Bei diesem Beitrag handelt es sich um einen wissenschaftlich begutachteten und freigegebenen Fachaufsatz („reviewed paper“).

doi.org/10.37544/1436-4980-2025-07-08-102

Applying computer vision and behavior trees to robotics

Status recognition for collaborative robotics

D. Kötter, F. Nolte, O. Petrovic, C. Brecher

ABSTRACT Collaborative robotics tries to combine the strengths of humans and robots. This article discusses incorporating object detection into a behavior tree (BT), a common hierarchical control structure in robotics. This integration allows robots to verify assembly status and provide feedback to workers during tasks such as gear assembly. Also, a pipeline for creating synthetic training datasets from real parts is introduced, demonstrating effective status recognition and feedback mechanisms.

Statuserkennung in der kollaborativen Robotik - Anwendung von Computer Vision und Behavior Trees in der Robotik

ZUSAMMENFASSUNG Die kollaborative Robotik versucht, die Stärken von Menschen und Robotern zu kombinieren. In diesem Beitrag geht es um die Integration der Objekterkennung in einen Behavior Tree (BT), eine in der Robotik übliche hierarchische Kontrollstruktur. Diese Integration ermöglicht es Robotern, den Montagestatus zu überprüfen und den Arbeitern bei Aufgaben wie der Getriebemontage Feedback zu geben. Darüber hinaus wird eine Pipeline zur Erstellung synthetischer Trainingsdatensätze aus realen Teilen vorgestellt, die eine effektive Statuserkennung und Feedbackmechanismen demonstriert.

KEYWORDS

Collaborative Robotics, status recognition, assembly

STICHWÖRTER

Kollaborative Robotik, Statuserkennung, Montage

1 Introduction

Recent trends in globalization have increased competitive pressure [1]. Moreover, consumers demand products that can be customized to their specific needs [2]. The assembly process, as the last step in production, is particularly affected by fluctuations in demand due to its market proximity. To master these challenges, processes need to be optimized with a view to efficiency and agility. Automating previously manual processes by machines or robots reduces labor costs and ensures a high level of quality [3]. Unlike assembly machines, robots are characterized by increased flexibility and lower investment costs. Robots can take on simple assembly tasks, which are repetitive or physically strenuous for human workers [4–6]. However, some tasks still require high-level reasoning and dexterity, which only humans can provide. Human-robot collaboration (HRC) aims to combine the strengths of automated and manual assembly. Expert knowledge in automation is needed to plan collaborative assembly processes. An expert must program the robots and ensure safe collaboration with humans. Most small and medium-sized businesses (SMB) lack this expertise. Simplifying the planning process makes the advantages of automation accessible to SMB, strengthening their resilience towards global competition.

A promising approach is to use behavior trees (BTs) as a high-level control structure for controlling robots [7]. BTs organize the behavior of a system in a hierarchical tree structure consisting of internal control flow nodes and leaves. BTs are modular since nodes can be easily added or removed. Condition nodes check if a specific condition is met before an action node is executed, which leads to the reactivity of the trees. BTs can be extended to distributing subtasks between humans and robots, or to automatically determining an optimum of order subtasks. Also, BTs are intuitive and human-understandable, and can therefore be used by non-experts. [8]

To optimize human-robot collaboration, this paper extends a BT by a node for a status recognition system (SRS) to recognize the assembly process, using convolutional neural networks. This node recognizes the actual status of the assembly process (e.g., “fastening screws”, “assembly shaft”), controls the robot depending on status, and, accordingly, allows the robot to work more independently. A dataset consisting of assembly parts is created to train the SRS.

The paper is structured as follows. Chapter 2 gives an overview on state-of-the-art approaches towards SRS for collaborative robotics. Chapter 3 presents the system design of status recognition embedded in BTs. Chapter 4 evaluates the status recognition

Table. Approaches towards a status recognition system in collaborative robotics.

