3rd Conference on Production Systems and Logistics

Augmented Virtuality Data Annotation And Human-In-The-Loop Refinement For RGBD Data In Industrial Bin-Picking Scenarios

Andreas Blank[1], Lukas Baier[1], Maximilian Zwingel[1], Jörg Franke[1]
[1]*Institute for Factory Automation and Production Systems, FAU Erlangen-Nürnberg, Germany*

Abstract

Beyond conventional automated tasks, autonomous robot capabilities aside to human cognitive skills are gaining importance. This comprises goods commissioning and material supply in intralogistics as well as material feeding and assembly operations in production. Deep learning-based computer vision is considered as enabler for autonomy. Currently, the effort to generate specific datasets is challenging. Adaptation of new components often also results in downtimes. The objective of this paper is to propose an augmented virtuality (AV) based RGBD data annotation and refinement method. The approach reduces required effort in initial dataset generation to enable prior system commissioning and enables dataset quality improvement up to operational readiness during ramp-up. In addition, remote fault intervention through a teleoperation interface is provided to increase operational system availability. Several components within a real-world experimental bin-picking setup serve for evaluation. The results are quantified by comparison to established annotation methods and through known evaluation metrics for pose estimation in bin-picking scenarios. The results enable to derive accurate and more time-efficient data annotation for different algorithms. The AV approach shows a noticeable reduction in required effort and timespan for annotation as well as dataset refinement.

Keywords

Data Annotation; Augmented Virtuality; Human-in-the-Loop; Machine Learning; Bin-Picking

1. Introduction

Short product life cycles, an increasing amount of product variants and more complex goods pose challenges to the manufacturing industry. Flexible automation, involving robot systems, contribute to improve the situation. However, conventional automation reaches limitations in scenarios with uncertainties, including industrial bin-picking for material supply, machine feeding and assembly. Deep learning (DL) is an enabler for autonomous robot capabilities able to cope with such complex tasks [1]. Yet, the required dataset generation is time-consuming and thus costly [2]. In addition, downtimes may occur on system ramp-up [3].

In this context, the contribution of this paper is an augmented virtuality (AV)-based real-world RGBD data annotation and human-in-the-loop (HuITL) dataset refinement method. Objectives of the method are to reduce the effort and time spent in initial dataset generation for industrial bin picking and tuning the dataset up to operational readiness during ramp-up phase. In a single operation, both data for object classification and localization as well as data for 6DoF pose estimation are annotated. When in online refinement mode, in addition, the cognitive skills of the human remote operator are exploited to provide fault intervention and proper task solution for autonomous handling from distance by a teleoperation interface. The method enables accurate and more time-efficient RGBD data annotation and refinement. Thereby a noticeable reduction in required effort for successful application deployment and adaptation in industrial bin picking is achieved.

DOI: https://doi.org/10.15488/12184

2. Related Work: Data Generation and Human-In-The-Loop in Industrial Bin-Picking

In this section, progresses made in data generation for DL-based object recognition and regarding HuITL approaches for improving autonomous robot skills within industrial bin-picking applications are reviewed.

2.1 Data Generation for Object Recognition

The classification of algorithms can be done by different criteria, such as the image processing or input data. This includes object classification, localization, segmentation and pose estimation as well as solving combinations of these. As input, 2D-RGB, depth data or both, as well derived point clouds are common [4].

Although DL proves to outperform traditional algorithms, efforts required for specific dataset generation and training parametrization are still high [2]. A large quantity of data is required to improve the performance as well as to reduce the risk of overfitting. Data itself proves to be both the constraining and the driving factor. For data generation multiple techniques exist (cf. Figure 1). Depending on the type of input data, annotation is performed by 2D- and 3D-bounding boxes, 6DoF pose specification or pixel-wise labeling.

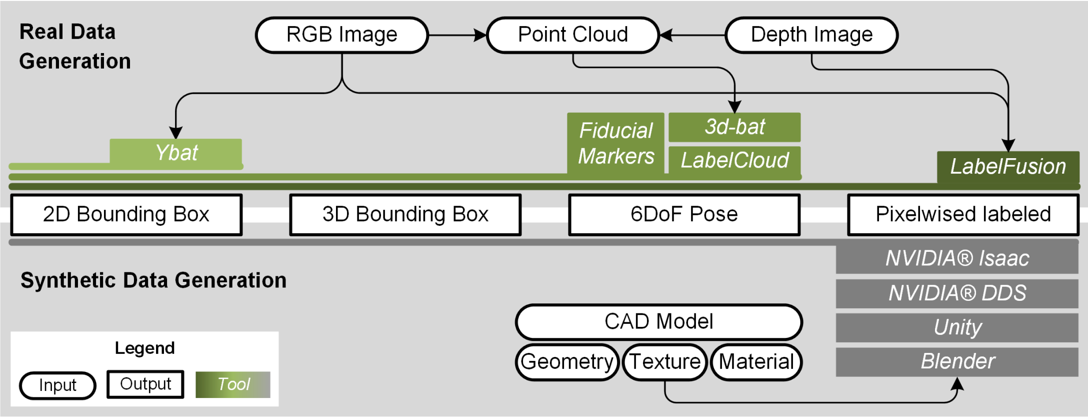


Figure 1: Data annotation types (center), procedures for annotating real-world data with common tools (top) as well as procedures and tools for automated generation of synthetic data (bottom)

Visual fiducial markers are state of art to determine ground truth pose for real-world objects [5]. Utilizing markers, however, is cumbersome and time-consuming. Components difficult to annotate exist due to geometric complexity. In addition, occlusions in multi-object scenarios complicate marker application.

