Model-free Bin-Picking: Food Processing and Parcel Processing Use Cases

Nicola Castaman, Alberto Cenzato, Stefano Tonello
IT+Robotics S.r.l., Italy
nicola.castaman@it-robotics.it

Emanuele Menegatti, Stefano Ghidoni, Alberto Pretto
Department of Information Engineering, *University of Padua*, Italy
alberto.pretto@dei.unipd.it

***Abstract*—Vision guided robots are enjoying growing success in industry and logistics, thanks to their adaptability to unstructured contexts and applications. The bin-picking problem is a prominent example in this trend. In typical bin-picking applications, a robot is guided to pick known rigid objects randomly placed inside a container. Given the objects' CAD models, it is possible to accurately estimate the object pose and to perform the grasp synthesis in a closed form. Unfortunately, domains such as the food industry and package delivery require to manipulate polymorphic and deformable objects; moreover, valid CAD models are often not available. A direct bin-picking technology transfer is here infeasible, due to the variability in shapes, size, and appearance of the elements to be manipulated. In this abstract, we present the SPIWI and the EACHPack research projects: both have the goal to bridge the gap between the current and desired capabilities of vision guided robots, by developing new bin-picking methodologies and systems well suited to deal with food processing and package delivery products. We address the challenges related to the variability of shapes of the objects by solving the object pose estimation and grasp synthesis problems in an unified way, inside a state-of-the-art instance segmentation data-driven model. The former will be modified to explicitly deal with polymorphic shapes. The latter will be exploited both to provide an initial estimate of the position of goods, and to infer in advance some features (e.g., object weight), useful for efficient object placing. The proposed systems will be easily adaptable to a wide range of applications, thus greatly improving their potential impact.**

***Index Terms*—Bin-Picking, Industrial Robotics, Object Pose Estimation, Instance Segmentation**

I. Introduction

The ability to locate and manipulate objects randomly placed inside containers and accurately insert them into production processes is an enabling technology that is becoming increasingly popular in many domains. This problem, often referred to as *random bin picking*, rely on robust 3D pose estimation algorithms that exploit either 2D or 3D vision technologies [1], [6], with an increasingly trend towards data-driven approcehs based on deep models [3], [5]. Most of these systems assume to deal with one or more objects of known and non-variable shape. This assumption limits the applicability of these technologies in new, very promising domains. In this abstract we present two research projects started to address these challenges: the SPIWI and EACHPack projects. Both projects rely on a shared backbone based on a robust data-driven pose estimation module able to localize and handle deformable and non-fixed size objects.

This work was supported by the EC under Grant H2020-780265-ESMERA (SPIWI project) and Grant H2020-825196-TRINITY (EACHPack project) and by MIUR (Italian Minister for Education) under the initiative "Departments of Excellence" (Law 232/2016).

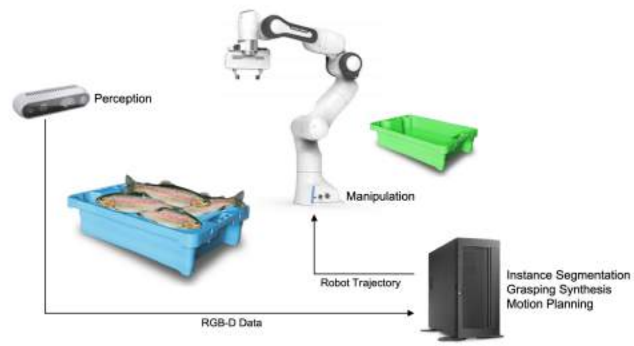

Fig. 1: SPIWI project concept: a robot accurately transfers fresh fish from large bins to small trays, ensuring with a predictive weighing procedure that the total weight of each tray is kept within specified limits.

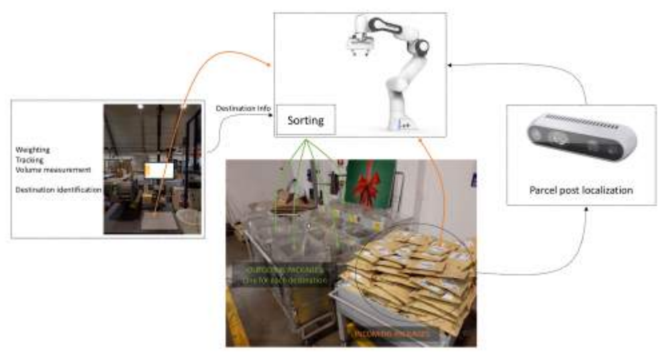

Fig. 2: EACHPack project concept: a robotized handling system for parcel posts able to pick each parcel post from a bin, weigh it, measure its volume, and sort it according to the destination in different outgoing containers.

A. The SPIWI Project

Primary food processing is an important example where industrial robots are increasingly exploited to improve cleanliness, safety, and efficiency. An emerging application in this context is represented by the vision-guided pick and place of fruits, vegetables, seafood, and beef. Unfortunately, these products are difficult to perceive and manipulate with a robot, due to their variable size and shapes. They also require

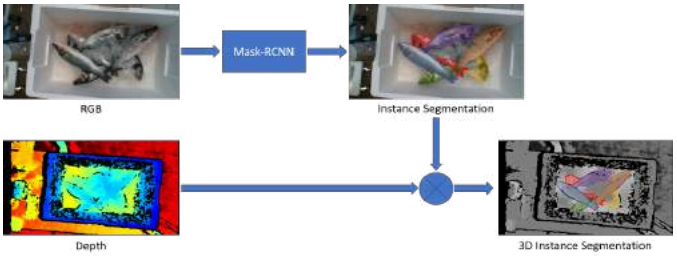

Fig. 3: The proposed object pose estimation pipeline.