Source	Model	Hands	Parts	Dataset	Summary
[9]	YOLOv5	No	Yes	Yes	Status recognition of a mold assembly by detecting parts, tools and actions
[10]	YOLOv3, Open Pose	Yes	Yes	Yes	Using Markov chains to recognize the assembly motion corresponding to a specific action
[11]	YOLOv3	Yes	No	Yes	Classification of different hand gestures
[12]	Kinect V2	Yes	No	No	Robot control by means of gestures and voice commands
[13]	YOLOv3	No	Yes	No	Object matching of assembly parts on a base plate
[14]	ANN, LSTM	Yes	No	Yes	Gesture-based human-robot interaction in manufacturing

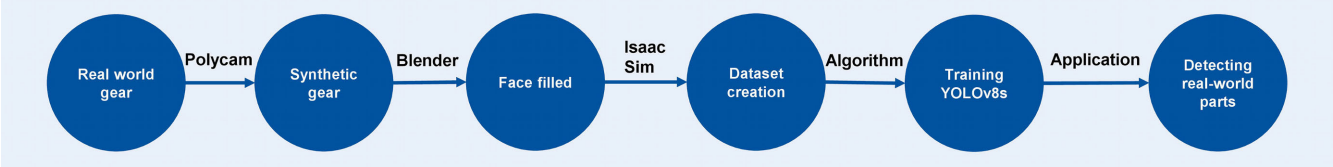

Fig. 1. Pipeline for detecting different parts of a gear in the real world. *Source: WZL, RWTH Aachen*

system and discusses the results. Chapter 5 provides a summary and outlook for future lines of investigation.

2 Related work

There is a range of different SRS varying with respect to the model used for detecting the assembled parts (and the kind of parts) as well as the hands of the worker, and also whether a dedicated dataset was created for the specific use case. [9] apply a YOLOv5 algorithm to create a status recognition system that can be applied to mold assembly. They decompose the assembly process into tasks and sub-tasks and define the actions to represent the status of sub-tasks. [10] propose a method for recognizing and segmenting assembly tasks into individual motions. They use a motion capture system with pose estimation to collect time-series motion data and an object detector to identify grasped parts and tools. The assembly motion is then segmented based on changes in the manipulated object and hand velocity, using Hidden Markov Models to recognize these segmented motions.

[11] introduces a lightweight gesture recognition model based on YOLOv3 and DarkNet-53. It is designed for people with disabilities to enable them to communicate more easily. The model does not require extra preprocessing steps and achieves high accuracy, even with complex environments and low-resolution images. [12] explores the Kinect v2 for industrial cobot control via gestures and voice commands. This software separates robot movement and communication into two independent threads, preventing interference. [13] develops a method using deep learning and object matching to detect missing and incorrect parts. The method uses an improved YOLOv3 network with smaller target detection scale and attention module, optimizing the prior anchor box size using the K-means++ algorithm. By matching 2D detection boxes with a standard assembly template, the method accurately locates assembly parts and identifies missing or wrong components, as validated by tests with the model.

[14] proposes a gesture-based human-robot interface framework where a robot assists a human coworker by delivering tools or parts and holding objects, utilizing wearable sensors to capture upper body gestures, which then are classified using an artificial neural network. The system uses a parameterized robotic task manager where coworkers use gestures to select or validate robot options based on speech and visual feedback, demonstrating efficiency in assembly operations. The **Table** below summarizes the approaches towards an SRS in collaborative robotics.

Since none of the approaches focuses on the reusability of the proposed system, this work utilizes behavior trees to ensure an easy integration and adaptation of the proposed SRS into different robotic control systems. Without the underlying control architecture being reusable, a lot of manual effort is required to integrate the approach into existing systems. Furthermore, the SRS should be able to recognize different parts of the process and provide feedback to workers in case the wrong part is moved into the workspace of the cobot.

3 Status recognition system design

3.1 Object detection

To detect different parts from a gear in the cobot's workspace, two IntelReal Sense D435i [15] cameras are used alongside the YOLOv8s model for object detection (see **Figure 1**).

Since synthetic gear data are not available, we use Polycam [16] to capture the 3D model of the object. Blender [17] is used to face fill, and Isaac Sim [18] to create and annotate a dataset of 100 images with bounding boxes for each part of the gear, which totals 4400 images. The algorithm is trained and evaluated with an 80/20 split and reaches a mean average precision of 91.82 % at around 2.2 frames per second, running on an AMD Radeon RX Vega10 graphics card. The two cameras publish their image stream to the respective topic */cam_*1 and */cam_*2, which are subscribed by the detection nodes 1 and 2, each running the trained YOLO algorithm, publishing the results to */YoloCoords_*1 and */YoloCoords_*2. The results are merged and published to the */AssemblyStatus.*

3.2 Behavior tree

Figure 2 shows the BT for the SRS. *Trinh et al* [19] provide more information about the safety subtree.