For annotating objects based on recorded sensor data, software such as YOLO BBox Annotation Tool (Ybat), 3d-bat [6] and labelCloud [7] are important tools. Further available point cloud labeling tools focus mainly on autonomous driving [8]. Another solution is LabelFusion [9], a pipeline utilizing recorded RGBD video-data to produce pixel-wise object masks and 6DoF object pose labels. For software tools it is common to perform labeling utilizing computer monitor based graphical user interfaces (GUI). Thereby, components are iteratively aligned through definition of distinctive object related points and subsequent object boundaries are manually defined. For LabelFusion, in addition, a preceding 3D reconstruction step is required. Current solutions for labeling are time-consuming and common GUIs lack in spatial scene representation, especially complicating 6DoF pose annotation. In contrast, VR shows potential for efficient labeling. While VR solutions exist for semantic segmentation of large landscapes [10], approaches for appropriate labeling of data required for accurate robotic grasping operations are not available.

Synthetic data represent an alternative requiring less effort [11]. Thereby, an increased variance of 6DoF poses and camera perspectives as well as control of the image rendering are achievable. Digital component models involving texture and material specifications as well as current R&D-results on generative adversarial networks, domain adaption and domain randomization enable a more realistic dataset generation. However, closing the domain gap between synthetic and real-world data is still challenging [12].

2.2 Human-in-the-Loop Intervention and Dataset Refinement

Although datasets are generated component-specific, fault incidents of autonomous robots still occur. Since a machine operator may not always be nearby or on-site fault clearance may be harmful, effective remote intervention solutions are necessary. Current research focusses on teleoperated intervention [3]. However, relying solely on remote manipulation does not enable system adaptation. This is a relevant aspect, since root causes of failures often recur as well as operator time is expensive. Once a model is trained, human cognitive skills remain valuable to interpret and review model predictions as well as to enable dataset refinement [13,14]. Yu et al. show the potential of iterative refinement, especially in a more application-specific manner [15]. However, in industrial bin-picking exist a lack of HuITL dataset refinement methods.

3. Augmented Virtuality Data Annotation and Human-in-the-Loop Refinement and Intervention

Following the AV based real-world data annotation and HuITL refinement and intervention are described. First, an overview on architectural level is given (cf. Figure 2), followed by a description of the method.

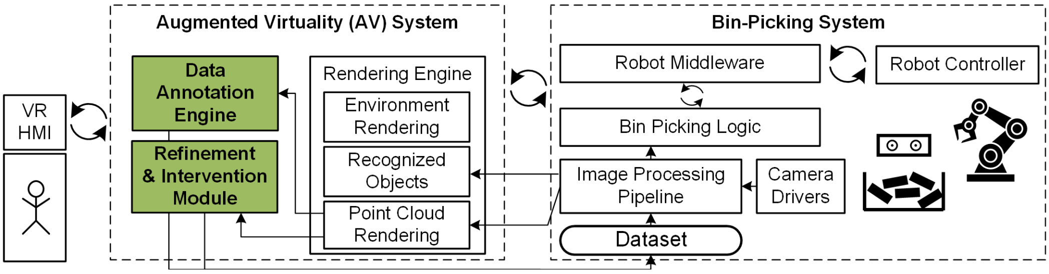


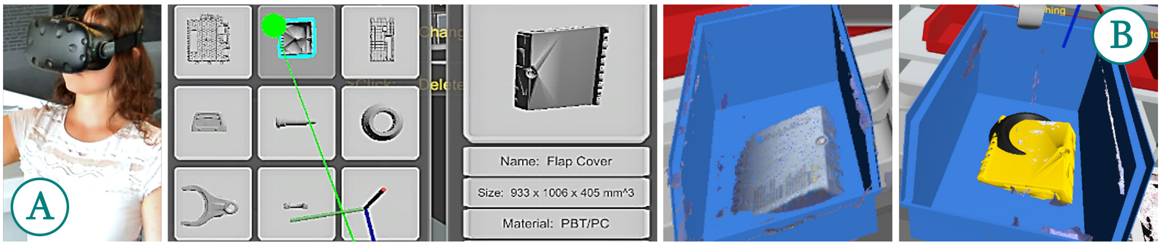
Figure 2: System overview as schematic technical architecture modelling (TAM)

