Semantic Segmentation for Flexible and Autonomous Manufacturing

Matteo Terreran*, Stefano Ghidoni*

* Department of Information Engineering, University of Padova, Padova, Italy Email: matteo.terreran@dei.unipd.it, stefano.ghidoni@unipd.it

Abstract-Customized mass production of boats and other vehicles requires highly complex manufacturing processes that involve a high amount of automation. Key elements to enhance the efficiency of such systems are represented by vision and sensing, which provide robots with detailed information about the working environment. In this paper, we focus on the sanding process of boat molding tools by means of a robot, proposing the use of semantic segmentation to detect the key elements involved in production and increase the automation of the production process. We demonstrate the potential of semantic segmentation in an industrial environment which differs from the domestic scenes typically considered in the literature: it features a lower degree of variability with respect to domestic scenarios, but higher performances are required in the production environment to address challenging manufacturing operations successfully. Our segmentation algorithm has been thoroughly validated on a industrial dataset that was created on purpose, whose acquisition and annotation were speeded up thanks to our optimized pipeline.

Index Terms—scene understanding, semantic segmentation, machine learning

I. INTRODUCTION

The manufacturing process of boats, as well as aircrafts and automobiles, requires the production and assembly of several components, with extensive use of light and strong materials like Glass Fiber Reinforced Polymers (GFRPs). Such composite materials are common choices being the perfect trade-off considering all these aspects and given that they can also be easily molded into complex shapes. The mass production of GFRP components comprehends the design and manufacturing of a molding tool for each component. The surfaces of the molding tools must be perfect, since every imperfection in the molding tools affects the final component. The whole process to smooth and polish the surface include many sanding operations which represent more than half of the total manufacturing time and despite being heavy tasks, they are generally performed by human workers.

Under this perspective, one of the aims of the EU project COROMA¹ (COgnitively enhanced Robot for flexible MAnufacturing of metal and composites) is the robotic automation of the sanding procedure in the production of boats to improve the quality of work and achieve better efficiency and performance in the production phase. One of the key contributions of the project is to use vision to make the robot fully autonomous and capable of navigating across the production



Fig. 1: Example of the industrial scenario considered in the COROMA project, including a boat hull and the sanding robot.

environment: once a boat to be sanded is recognized, the robot moves towards the boat and begin the sanding procedure as depicted in Figure 1. To achieve this goal, the robot needs to analyze and understand the environment, decomposing the scene into its most meaningful parts, like the boat components or structural elements (e.g. floor and walls), producing a pointwise labeling of the scene [1].

In the robotics community such problem is known as semantic segmentation: given an input image, the objective is to assign to each pixel in the image a label representing the class of interest which the pixel belong. Semantic segmentation is often tackled using supervised machine learning, where an algorithm learns to map an input to an output based on example input-output pairs. This requires huge datasets of pixel-wise annotated frames where each pixel is associated to a class label. Unfortunately, popular datasets for semantic segmentation are collected in indoor scenario like office and apartments [2] and do not contain the classes of interest for the particular industrial application (e.g. the boat components). This highlights the need to acquire a dataset specific for the industrial application, which is usually a time-consuming task especially regarding the annotation phase that is generally done manually for each single image in the dataset. To address this problem we proposed an optimized pipeline to quickly collect a dataset in an industrial setting.

¹https://www.coroma-project.eu



Fig. 2: Qualitative results of 3DEF on the acquired dataset for different views of the main object of interest (the boat). Colors represent the predicted class: boat hull (light blue), tech. area (blue), frame (yellow), wall (pink), floor (red) and clutter (green).