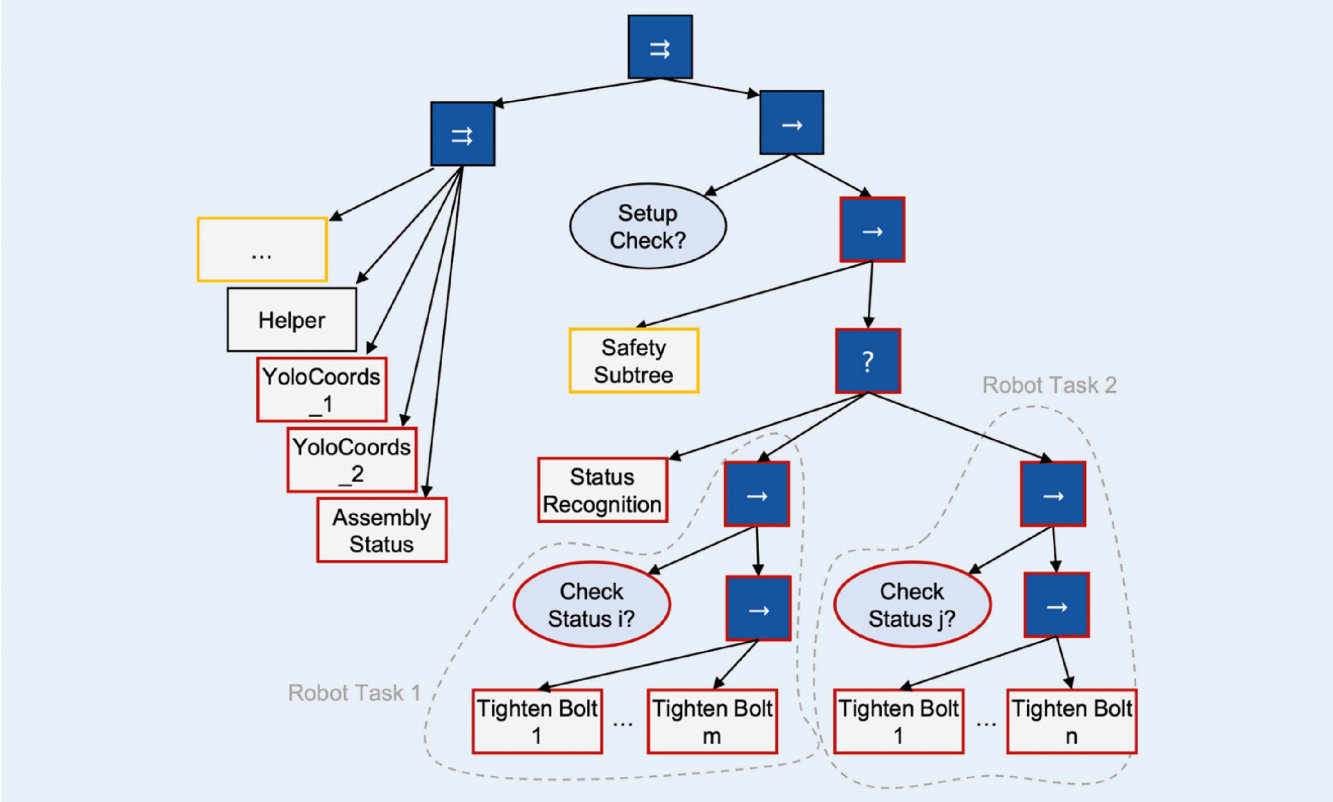

Fig. 2. Behavior Tree with integrated status recognition system (shown in red). *Source: WZL, RWTH Aachen*

If the safety subtree detects that the distance between human and robot is bigger than a given threshold, the status recognition subtree is ticked. Herein, both camera-streams are evaluated to detect parts of the gear, and the results are written on the blackboard as /*YoloCoords*_1 and /*YoloCoords*_2. Then, the BT checks if the assembly is in a certain state, triggering the respective cobot action associated with the state. If the worker moves the wrong part of the gear into the workspace of the cobot (e.g., the worker moves the cover into the workspace, although the assembly requires the circuit board as next part), the system provides feedback to the worker, via monitor, that a wrong part has been moved into the workspace. Additionally, the system gives feedback on which part it expects instead of the wrong part (see Figure 3).

4 Evaluation

Figure 3 shows the image stream of one of the cameras and the feedback from the SRS to the worker.