3.1 Overview: Schematic Technical Architecture Modelling (TAM)

The AV rendering engine (cf. Figure 2) is described in previous work [16,17]. It involves rendering of the known environment, recognized objects as well as rendering of a pre-segmented point cloud. The illustrated image processing pipeline serves both systems for component classification, localization and 6DoF pose estimation. The exemplary utilized pipeline within this work is described in [4]. A robot control middleware is required to automate parts of the data annotation process (e. g. by changing camera perspective) as well as for HuITL intervention. The utilized middleware based on the Robot Operating System (ROS) is available open-source and is described in previous work [18]. Regarding the bin-picking system, the method requires access to the DL-dataset. At least weights should be interchangeable. To achieve a modular bin-picking system, a skill based logic using ROS actions (e. g. for motion control) is implemented [19]. Thereby, triggering and composition of skills as orchestration to a bin picking task is performed by a state machine.

3.2 Data Annotation Engine

The data annotation (Figure 2) is structured in: system parameterization and model selection, labeling and data generation (3.2.1) as well as automated data complementation (3.2.2). Although the descriptions address online data processing, the method is suitable for data generation based on once recorded offline sensor data.


Name: Flap Cover

Size: 933 x 1006 x 405 mm^3

Material: PBT/PC

Figure 3: AV Annotation: model selection (A) and labeling process (B) for data annotation

3.2.1 System Parameterization and Model Selection as well as Labeling and Data Generation

According to the bin-picking system, specifically the type of camera input data, the corresponding resolution and the type of desired output data (e. g. color images, point cloud, segmentation map, etc.) are determined. The camera transformation $\mathbf{T}_{\text{world}}^{\text{camera}}$ with respect to the world frame is calculated by the AV.

Within the annotation engine, new or not recognized components are rendered using their point cloud. The objective is to label these with matching models. For this purpose, the corresponding model is initialized (cf. Figure 3 (A)). The transformation $\mathbf{T}_{\text{world}}^{\text{object}}$ of components with respect to the AV world frame is determined within initialization step as well as the transformation $\mathbf{T}_{\text{camera}}^{\text{object}}$ is calculated. Thus, through visual alignment of the virtual component within the point cloud, the required ground truth is obtained (cf. Figure 3 (B)).

Since the coordinate system in rendering engines are based left-handed, whereas the camera coordinate system in robotics is often right-handed, transformations are applied. Finally, the labeling results are stored.

3.2.2 Automated Robot-Assisted Data Complementation

To further reduce effort in data generation, an optional automated robot-assisted data complementation is provided. This is applicable for setups with a robot-wrist mounted camera. Due to the previous labeling, the 6DoF poses of objects in space are known. Subsequently, the robot arm is remotely manipulated by the operator or automatically moved according to a pre-defined trajectory above and along the components for gaining further perspective object views. Thereby, new real-world data is annotated automatically.

3.3 Refinement and Intervention Module (HuITL)

Whereas AV annotation is employed for initial data generation, the refinement module (Figure 2) addresses system ramp-up enabling continuous dataset improvement. The method is structured in: interconnection and environment rendering (see 3.3.1), fault clearance (3.3.2) as well as annotation and refinement (3.3.3).

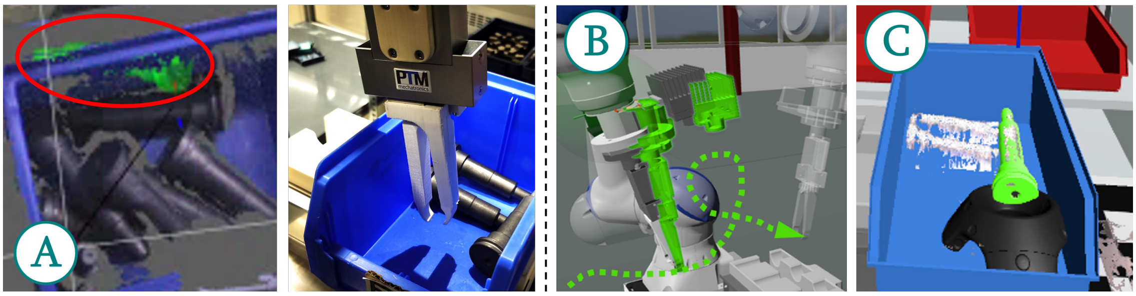


Figure 4: HuITL refinement and intervention: fault scenario caused by low confidence pose estimate (A), schematic illustration of teleoperated motion control (B) as well as hybrid component pose determination and annotation (C)

3.3.1 Interconnection and Environment Rendering

Once a fault is indicated by a linked system, an operator interconnects via the AV system. Subsequent, the robot environment involving system models and camera data is rendered. The fault condition itself might be triggered through a state-machine of the bin-picking system, caused by a repeatedly incorrect or low-confidence pose estimation as well as by multiple failed grasping attempts (cf. Figure 4 (A)).

