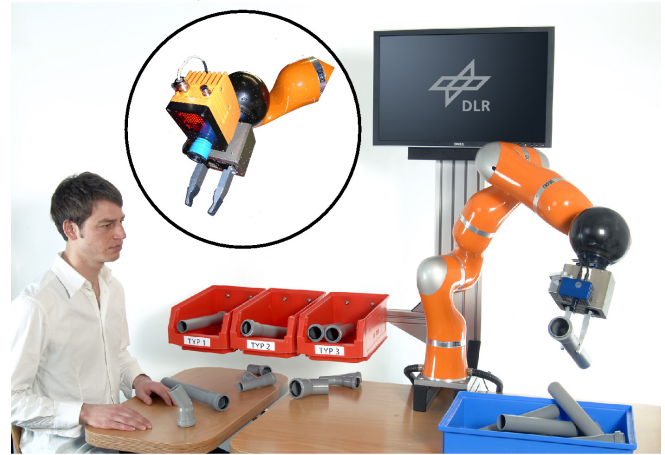


Fig. 1. Cooperative bin-picking scenario with DLR light-weight robot (LWR-III) and ToF camera. Two types were used: the IFM ToF camera with 204×204 pixel resolution is displayed encircled. On the right hand side, the SwissRanger SR3100 with 176×144 pixel resolution is attached to the robot.

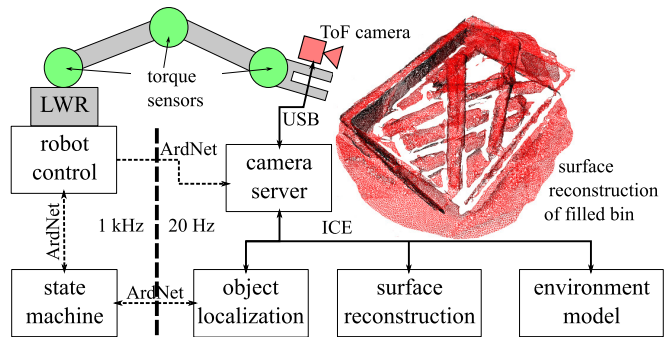


Fig. 2. Schematic illustration of the system architecture. The *state machine* is the interface between robot control and computer vision. The camera server provides intensity and depth images to the object localization, surface reconstruction and environment model modules.

It weighs only 14 kg and handles a payload up to the same weight¹. The LWR-III is especially characterized by its soft-robotics features [1]. As it is equipped with torque sensors in every joint, it makes possible both impedance control and accurate position control at the same time. Together with a very accurate dynamic model, the torque sensors enable the detection of contacts and the estimation of contact forces along the entire robot structure [6]. Furthermore, the robot can identify the payload and detect payload loss online without additional force/torque sensing in the wrist.

B. Time-of-Flight Camera

Nowadays, ToF sensors measure full-range distance information by estimating the elapsed time between emission and receiving of active light. Such sensors are inexpensive, compact and have a high performance. The specific type of ToF sensors which is used in the bin-picking application is based on the technology of *Photonic Mixing Devices (PMD)*. Its main component is an array of smart pixels (see

¹The nominal payload according to industrial long range testing is 7 kg. However, for research purposes the robot can carry its own weight.

Lange [14] and Xu et al. [16]), which provides resolutions of 64×48 to 204×204 pixels at 20 Hz. Due to an automatic suppression of background light, current devices are suitable for both indoor and outdoor scenes. The ToF camera emits sinusoidal modulated Near-Infrared (NIR) light. The NIR light is reflected by the observed scene and the camera optics project the light onto the smart pixel array. Distances and intensity values are then computed within every pixel by sampling the individual auto-correlation four times.

The PMD technology causes several error sources, which influence the accuracy of the distance measurement. Dynamic motion blurring and resolution artifacts cause so-called flying pixels. The typical distance-related error is also an important effect of such sensors. Flying pixels arise when an area of inhomogeneous depth is covered by a single PMD-pixel. The motion blurring occurs in dynamic scene setups since the four correlation samples lead to varying object points observed by the respective smart pixels during the temporal integration. The distance-related error is based on the non-harmonic properties of the optical signal, i.e. the signal is not perfectly sinusoidal. Fuchs et al. [4] describe an appropriate calibration method which identifies this error component. Using amplitude and depth images of a checkerboard, which are captured from different points of views, a common photogrammetric intrinsic and extrinsic calibration is performed initially. The intrinsic parameters characterize aberrations caused by the optics, such as skew, focal length, and optical center. Given an external positioning system, e.g. a robot, the extrinsic calibration yields the hand-eye-transformation. Then, the distance-related error is identified in a depth calibration step. As a result, an overall mean accuracy of 3 mm is achieved in distance measurements.