The camera is mounted 400 mm above the table that is used for assembling. As a use case for assembly, a gear for an electric bicycle is taken. Assembling the gear requires ten different steps:

1. Positioning the base frame
2. Positioning the gears
3. Pressing on the metal cover
4. Overlaying the rubber cover
5. Inserting the circuit board
6. Inserting bolts
7. Tightening bolts
8. Placing the cover
9. Inserting bolts
10. Tightening bolts

Except for the process steps of tightening bolts (as the camera's bird's-eye view does not allow for recognizing if the bolts are tightened or not – which would even be impossible for humans from that position), the system is able to update the assembly status, from the initial step of positioning the base frame to the last step, i. e., inserting the bolts.

If a wrong part is moved into the cobot's workspace, the system gives feedback via a monitor to the worker and shows what part the system expects instead of the given part. If the right part is moved into the workspace, the status of the assembly updates by +1 (see Figure 3). Using an AMD Radeon RX Vega10 graphics card, the frame rate of the evaluated images is 0.45 frames per second.

5 Conclusion

This paper develops a status recognition system for collaborative assembly processes. As a use case, a gear from Bosch is assembled. Polycam, alongside Blender and Isaac Sim, is used for creating a dataset to train the YOLOv8s object detection algorithm. The detection model is embedded in the BT of the cobot, enhancing the reusability of the proposed system. The evaluation shows that the logic of the SRS within the BT works well. Nevertheless, the respective cobot actions need to be implemented to deploy the system in the industry.

In the future, the authors want to program the respective cobot action. Furthermore, they plan to focus on detecting the assembled part instead of single parts. Additionally, they want to evaluate the proposed system with different cobots to demonstrate that the proposed system can be generalized. Furthermore,

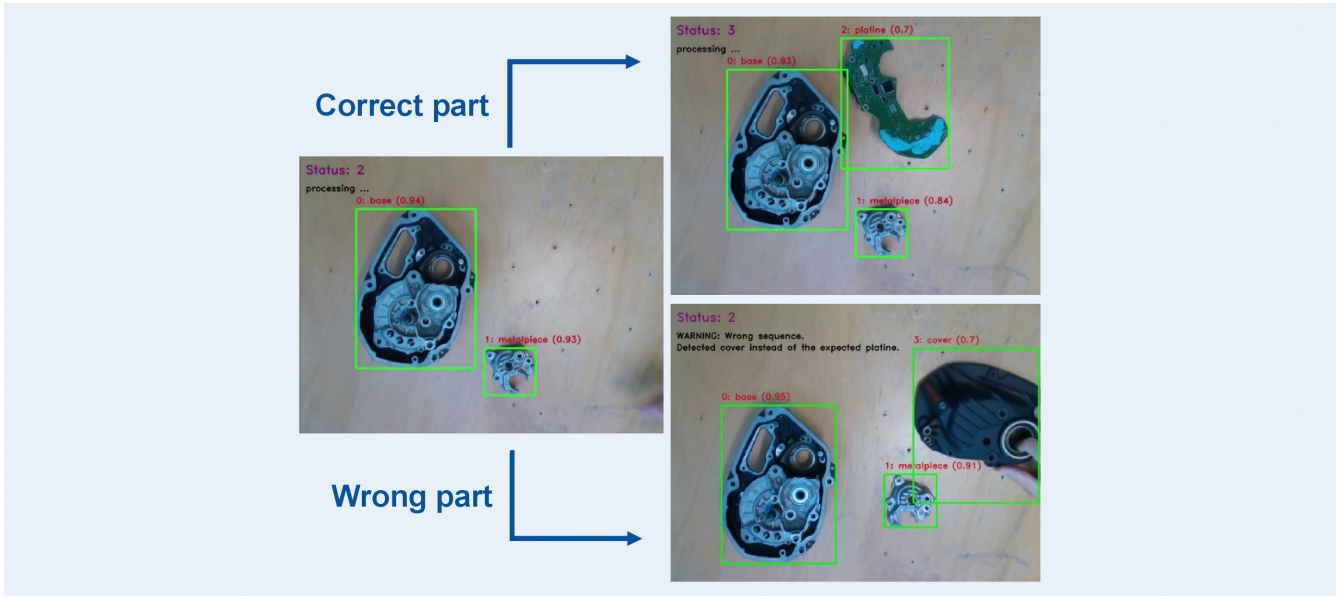

Fig. 3. Evaluation of the status recognition system. *Source: WZL, RWTH Aachen*

the evaluation should be carried out with a more powerful graphics card, improving the number of evaluated frames per second.