3.3.2 Fault Clearance

The intervention module initially visualizes the last pose estimate (cf. Figure 4 (A)). In case a robot movement is required for solving (e. g. for component grasping or to realign camera, a component or the manipulator), the teleoperation interface is utilized (B). If the failure is caused by component localization or pose estimation, operators determine a reasonable solution (e. g. by target pose specification) based on the corresponding object model (C). This serves the bin-picking system as input for automated task completion.

3.3.3 Data Annotation and Network Refinement

The provided solution by the operator is stored as input for dataset refinement. The procedure is identical to AV annotation process. It differs in not generating an initial dataset, instead extending the existing dataset with additional more application-specific labels. The network model is updated based on the pre-trained model. Thereby, the latent network parameters are optimized towards the bin-picking application scenario.

4. Setup and Procedure of Experiments

The system setup, design of experiments as well as evaluation metrics are described in the following sections.

4.1 Demonstrator System Setup

To evaluate the method, a bin-picking testbed is implemented (cf. Figure 5 (A)). Visual perception is performed by a roboception rc_visard 65 stereo camera. The camera is mounted at the robot wrist. The camera delivers a depth image with a resolution of 640 × 480 pixels at a frame rate of 25 Hz. The highest available depth image resolution of 1280 × 960 pixels is not utilized due to lower frame rates. The camera depth deviation is specified with ±0.5 mm at 200 mm object distance and ±15 mm deviation at 1000 mm. Within the testbed a Yaskawa HC10 articulated robot is used. The system is based on ROS Melodic and Unity3D. ArUco markers are utilized for calibration between robot, camera and workspace.

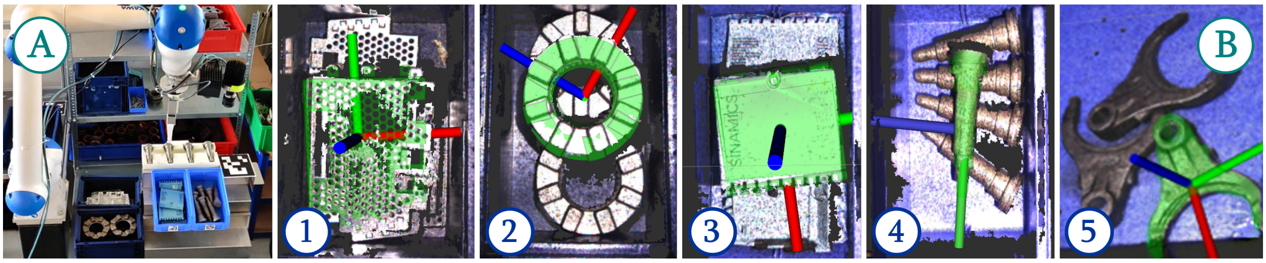


Figure 5: Bin-picking testbed (A) with evaluation components (B) – I/O shield (1), shifting sleeve (2), flap cover (3), shifting rod (4) and shifting fork (5); qualitative visual pose estimation results based on evaluation components (B).

The components under investigation are metallic and non-metallic parts of differing characteristics (i. a. material, shape, texture and surface) and will be referred to as: I/O shield (1), shifting sleeve (2), flap cover (3), shifting rod (4) and shifting fork (5).

Regarding materials, the flap cover (3) is made of polymer blend while the other parts (1,2,4,5) are metal. Geometries are a mixture of cylindrical and planar parts. Additionally, flap cover is matt and slightly textured while all metal parts are texture-less. I/O shield (1), in addition, challenges with a perforated surface.

4.2 Procedure of Experiments

Three experiments are performed for evaluation. First, the required timespan and the accuracy achieved when labeling are assessed. Here AV annotation is compared to the established AprilTag marker technique, utilized as standard for real-world data annotation in 6DoF pose estimation as well as been proven for high accuracy regarding annotation output [4,5]. This experiment serves to compare the performance of the proposed method in terms of accuracy and time effort to the state of the art. Further, to demonstrate feasibility and successful operability of the annotation outcomes provided by the proposed method, training results are tested by an image processing pipeline involving YOLOv3 [20] for object localization as well as Frustum PointNets [21] (FPN) for pose estimation within the described real-world bin picking testbed (Figure 5). Here, achieved object classification and localization success rates as well as 6DoF pose estimation accuracies obtained are assessed based on established evaluation metrics (cf. 4.3). Finally, a small user study is carried out to determine the Mean Time to Repair (MTTR) of the proposed teleoperation based fault intervention.