C. Communication

There are two communication spaces. The robot control kernel runs a cycle of 1 kHz on VxWorks and is controlled by an asynchronous communication protocol [8]. For stability reasons, the data between robot and camera has to be exchanged online, while keeping latencies small. Therefore, the data (also among network nodes) is stored in shared memories and exchanged via the proprietary *ARDNET* protocol (see Bäuml et al. [2]).

The ToF camera produces a large amount of data: 800 kB per frame at 20 Hz. The camera server polls distance and amplitude images from the ToF camera, fuses them with the robot pose and provides them for registered modules as e.g. localization, surface reconstruction and environment model. The registered modules run on both Windows and Linux, and are connected via the *Internet Communications Engine (ICE)* (see Henning [10]). This object-oriented middleware is designed for high performance (by using a binary protocol) and low verbosity. Since ICE takes care of all interactions with low-level network programming interfaces, we are able to run modules such as object localization as well as the environment model on a remote computer connected via Internet.

IV. METHODS

The central element of the bin-picking scenario is a hybrid state machine which monitors the entire process. It

processes object poses from the localization and commands the robot. The state machine decides, for example whether an object shall be grasped or the robot needs to stop due to a recognized collision. In the following paragraphs we describe the elementary methods, which support the state machine.

A. Environment Model

The major purpose of the environment model is the temporal accumulation of sensor data into a consistent data basis, which is then available for further processing algorithms, e.g. the bin-localization (see Sec. V-B). In our particular application of bin-picking with ToF cameras the environment model is used for the observation of the dynamic workspace.

1) *Data accumulation:* Due to the large temporal resolution of ToF cameras it is necessary to process the input data with approximately the same frequency. Therefore, we chose the underlying data structure to rely on an implementation of *Dynamic Volume Trees (DVT)* (see Keller et al. [12]), which can be modified in real-time. This adaptive and hierarchical volume data structure achieves its online capability by realizing both, the hierarchical data structure and the manipulation of the data structure solely on the *Graphics Processing Unit (GPU)*. The following steps describe the hierarchical rasterization of the range images provided by the ToF sensor:

1. Perform hierarchical rasterization by traversing the volume tree and select the nodes which are intersected by the respective depth values of the provided range image.
2. Test only those depth values for their node intersection according to the space which is covered by the specific node. The transformation of the range image data into the tree coordinate system relies on the robot pose.
3. Refine the selected nodes until the desired target resolution is reached.

Finally, the resulting subtree is merged with the geometry represented in the current data structure and the next range image is ready to be processed.

2) *Virtual Views:* The environment model offers an interface for the request of so-called *virtual views*. This supports fast access to the data of the accumulated scene for further processing steps. For this purpose, the environment model generates scene views based on the information of the requested pose and the resulting image properties such as the resolution. The graphics hardware is used for the synthesis of the respective view based on a virtual camera, which is then available in terms of a range image (see Fig. 3). The further processing algorithms benefit from the possibilities of placing a virtual camera everywhere in the scene and obtaining images with the desired resolution from a complete workspace overview to a close up, for instance.

B. Object Localization and Tracking

The localization is able to cope with a number of basic geometric objects that feature edges and surfaces. The objects are described by a combined point and normal model. The model can either be generated from CAD models or

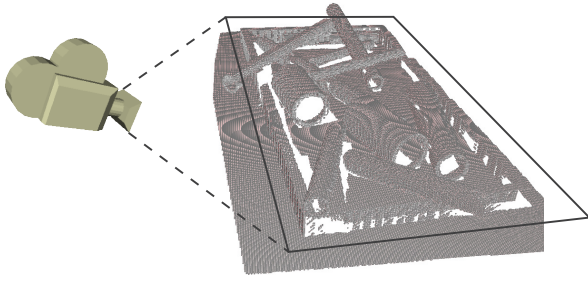


Fig. 3. A virtual camera within the environment model renders a virtual view as shown in Fig. 6(a).

from surface reconstruction. An Iterative Closest Point (ICP) algorithm estimates the object pose in 6 degrees of freedom (DoF). In order to accelerate the registration process and reduce erroneous correspondences which lead to incorrect pose estimations, a three staged localization architecture is implemented (see Fig. 4).