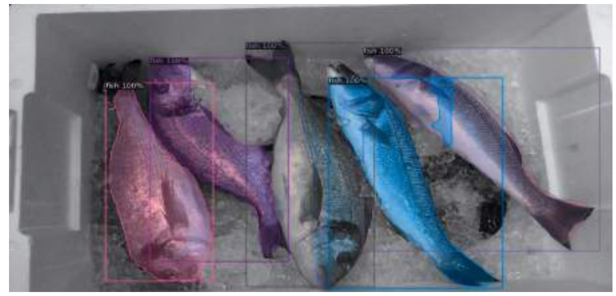

Fig. 4: An example of instance segmentation.

gentle manipulation to avoid damage. The SPIWI research project, acronym for “Smart Fish Picking Enhanced with Weight Inference”, aims to tackle some of these challenges by proposing a system capable to robustly and automatically transfer fresh fish from large trays to smaller trays ensuring that the total weight of the packed second tray is kept within specified limits (Fig. 1). The task involves localizing the suitable fish to pick, possibly selecting the one with the weight more convenient to correctly fill the output tray, and achieving a stable grasp that does not damage the fish.

B. The EACHPack Project

The growth of e-commerce is leading to the explosion of shipments of small postal packages weighing less than 5kg. Small parcel sorting is still a manual process that requires to pick up each package from a container, put it under a tracking system that checks the package transit, measures weight and volume, and places them on specific outgoing containers according to its destination. Parcels do not have a reference model and may have different shapes, sizes, and colors. The goal of the EACHPack project (acronym for “End-to-end Automatic Handling of Small Packages”, Fig. 2) is to completely automate the package tracking process, by picking each parcel post from a bin with a manipulator, weigh it, measure its volume, recognizing the shipping label (bar code or data matrix, which could be found anywhere in the package), and sort it according to the destination in different outgoing containers.

II. Object Pose Estimation

In both the SPIWI and EACHPack projects we use the same object pose estimation algorithm and a vision system based on an Intel Realsense D435 RGB-D camera. The algorithm exploits a model-less, data-driven approach using RGB-D data to identify both object instances (fish or parcel posts) and initial guesses of the grasping points. As depicted in Fig. 3, the first step consists in finding each instance in the image. To perform this task, we identified Mask-RCNN [2] as a suitable model-less algorithm. Mask-RCNN is a modular algorithm for instance segmentation that predicts segmentation masks on regions of interest in parallel with the classification and bounding box recognition. In order to improve the segmentation performance, we enabled the Mask-RCNN network to work with four-channel RGB-D images. We also exploited the capability of Mask-RCNN to infer keypoints positions by encoding suitable 2D x, y grasping points in the training data. In this way, the model provides for each segmented instance also a possible grasping point. The objects segmentation obtained with the RGB-D image (e.g., Fig. 4) is then used to select a corresponding point-cloud cluster for each instance. 3D-segmented data along with the inferred 2D grasping point are then processed in the manipulation phase to detect suitable 3D grasping points.

III. Self-Supervised Predictive Weighing

A further contribution we introduced in SPIWI is a self-supervised procedure that enables the system to infer in advance the weight of the objects. The total weight of the destination trays (green tray in Fig. 1) should be kept within specified limits. To reduce the number of re-grasps due to pick-ups of fish with a wrong weight, we implemented a visual weight estimator we call *predictive weighing*, implemented as a self-supervised data-driven learning module, constantly fine-tuned in production by using the load cell placed under the input tray (light blue tray in Fig. 1) to automatically label data. A convolutional neural network (CNN) trained for regression [4] takes as input the portion of the depth map that belongs to a fish and infers its weight. After each successful picking, the system can collect both the RGB-D data and the actual fish weight: we use these samples to constantly re-train the regression model.

References

[1] E. Brachmann, A. Krull, F. Michel, S. Gumhold, J. Shotton, and C. Rother, “Learning 6d object pose estimation using 3d object coordinates,” in *Computer Vision – ECCV 2014*, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds., 2014, pp. 536–551.

[2] K. He, G. Gkioxari, P. Dollár, and R. Girshick, “Mask r-cnn,” in *Proceedings of the IEEE international Conference on Computer Vision (ICCV)*, 2017, pp. 2961–2969.

[3] W. Kehl, F. Milletari, F. Tombari, S. Ilic, and N. Navab, “Deep learning of local rgb-d patches for 3d object detection and 6d pose estimation,” in *Computer Vision – ECCV 2016*, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., 2016, pp. 205–220.

[4] S. Lathuilière, P. Mesejo, X. Alameda-Pineda, and R. Horaud, “A comprehensive analysis of deep regression,” *CoRR*, vol. abs/1803.08450, 2018.

[5] S. Peng, Y. Liu, Q. Huang, X. Zhou, and H. Bao, “Pvnet: Pixel-wise voting network for 6dof pose estimation,” in *CVPR*, 2019.

[6] A. Pretto, S. Tonello, and E. Menegatti, “Flexible 3d localization of planar objects for industrial bin-picking with monocamera vision system,” in *Proc. of: IEEE International Conference on Automation Science and Engineering (CASE)*, 2013, pp. 168 – 175.