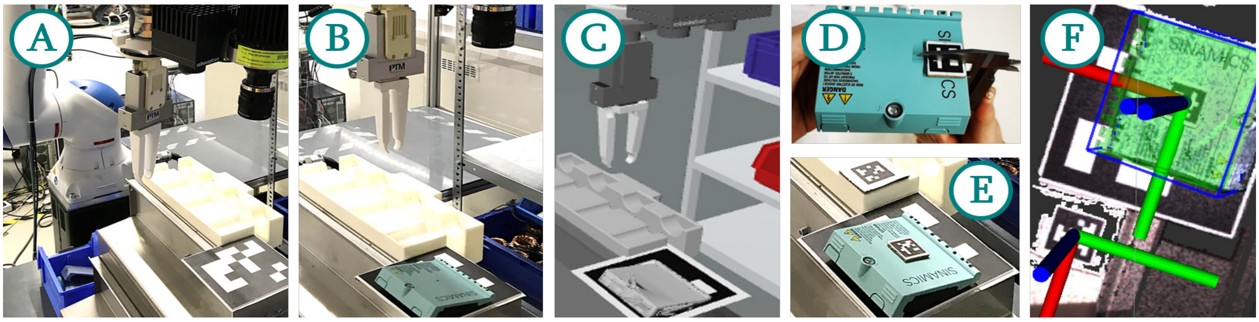


Figure 6: Ground truth determination (A) – (C) as well as technique used for preparing AprilTag comparison (D) – (F)

For accuracy evaluation, ground truth is required. First, ArUco markers serve for calibration (Figure 6 (A)). Afterwards, the test component is placed upon the ArUco marker and aligned (B). At the same time, the component is configured within a simulation (C). As a result digital and real-world components coincide. Finally, the ground truth 6DoF pose acquired is stored. Several steps are required for data annotation via AprilTag markers ((D)-(F)). First, for preparation the component must be measured to obtain center of gravity and surface middle (D). Subsequent, two markers are placed, one as reference within the scene and another in the component surface middle (E). Specifications obtained are stored for parametrization. Finally, the pose, the RGB image and the point cloud are recorded (F).

For comparison, AV-based annotation and AprilTag both are performed under consideration of the main influencing factors: namely the person conducting the annotation and the component to be annotated. For this purpose, a user study is carried out, involving five volunteers with annotation experience. Thereby, the five components ((1)-(5)) are labelled. As resulting measures, the timespan required for preparing and conducting labeling is recorded. The achieved accuracy is evaluated according to the metrics described (4.3).

To determine the MTTR, the study is extended. First, two failure types are defined – low-confidence pose estimation and failed grasping. A random selection between both is performed for each intervention. Within a session, the timespan between the following steps is measured for evaluation: putting on the VR headset, interconnection, spatial acquisition of the fault case, teleoperating the robot and target pose determination.

4.3 Evaluation Metrics

Throughout the literature, several different metrics for evaluating 6DoF pose accuracy have been proposed. To quantify the labeling as well as the resulting pose estimation accuracy, the following metrics are chosen.

Average Distance (ADD) [22] computes average distance between ground truth 6DoF pose and labeled or estimated pose utilizing the component model $\mathcal{M}$. With given ground truth pose $\widehat{\mathbf{P}}$ and estimated pose $\overline{\mathbf{P}}$, the average distance of model points is calculated (cf. [23]) as – utilized for test components (1), (3) and (5).

$$e_{ADD}(\widehat{\boldsymbol{P}}, \overline{\boldsymbol{P}}; \mathcal{M}) = \underset{x \in \mathcal{M}}{avg} \left\| \overline{\boldsymbol{P}}x - \widehat{\boldsymbol{P}}x \right\|_2. \tag{1}$$

Average Distance for objects with Indistinguishable views (ADI) [22] is similar to ADD-metric, but adapted for components with symmetric rotation shape and is defined as – utilized for test components (2) and (4).

$$e_{ADI}(\widehat{\boldsymbol{P}}, \overline{\boldsymbol{P}}; \mathcal{M}) = \underset{i \in \mathcal{M}}{avg} \ \underset{j \in \mathcal{M}}{min} \left\| \overline{\boldsymbol{P}}i - \widehat{\boldsymbol{P}}j \right\|_2. \tag{2}$$

Visible Surface Discrepancy (VSD) [23] is proposed to deal with cases of pose ambiguity. VSD is suitable for both symmetric as well as non-symmetric objects and defined as

$$e_{VSD}(\widehat{\boldsymbol{P}}, \overline{\boldsymbol{P}}; \mathcal{M}, I, \delta, \tau) = \underset{p \in \widehat{V} \cup \overline{V}}{avg} c(p, \widehat{D}, \overline{D}, \tau) \tag{3}$$

$\hat{V}$ and $\bar{V}$ represent 2D masks of the visible surface rendered from $\mathcal{M}$ ($\hat{\mathcal{M}} = \hat{\mathbf{P}}\mathcal{M}$ and $\bar{\mathcal{M}} = \bar{\mathbf{P}}\mathcal{M}$) with estimated pose $\hat{P}$ and ground truth $\bar{P}$. The tolerance δ is defined to determine visibility.

With distance images $\hat{D}$ and $\bar{D}$ rendered at the estimated and ground truth, matching cost c is calculated as

$$c(p, \hat{D}, \bar{D}, \tau) = \begin{cases} d/\tau & if\ p \in \hat{V} \cap \bar{V} \wedge d < \tau \\ 1 & otherwise, \end{cases} \tag{4}$$

with $d = |\hat{D}(p) - \bar{D}(p)|$ as distance between the surfaces of $\hat{\mathcal{M}}$ and $\overline{\mathcal{M}}$ at pixel p. Thereby, τ denotes the misalignment tolerance limiting the allowed range of d.