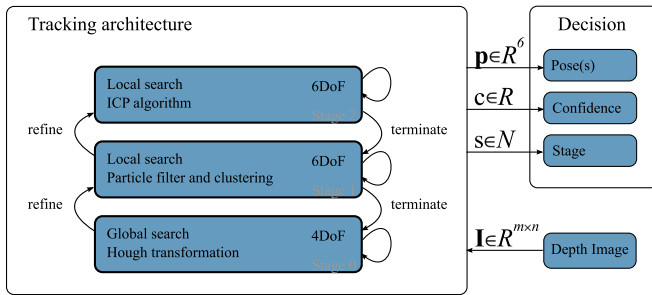


Fig. 4. Multi-stage tracking architecture. The incoming real or virtual depth images are processed by one of these stages. Each stage provides a list of hypotheses tagged with a confidence value. Depending on the confidence value the stages are refined or terminated.

At each stage a different algorithm processes an incoming depth image and provides a list of pose hypotheses, which are additionally tagged with a confidence value for the potential object. The stages are continuously monitored and executed according to suitable termination criteria or reentered for refinement.

The first stage performs a global search, consisting of edge filtering and a Hough transform for identifying lines, which are used as initial hypotheses for the pipe poses. Crossing lines indicate a bin pose. At the second stage, these hypotheses are locally consolidated and clustered by a particle filter. Thirdly, the ICP provides an accurate 6 DoF pose of the target object in 20 Hz according to the frame rate of the ToF camera. Both ICP and particle filter directly process 3D data and a 3D model of the target.

C. Compliant Control and Disturbance Observer

The control of the LWR-III is significantly involved in realizing the bin-picking while providing the relevant control features for physical human-robot interaction. The LWR-III has several modes of operation. In this paper we use Cartesian impedance control and zero-gravity torque control (see Albu-Schäffer et al. [1]). The Cartesian impedance control allows the task space impedance to be freely assigned

and causes the robot to behave compliant during physical interaction with the environment. Thus, dexterous grasps and assembling tasks are feasible despite uncertainties in the localization of target objects. Furthermore, the controller considers obstacles, which are represented as geometric primitives, as spheres or planes, for instance, whose surface normals describe virtual forces affecting the robot through its direct torque control interface. Thus, unexpected collisions with these known items are prevented. The zero-gravity torque control enables the human operator to intuitively configure the robot, especially during teaching operations.

Apart from coping with known environments, the robot is capable of coping with unforeseen contact. If such a physical collision is detected, the common reaction is usually to urgently stop the robot. However, there are two drawbacks: Firstly, the robot freezes in a probably inappropriate configuration and secondly the efficiency of the performed task is drastically reduced. For that reason a disturbance observer [6] does not only detect contacts but also classifies them according to their “severity level”. This is defined according to the evaluated severity levels for humans during worst-case robot-human impacts (see Haddadin et al. [7]).

V. OVERALL CONCEPT AND IMPLEMENTATION

Figure 5 displays the flow chart of the entire implementation, which can be divided into three sub loops: exploration, bin-picking and human-robot interaction. All loops are managed by a state machine and supervised by safety layers using e.g. the aforementioned disturbance observer.

A. Safety Architecture

The safety architecture is composed of four hierarchical layers, which interpret the robot sensor data permanently. Thus, not only physical contacts with the environment are detected and classified, but also malfunctions of the robot are recognized. Appropriate reactions are triggered depending on the different *collision severity stages* and current states. In the worst case the bottom *emergency layer* activates the emergency brakes. The second layer commands a safe stop. The third layer switches from impedance control to zero-gravity torque control and lets the robot react in a convenient compliant manner. The top layer supports the bin-picking process by indicating failed grasps due to excessive process forces or providing break criteria, for example. During the transport phase gentle contacts at the robot structure are recognized as a command to interact with the user. Then, the robot hands the item over to the user. Nudging the robot again is recognized as a signal for taking back the item. This example demonstrates the fault tolerant and situation suited system behavior.