ACKNOWLEDGEMENTS

Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC-2023 Internet of Production – 390621612 and funded as IGF-project 22648 N/2 (ROOKIE) of the research association FVP via AiF within the funding program "Industrielle Gemeinschaftsforschung und -entwicklung (IGF)" by the Federal Ministry for Economic Affairs and Climate Action (BMWK), due to a decision of the German Parliament.

REFERENCES

[1] Bang, K.; Markeset, T.: Impact of globalization on model of competition and companies' competitive situation. In: Frick, J.; Laugen, B. T. (eds.): Advances in Production Management Systems. Value Networks: Innovation, Technologies, and Management. APMS 2011. IFIP AICT 384. Heidelberg. Springer-Verlag 2012, pp. 276–286

[2] Pallant, J.; Sands, S; Karpen, I.: Product customization: a profile of consumer demand. Journal of Retailing and Consumer Services 54 (2020), #102030

[3] Xiang, J.; Kong, D.; Zhang, F.: Labor cost, robots, and product quality. China Economic Review Volume 90, (2025), #102373

[4] Sun, Q.; Gao, D.; Chen, X.; Sun, J.; Huang, N.; Gao, F.: A collaborative robotic screw assembly system using 6-PUS parallel mechanism. ASME 2023 International Design Engineering Technical Conferences and Computers and Information in Engineering (2023) pp. 179–186, doi.org/10.1115/DETC2023–114479

[5] Liau, Y.; Ryu, K.: Task allocation in human-robot collaboration (HRC) based on task characteristics and agent capability for mold assembly. Procedia Manufacturing 51 (2020), pp. 179–186

[6] Kötter, D.; Wiedon, G.; Meierkord, D. et al.: Development of an augmented reality user interface for collaborative robotics in quality inspection for manufacturing. 5th International Conference on Control and Robotics (ICCR), 2023, pp. 107–112

[7] Ögren, P.; Sprague, C.: Behavior Trees in Robot Control Systems. Annual Review of Control, Robotics, and Autonomous Systems (2022), pp. 81–107, doi.org/10.48550/arXiv.2203.13083

[8] Colledanchise, M.; Ögren, P.: Behavior Trees in Robotics and AI: An Introduction. CRC Press, doi.org/10.48550/arXiv.1709.00084

[9] Liau, Y.; Ryu, K.: Status Recognition Using Pre-Trained YOLOv5 for Sustainable Human-Robot Collaboration (HRC) System in Mold Assembly. Sustainability 21 (2021) 13, #12044

[10] Fukuda K. et al. Assembly Motion Recognition Framework Using Only Images. IEEE/SICE International Symposium on System Integration (SII) (2020), pp. 1242–1247

[11] Mujahid, A. et al.: Real-Time Hand Gesture Recognition Based on Deep Learning YOLOv3 Model. Applied Sciences 11 (2021) 9, #4164

[12] Kaczmarek, W. et al.: Industrial Robot Control by Means of Gestures and Voice Commands in Off-Line and On-Line Mode. Sensors 20 (2020) 21, #6358

[13] Zhao, S. et al.: Assembly state detection based on deep learning and object matching. IEEE 18th International Conference on Automation Science and Engineering (CASE) (2022), pp. 1695–1700

[14] Neto, P. et al.: Gesture-based human-robot interaction for human assistance in manufacturing. The International Journal on Advanced Manufacturing Technology 101 (2019) 1–4, pp. 119–135

[15] Intel RealSense: D435i camera. Product description. Date: 2024. Internet: www.intelrealsense.com/depth-camera-d435i/. Accessed: 22.07.2025

[16] Polycam Learn: Homepage. Date: 2024. Internet: poly.cam/. Accessed: 22.07.2025

[17] Blender: Software description. Date: 2024. Internet: www.blender.org/about/. Accessed: 22.07.2025

[18] NVIDIA Isaac Sim. Framework description. Date: 2024. Internet: developer.nvidia.com/isaac/sim . Accessed: 22.07.2025

[19] Trinh, M. et al.: Safe and Flexible Planning of Collaborative Assembly Processes Using Behavior Trees and Computer Vision. Intelligent Human Systems Integration (IHSI) 69 (2023), pp. 869–879

David Kötter, M.Sc.
d.koetter@wzl.rwth-aachen.de

Fabian Nolte, M.Sc.

Oliver Petrovic, M.Sc.

Prof. Dr.-Ing. Christian Brecher

Laboratory for Machine Tools and Production Engineering WZL, RWTH Aachen University
Steinbachstr. 25, 52074 Aachen
www.wzl.rwth-aachen.de