5. Results and Discussion

Following, the results regarding time efficiency and accuracy for data annotation utilizing the proposed AV-based approach in comparison to the established AprilTag marker technique are presented in Figure 7.

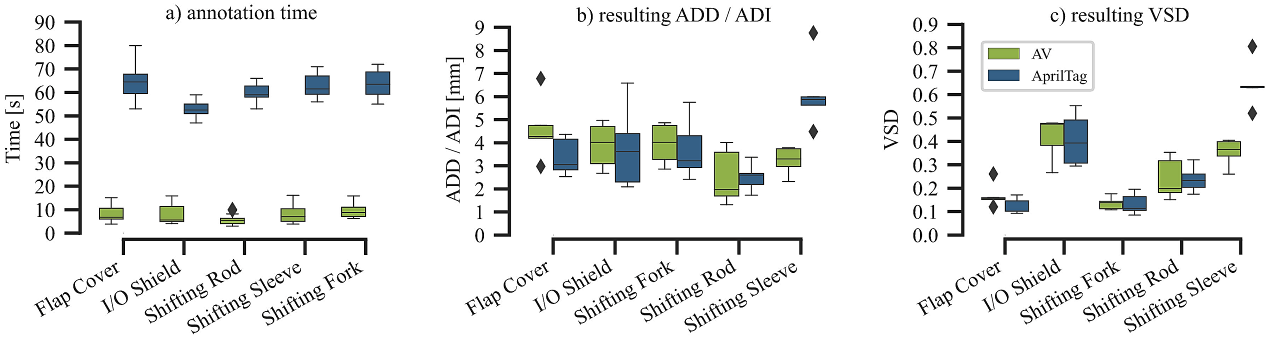


Figure 7: AV annotation versus AprilTag regarding a) the timespan required to create a single annotation, b) the resulting ADD (components: (1), (3) & (5)) and ADI ((2) & (4)) measures and c) the resulting VSD vs. ground truth

5.1 Analysis of the Annotation Time

The experiments show that fiducial marker annotation requires about one minute (cf. Figure 7 (a)). Variance in timespan is mainly caused by the subject experience and the component complexity. The main portion of time consumption is caused by the procedure of marker usage. Additionally, there is geometry dependent preparation time, which increases the required timespan, but scales with annotation numbers. On the other hand, timespans required for AV annotation are 85 % less. On average annotating a component requires about eight seconds as there are no preparation procedures required as well as the faster, immersive interaction with the HMI. Part complexity does not take an equally important role as it does with AprilTag labeling. However, subjects and their experience with using the HMI do have a slight impact.

5.2 Analysis of the Annotation Accuracy

To evaluate the accuracy of the pose estimation of AV versus AprilTag annotations, ground truth is obtained by duplicating the real-world scene and placing models with known poses at measured locations within the scene. Applying the described 6DoF pose evaluation metrics, resulting annotation quality can be quantified.

Regarding ADD or ADI, as applicable, quite similar results are achieved with each annotation method (cf. Figure 7 (b)). The proposed approach is nearly as accurate as annotating with AprilTags, however, requires only about 12 % of the time. For shifting sleeve, which is a hollow part, creating pose labels with AprilTags is more imprecise since measurement of component geometry and center of gravity is complicated. Results for VSD are nearly identical for both approaches (c). With respect to the VSD metric, resulting values below 0.5 are considered as good. Here, AprilTag is outperformed for 6DoF poses obtained for shifting sleeve.

5.3 Demonstration of Operability using Annotated Data under Industrial Application Conditions

The shifting fork component serves as chosen example to demonstrate operability of the annotation method to generate different annotation data types required for an mixed image processing pipeline in a single step (e. g. YOLO + FPN). Final datasets were automatically complemented through varying object and camera poses (see 3.2.2). In the end, 150 component specific measurements are taken into account for YOLO as well as 700 for FPN. Thereby, single- and multi-object setups are evaluated separately. Utilized and evaluated is an initial, unrefined dataset, generated for the system ramp-up stage.

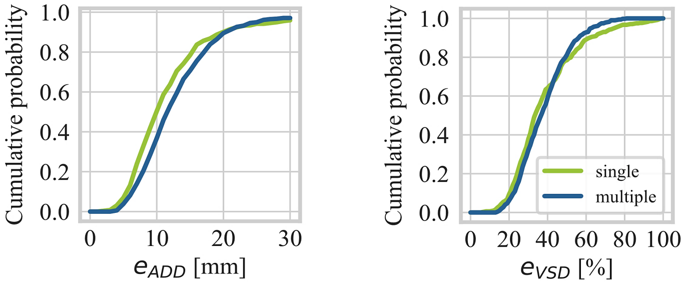


Figure 8: Pose estimation errors retrieved for scenarios with single and multiple objects, visualized as the empirical cumulative distribution function in terms of ADD and VSD metric for shifting fork RGBD real-world dataset