B. Scene Exploration

Before starting the bin-picking task the entire workbench is explored due to the fact that the path planning and collision avoidance have to consider appearing and disappearing obstacles. Since ToF cameras provide a small apex angle of about 45° combined with a comparatively low lateral resolution, the localization of small objects is feasible only at close distances. Furthermore, the working range of the robot

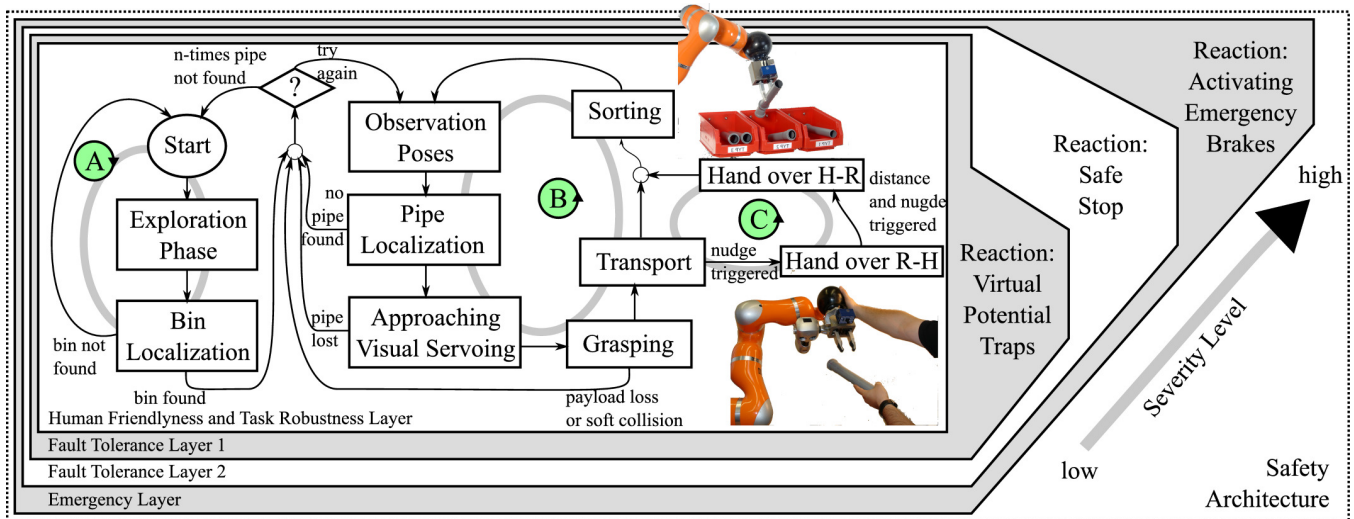


Fig. 5. The state machine (SM) represents the entire bin-picking process and is embedded into the topmost layer of the safety architecture, the *human-friendliness and task robustness layer*. Basically, the SM manages three sub-loops. In loop A the workbench is explored, and the environment model is created. In loop B the bin is cleared. Furthermore, we integrated various interaction schemes during task execution. In the phase of transporting the grasped item, the user can e.g. initiate a hand-over, which is represented by loop C.

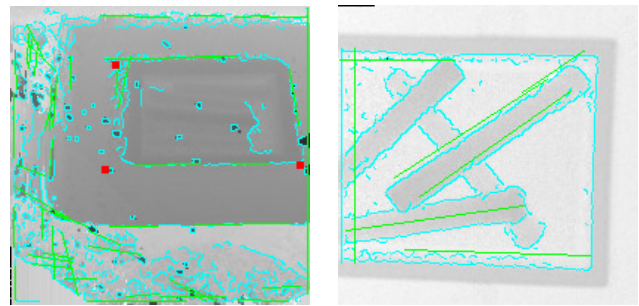
does not allow the observation of the complete workbench at one point in time. Thus, capturing the entire scene can only be achieved by accumulating several images into the environment model, which is used to localize the bin and feeds the Cartesian impedance control (see IV-C). More precisely, the robot is moved to various positions in order to explore the overall workspace. Since all captured images are transferred into the environment model directly, virtual views are requested in order to evaluate the complete workbench in a single depth image.

C. Bin-Picking

The tracking framework (TF) is able to localize two object types: the bin and the pipe. The bin's size is $430 \text{ mm} \times 280 \text{ mm} \times 100 \text{ mm}$. The pipe has a diameter of 40 mm and a length of 300 mm .

In loop A the TF queries the environment model for a virtual depth image. The virtual point of view, the resolution and the apex angle are tuned for covering the entire workbench with a single depth image. The TF stages consecutively process the virtual depth image for localizing the bin (see Fig. 6(a)). If the bin is not found, then either the virtual point of view is changed or the scene is explored again.