II. METHODS

Among the supervised machine learning methods, deep learning networks represent state-of-the-art methods to solve the semantic segmentation task on many popular outdoor and indoor datasets [2]. However, deep learning solutions required huge amount of pixel-wise annotated data to be trained and need to be implemented on high-end GPUs in order to achieve real-time performances. This is impractical in many mobile robotics applications where robots cannot load high performance and high-power GPUs or large batteries, hence limiting the available energy. For these reasons, in our work we avoid the use of deep learning and focus on a machine learning algorithm, named 3D Entangled Forest (3DEF) [3], which runs on CPU and it is simpler to integrate on a mobile robot. Moreover, we propose an optimized pipeline to quickly acquire and annotate a big dataset to train the algorithm in specific scenarios with object classes not included in popular existing datasets. For data acquisition we rely on a mobile robot equipped with an RGB-D sensor (i.e. Kinect One V2) which provides RGB and depth images. The mobile robot is teleoperated by means of a joystick and moved around the environment while recording the data from the RGB-D sensor. Since the data are acquired in a contiguous manner in the same environment, each single-view frame contains the same objects only from a different point of view. Instead of annotating each frame individually, we decide to annotate a 3D reconstruction of the environment and then projecting back the annotations to the original frames; this allow to speed-up the annotation phase since each single object is labeled only once. In particular, a 3D reconstruction of the scene is computed from the RGB-D data using ElasticFusion [4], a Simultaneous Localization and Mapping (SLAM) algorithm. The 3D scene is then annotated using SceneNN annotation tool [5] according to the classes of interest: the boat component is annotated with three different colors in order to differentiate among the proper boat, the technical areas below, and the supporting frame. Other classes we considered are wall, floor and clutter. The clutter class wraps lots of different elements, like machines, robots, boxes and tools. Finally, the pixel-wise annotation of each singleview frame can be retrieved automatically by re-projecting the annotation from the 3D scene to its point-of-view.

III. RESULTS

We tested our algorithm on a real industrial scenario. First, a task-specific dataset is collected and annotated with our optimized pipeline. The dataset is composed of several sequences, where each sequence corresponds to a single run of the mobile robot in the environment and differs from the others in the boat location inside the environment. The set of classes of interest is chosen in order to decompose the scene into its most interesting elements regarding the sanding operation: the boat components (boat hull and technical area), its support (frame) and the structural elements like floor and walls, while all the other objects are grouped in a heterogeneous clutter class. As shown in Figure 2, qualitative results are very good and the 3DEF classifier is able to correctly recognize the main elements of the scene. In conclusion, we present an application of semantic segmentation to detect the key elements involved in production and boost automation in the production process. In particular, we demonstrate the potential of the 3DEF classifier in an industrial environment featuring a lower degree of variability with respect to the domestic scenes typically considered in the literature but requiring higher performances to address challenging manufacturing operations successfully.

ACKNOWLEDGMENT

Part of this work was supported by MIUR (Italian Minister for Education) under the initiative "Departments of Excellence" (Law 232/2016).

REFERENCES

- M. Terreran, M. Antonello, and S. Ghidoni, "Boat hunting with semantic segmentation for flexible and autonomous manufacturing," in 2019 European Conference on Mobile Robots (ECMR), pp. 1–8, IEEE, 2019.
- [2] N. Silberman and R. Fergus, "Indoor scene segmentation using a structured light sensor," in 2011 IEEE international conference on computer vision workshops (ICCV workshops), pp. 601–608, IEEE, 2011.
 [3] D. Wolf, J. Prankl, and M. Vincze, "Enhancing semantic segmentation
- [3] D. Wolf, J. Prankl, and M. Vincze, "Enhancing semantic segmentation for robotics: The power of 3-d entangled forests," *IEEE Robotics and Automation Letters*, vol. 1, no. 1, pp. 49–56, 2015.
- [4] T. Whelan, S. Leutenegger, R. Salas-Moreno, B. Glocker, and A. Davison, "Elasticfusion: Dense slam without a pose graph," Robotics: Science and Systems, 2015.
- [5] B.-S. Hua, Q.-H. Pham, D. T. Nguyen, M.-K. Tran, L.-F. Yu, and S.-K. Yeung, "Scenenn: A scene meshes dataset with annotations," in 2016 Fourth International Conference on 3D Vision (3DV), pp. 92–101, IEEE, 2016.