The performance of YOLO in terms of mean average precision (mAP@50IOU) reaches 91.42 % for all five components. Specifically, for shifting fork, an average precision (AP@50IOU) of 95.8 % is obtained. For ADD, a pose estimate is considered correct if error $e_{\mathrm{ADD}} \leq k \cdot d$. Thereby, k is a constant to be chosen and d the largest distance between any pair of model vertices, i. e. the diameter or length. Shifting fork features a dimension of 167 mm × 130 mm × 28 mm (l × w × h). For VSD, a pose estimate is considered correct if $e_{\mathrm{VSD}} \leq 0.5$ is met. Within this evaluation, parameters are set to $k = 0.1$ as well as to $\tau = 30$ mm and $\delta = 30$ mm. In single object scenario a recall$_{\mathrm{ADD}}$ of 85.3 % and a recall$_{\mathrm{VSD}}$ of 78.5 % are obtained. In multi-object scenario, a recall$_{\mathrm{ADD}}$ of 78.7 % and a recall$_{\mathrm{VSD}}$ of 81.2 % are achieved. e_{ADD} and e_{VSD} visualized as empirical cumulative distribution function are shown in Figure 8. In conclusion, the results obtained are rated as good, especially since solely real-world data without any additional mixed synthetic data is used.

5.4 Analysis of Fault Clearance Potential

To determine the advantage generated by the intervention strategy, a small user study is performed. Therefore, a bin-picking failure is simulated and the timespan resolving the issue is recorded. On average it took about 55 seconds to return the robot into autonomous operation. Compared to manual fault clearance, where the operator walks to the robot, triggers emergency safety circles before taking action, the AV approach saves time. This applies even more for supervising multiple bin-picking systems, since no transit time is required. Besides teleoperated solving, our approach provides direct target pose specification and subsequent automated object handling. Simultaneously, annotated data for dataset refinement is generated.

6. Conclusion and Outlook

In this work, an annotation and refinement method for RGBD data in rigid objects industrial bin-picking is described. The results obtained by the proposed method show a noticeable reduction in effort, measured by the required timespan for annotation, while maintaining high annotation accuracy at least competitive to the state of the art. It provides support for multi-object- and automated annotation as well as different I/O-formats. In addition to pure remote intervention, the HuITL approach enables optimization and enhanced adaptation during system ramp-up through dataset refinement. Although, the method focusses on online data processing, the method is also suitable for dataset generation based on once recorded offline sensor data. Future research will concentrate on hybrid reinforcement- / imitation learning for component grasping.

References

[1] Obermeier, B., Treugut, L. Learning Systems in Hostile Environments, in: , Lernende Systeme 2019, Munich.

[2] Jo, H.-J., Min, C.-H., Song, J.-B., 2018. Bin Picking System using Object Recognition based on Automated Synthetic Dataset Generation, in: 2018 15th International Conference on Ubiquitous Robots (UR). 2018 15th International Conference on Ubiquitous Robots (UR), Honolulu, HI. IEEE, pp. 886–890.

[3] Blank, A., Berg, J., Zikeli, G.L., Lu, S., Sommer, O., Reinhart, G., Franke, J., 2020. Intervention strategy for autonomous mobile robots. wt Werkstattstechnik online 110 (09), 613–618.

[4] Blank, A., Hiller, M., Zhang, S., Metzner, M., Lieret, M., Thielecke, J., Franke, J., 2019. 6DoF Pose-Estimation Pipeline for Texture-less Industrial Components in Bin Picking Applications, in: , IEEE ECMR, Prague, Czech.

[5] Wang, J., Olson, E. AprilTag 2: Efficient & robust fiducial detection, in: , 2016 IEEE IROS 2016, pp. 4193–4198.

[6] Zimmer, W., Rangesh, A., Trivedi, M., 2019. 3D BAT: A Semi-Automatic, Web-based 3D Annotation Toolbox for Full-Surround, Multi-Modal Data Streams, in: 2019 IEEE Intelligent Vehicles Symposium (IV). 2019 IEEE Intelligent Vehicles Symposium (IV), Paris, France. IEEE, pp. 1816–1821.

[7] Sager, C., Zschech, P., Kühl, N., 2021. labelCloud: A Lightweight Domain-Independent Labeling Tool for 3D Object Detection in Point Clouds, in: CAD’21 Conference.

[8] Arief, H.A., Arief, M., Zhang, G., Liu, Z., Bhat, M., Indahl, U.G., Tveite, H., Zhao, D., 2020. SAnE: Smart Annotation and Evaluation Tools for Point Cloud Data. IEEE Access 8, 131848–131858.

[9] Marion, P., Florence, P.R., Manuelli, L., Tedrake, R., 2018. Label Fusion: A Pipeline for Generating Ground Truth Labels for Real RGBD Data of Cluttered Scenes, in: 2018 IEEE ICRA. 2018 IEEE International Conference on Robotics and Automation (ICRA), Brisbane, QLD. IEEE, pp. 3235–3242.