If the bin is found, the state machine switches to loop B and the camera moves to different observation points that are located 500 mm above the bin (see Fig. 6(b)). The search space is reduced and the image processing is accelerated by using the known bin pose. The pipe diameter of the narrow pipes is only represented by 20 pix in the depth images because of the low resolution. Together with the measurement noise of $\sigma \approx 10 \text{ mm}$, the pipe poses can only be localized with a significant uncertainty. For this reason, the camera is moved towards the pipe in an *open-loop visual servoing* manner. At a distance of only 300 mm the pipe completely fills the depth image with a diameter of $\approx 30 \text{ pix}$. If the increased reliability of the pose hypothesis falls below



(a) Edge Filtered Virtual View (b) Edge Filtered Real Range Image

Fig. 6. (a) The point of view for this virtual depth image is tuned in order to overview the entire workbench. Firstly, an edge filter (cyan) and a Hough transform identify lines (green). The three-dimensional line parameters are estimated and orthogonally crossing lines indicate bin corners (red dots). (b) This real depth image is captured from $\approx 500 \text{ mm}$ above the bin. The identified green lines are initial pose hypotheses of the pipes.

a given threshold, the state machine calls on the robot to grasp the pipe. Otherwise, the localization is restarted from a different observation point.

The accuracy in localization was experimentally evaluated. An object laying on the workbench is tracked by the moving camera. Given the robot positioning system and the hand-eye calibration, the global object pose is computed in each frame. Ideally, this global object pose should be constant, but apparent motion bares measurement errors. In this case, the object is localized and tracked with an accuracy of $\approx 7 \text{ mm}$.

D. Human-Robot Interaction

In loop C the TF is not used and the ToF camera simply serves as a proximity sensor. After handing the object over to the human operator, the state machine is idle. The object release is indicated, as the operator gives the pipe to the gripper. If the ToF camera recognizes an object between the

jaws or physical contact is made with the robot, the state machine calls on the robot to grasp the pipe. Returning to loop B the pipe is sorted in.

E. Impedance Control

The robot's compliant behavior allows for robust grasping despite the aforementioned recognition uncertainties. The grasping strategy shown in Fig. 7 successfully copes with possible translational deviations in the range of the jaw distance, which is 55 mm, before the grasp fails. This is far beyond the provided uncertainty of the computer vision. The last image in Fig. 7 shows a case expected to be a failure. We implemented rotational stiffness along the axis perpendicular to the image plane. Due to the compliant behavior of the robot and gripper-object as well as object-ground friction, the object is rotated into the firm grasp.

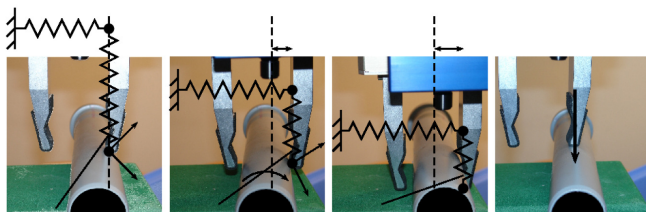


Fig. 7. Impedance controlled grasp. The springs illustrate the impedance control in the principal axis of the pipe. The gripper is not accurately positioned on purpose. Still, the gripper is shifted to the right, the pipe is slightly rotated during downward motion, and a firm grasp is performed. The rightmost image demonstrates a failure, where the surface normal of the pipe is in parallel with the moving direction of the gripper. However, due to advantageous placing of the reference frame for the Cartesian compliance, we are able to grasp this very difficult situation, too.

VI. QUANTITATIVE RESULTS

The efficiency and robustness of our approach was tested in a series of autonomous grasps. For this evaluation we replenished the bin after each successful grasp in order to have independent trials. On average, the robot needed 6.4s for one grasping process, which comprises object detection from an arbitrary viewing position, approaching and grasping, unbagging, and moving back to the initial viewing position. The robot was able to grasp an object in every cycle for 80 trials, i.e. the overall cycle success rate was 100%. This result could only be achieved due to the fault tolerance capabilities of the system during the entire process, such as the detection of a physical impossibility of a planned grasp, the non-successful grasp (overall 3 times), losing an object in tracking, or localization without any result. All of these failure modes were detected or realized by the system and induced a restart of the grasping process. As a result, the number of average views to recognize an object was $N_{\text{view}} = 2.2$.

VII. CONCLUSION

This paper presents an integrated bin-picking concept for the latest state-of-the-art technology which is commercially available. An impedance controlled lightweight robot and a ToF camera combine their respective capabilities in order to tackle the problem efficiently. In connection with innovative computer vision, the ToF camera is used for fast modeling

of the dynamic environment and for localizing the bin plus the objects therein. The soft-robotics concepts of the lightweight robot enable collision detection, safe human-robot interaction and provide high robustness. The fusion of soft-robotics and computer vision leads to high fault tolerance. Thus, the grasping is performed completely successfully, although the computer vision partially provides inaccurate object poses.

VIII. ACKNOWLEDGMENTS

This work has been partially funded by the German Federal Ministry of Education and Research (BMBF) under Contract No. 16SV2298 as part of the project Lynkeus (www.lynkeus-3d.de) and by the Kuka Roboter GmbH, Augsburg, Germany.

REFERENCES

- [1] Alin Albu-Schäffer, Christian Ott, and Gerd Hirzinger. A Unified Passivity-based Control Framework for Position, Torque and Impedance Control of Flexible Joint Robots. *Int. J. of Robotics Research*, 26:23–39, 2007.
- [2] Berthold Bäuml and Gerd Hirzinger. When hard realtime matters: Software for complex mechatronic systems. *Robotics and Autonomous Systems*, 56(1):5–13, 2008.
- [3] Faysal Boughorbel, Yan Zhang, Sangkyu Kang, Umayal Chidambaram, Besma Abidi, and Andreas Koschan. Laser ranging and video imaging for bin picking. In *Assembly Automation*, volume 1, pages 53–59, 2003.
- [4] Stefan Fuchs and Gerd Hirzinger. Extrinsic and Depth Calibration of ToF-Cameras. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR2008)*, Anchorage, USA, pages 1–6, 2008.
- [5] Ovidiu Ghita and Paul F. Whelan. Robust robotic manipulation. In *Proceedings of the SPIE - Intelligent Robots and Computer Vision XVII*, volume 3522, pages 244–254, Boston, USA, 1998.
- [6] Sami Haddadin, Alin Albu-Schäffer, Alessandro De Luca, and Gerd Hirzinger. Collision Detection & Reaction: A Contribution to Safe Physical Human-Robot Interaction. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2008)*, Nice, France, 2008.
- [7] Sami Haddadin, Alin Albu-Schäffer, and Gerd Hirzinger. Safe Physical Human-Robot Interaction: Measurements, Analysis & New Insights. In *International Symposium on Robotics Research (ISR2007)*, Hiroshima, Japan, pages 439–450, 2007.
- [8] Sami Haddadin, Sven Parusel, Rico Belder, Jörn Vogel, Tim Rokahr, Alin Albu-Schäffer, and Gerd Hirzinger. Holistic design and analysis for the human-friendly robotic co-worker. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS2010)*, Taipei, Taiwan, 2010.
- [9] Sami Haddadin, Michael Suppa, Stefan Fuchs, Tim Bodenmüller, Alin Albu-Schäffer, and Gerd Hirzinger. Towards the robotic co-worker. In *International Symposium on Robotics Research (ISR2007)*, Lucerne, Switzerland, 2009.
- [10] Michi Henning. A new approach to object-oriented middleware. *IEEE Internet Computing*, 8(1):66–75, 2004.
- [11] B. K. Horn and K. Ikeuchi. Picking parts out of a bin. Technical report, Massachusetts Institute of Technology, 1982.
- [12] Maik Keller, Nicolas Cuntz, and Andreas Kolb. Interactive Dynamic Volume Trees on the GPU. In *14th International Workshop on Vision, Modeling, and Visualization*, 2009.
- [13] Steen Kristensen, Stephane Estable, Matthias Kossow, and Ralf Brösel. Bin-picking with a solid state range camera. *Robotics and Autonomous Systems*, 35(3-4):143–151, 2001.
- [14] R. Lange. *3D Time-Of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology*. PhD thesis, University of Siegen, 2000.
- [15] S. Leonard, A. Chan, E. Croft, and J.J. Little. Robust motion generation for vision guided robot bin-picking. Seattle, Washington, USA., November 2007.
- [16] Z. Xu, R. Schwarte, H. Heinol, B. Buxbaum, and T. Ringbeck. Smart pixel – photonic mixer device (PMD). In *Proc. Int. Conf. on Mechatron. & Machine Vision*, pages 259–264, 1998.