[10]Ramirez, P.Z., Paternesi, C., Luigi, L. de, Lella, L., Gregorio, D. de, Di Stefano, L., 2020. Shooting Labels: 3D Semantic Labeling by Virtual Reality, in: 2020 IEEE AIVR. 2020 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR), Utrecht, Netherlands. IEEE, pp. 99–106.

[11]Blank, A., Baier, L., Kedilioglu, O., Zhu, X., Metzner, M., Franke, J., 2021. Efficient AI Adaption using Synthetic Data. wt Werkstattstechnik online 111 (10), 759–762.

[12]Sankaranarayanan, S., Balaji, Y., Jain, A., Lim, S.N., Chellappa, R., 2018. Learning from Synthetic Data: Addressing Domain Shift for Semantic Segmentation, in: 2018 IEEE CVPR. 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA. IEEE, pp. 3752–3761.

[13]Budd, S., Robinson, E., 2019. A Survey on Active Learning and Human-in-the-Loop for Medical Image Analysis.

[14]Maadi, M., Aickelin, U., 2021. A Review on Human-AI Interaction in Machine Learning and Insights for Medical Applications. International journal of environmental research and public health 18 (4).

[15]Yu, F., Seff, A., Zhang, Y., Song, S., Funkhouser, T., Xiao, J., 2015. LSUN: Construction of a Large-scale Image Dataset using Deep Learning with Humans in the Loop.

[16]Blank, A., Kosar, E., Karlidag, E., Guo, Q., Kohn, S., Sommer, O., Walter, J., Herbert, M., Sessner, J., Querfurth, F., Metzner, M., Franke, J., 2021. Hybrid Environment Reconstruction Improving User Experience and Workload in Augmented Virtuality Teleoperation. Procedia Manufacturing 55, 40–47.

[17]Kohn, S., Blank, A., Puljiz, D., Hein, B., Franke, J., 2018. Towards a Real-Time Environment Reconstruction for VR-Based Teleoperation Through Model Segmentation, in: 2018 IEEE/RSJ IROS, Madrid, pp. 1–9.

[18]Blank, A., Karlidag, E., Zikeli, L., Metzner, M., Franke, J., 2021. Adaptive Motion Control Middleware for Teleoperation Based on Pose Tracking and Trajectory Planning, in: , Annals of Scientific Society for Assembly, Handling and Robotics, vol. 55. Springer, Berlin, Heidelberg, pp. 153–164.

[19]Heuss, L., Blank, A., Dengler, S., Zikeli, G.L., Reinhart, G., Franke, J., 2019. Modular Robot Software Framework for the Intelligent and Flexible Composition of Its Skills, in: , Advances in Production Management Systems, vol. 566. Springer International Publishing, Cham, pp. 248–256.

[20]Redmon, J., Farhadi, A., 2018. YOLOv3: An Incremental Improvement. Computing Research Repository (CoRR).

[21]Qi, C.R., Liu, W., Su, H., Guibas, L.J., 2018. Frustum PointNets for 3D Object Detection from RGB-D Data, in: 2018 IEEE/CVF CVPR, Salt Lake City, USA, pp. 918–927.

[22]Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., Navab, N., 2013. Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes, in: , Computer Vision – ACCV 2012, vol. 7724. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 548–562.

[23]Hodaň, T., Matas, J., Obdržálek, Š., 2016. On Evaluation of 6D Object Pose Estimation, in: Computer Vision ‐ ECCV 2016 Workshops. Springer International Publishing, Cham, pp. 606 ‐ 619.

Biography

Andreas Blank has been a research associate at the Institute for Factory Automation and Production Systems (FAPS) and head of the Industrial and Service Robotics technology field (2014-2021). As PhD student his research is focused on immersive teleoperation, autonomous robot capabilities and human-to-robot skill transfer. He is Production Manager at BIG within Simba-Dickie-Group since 2021 leading factory automation and lean production integration.

Lukas Baier has been a research associate at the Technische Hochschule Ingolstadt (THI) and head of the Production Control and Intralogistics technology field at FAPS (2016-2021). As PhD student at FAPS his research is focused on sensor technologies and data processing for label-free goods identification in intralogistics. He is Senior Expert Operational Excellence at SIEMENS since 2021.

Maximilian Zwingel is a research associate at the Institute for Factory Automation and Production Systems at the FAU Erlangen-Nuremberg with prior research at the THI since 2018. His research is focussed on autonomous mobile robots in intralogistics environments, especially their sensors and navigation.

Prof. Dr.-Ing. Jörg Franke is Head of the Institute for Factory Automation and Production Systems (FAPS) at the FAU. His research focusses on innovative manufacturing processes for mechatronic products. He is involved in leading functions in scientific organisations such as IEEE, CIRP, WGP and 3D-MID. Before his professorship, he led different management positions in industry e. g. at McKinsey & Co., Robert Bosch GmbH, ZF Lenksysteme GmbH and Schaeffler AG.