ECAI 2024
U. Endriss et al. (Eds.)

doi:10.3233/FAIA241046

An Intelligent Pipeline for Localization of Industrial Components in Robotic Manufacturing Applications

Parth Rawal[a], Daniel Valencia[a] and Wolfgang Hintze[a,b]

[a]Fraunhofer Institute for Manufacturing Technology and Advanced Materials (IFAM), Ottenbecker Damm 12, Stade, 21684, Germany
[b]Institute of Production Management and Technology (IPMT), Hamburg University of Technology TUHH, Denickestraße 15, Hamburg, 21071, Germany
ORCID (Parth Rawal): https://orcid.org/0000-0001-8469-5783, ORCID (Daniel Valencia): https://orcid.org/0009-0005-1128-8198, ORCID (Wolfgang Hintze): https://orcid.org/0000-0001-9025-8803

Abstract. The rising skill shortage problem in Europe threatens the economic slowdown in the manufacturing sector. Approaches based on artificial intelligence can play a crucial role in bridging the shortage gap if they can be integrated into robot-assisted production to simplify repetitive manual tasks. Localizing components in a production cell is a familiar problem of robot-assisted production. Robots are often taught trajectories manually, which requires expertise in robot programming. Some of the existing feature-based computer vision solutions can localize a component in 3D space. However, these solutions are not versatile enough to be integrated across different components and production cells. This paper proposes an AI-based solution in the form of a pipeline for the 6D localization of components that can be integrated into multiple industrial use cases. The pipeline encompasses flows for generating synthetic images of components from their CAD model, training deep neural networks to estimate component poses, and improving their accuracy for manufacturing applications. The performance of the pipeline has been validated for components in a production-related environment. The paper also demonstrates the versatility of the pipeline by deploying it for a robotic spray coating use case. Such AI skills can empower the skilled workforce on the shop floor so that they can focus on the overall manufacturing process.

1 Introduction

Despite the numerous technological advancements made over the past two centuries, the human workforce has consistently held a vital position in ensuring the continuous production of manufactured goods. However, the manufacturing sector faces a significant challenge due to recent reports of skilled workforce shortages in Europe [7]. The problem is particularly severe in Germany, where the shortage has reached an all-time high, with nearly half of the companies reporting it [20]. Moreover, the necessity to transition towards a climate-neutral economy has added to the challenge of increasing production in the near future.

At the same time, there is a notable surge in global investments in artificial intelligence (AI). This trend is also evident in Europe, where total AI investment is projected to reach approximately 186 billion Euros by 2030 [18]. As companies begin to implement AI solutions to address the challenges posed by the shortage of skilled workers, there is a potential to enhance labor productivity by up to 40% in Europe by 2035 [16]. This can be done by assisting or augmenting the skilled workforce using AI technologies and helping them perform repetitive, recurrent tasks more efficiently, thereby focusing more on higher value-adding activities. Moreover, AI is also paving the way for intelligent automation solutions that are different from traditional ones in terms of adaptibility, cognition, and decision making. In the manufacturing sector, robotics, smart sensors, and smart automation are the high demand areas for AI technologies [24].

One such application of AI technologies is robot-based manufacturing, where a component's 3D position and 3D orientation (together known as 6D pose) is needed before the manufacturing process starts. Robots in a production cell need a precise description of a component's position and orientation to execute the planned tasks. In many cases, the position of a component is not fixed and cannot be taught directly to the robot. In this case, a component's position and orientation must be determined dynamically before starting the process. Solutions addressing these problems include using measurement probes and vision sensors. They are mainly developed focusing on the production setup and environment, such as the geometry of the components, local features, lighting, etc. The local features of a component, such as holes, pockets, edges, and corners, may provide an advantage in that the overall geometry of a component must not be known. While this approach works as desired, integrating it into another use case is extremely difficult without reprogramming and revalidation. Surrounding light often causes problems if the illumination conditions inside the production environment change. On the other hand, best-fit approaches based on the CAD model allow the algorithms to be reused but do not guarantee the final accuracy achieved due to the solution getting trapped in local minima.

With numerous valuable developments in AI over the last few years, extracting 3D information from a scene, such as the 6D pose of a component can be AI driven. In deep learning, convolution neural networks (CNNs) have emerged as one of the most successful tools for easy and scalable AI implementation in different domains. However, to train AI, annotated datasets are needed. Furthermore, these AI implementations are mostly validated for standard datasets in computer vision and do not consider the challenges related to textureless industrial components and varying production environments. If these problems can be solved, AI-based solutions can be deployed

for multiple use cases without much rework just by training them on a different dataset.

2 Related work

In manufacturing, the critical task of localizing a component involves determining its pose relative to the robot's location within the manufacturing cell. This is often accomplished by integrating vision-based metrology systems that utilize cameras or other visual sensors to capture data of the component in the form either images, depth maps or a point cloud. Advanced computer vision algorithms and techniques, including deep learning and point cloud processing, are employed to tackle the challenges of accurately estimating the 6D pose of objects from various sensor inputs such as RGB images, depth maps, or point clouds.

Marker-based pose estimation is a widely adopted approach that utilizes targets and cameras for accurate pose estimation and component identification. Its main advantage is that precise measurements rely on robust feature extraction and subpixel-accurate target detection. By using coded unique identifiers, component identification can be achieved with precision, allowing for associations with instructions and product passes. However, the installation of targets increases lead time and can potentially introduce errors. In [4], a multi-camera system with ten cameras was employed to simultaneously validate the shape fidelity of components on an entire holding fixture.

The estimation of a pose without the use of markers relies on the features of the component. Traditional objects pose estimation methods, based on image and depth data, typically involve matching local features. For instance, the Scale-Invariant Feature Transform (SIFT) [15] method is used to extract local features such as corners or blobs and represent them as descriptors. An example of a large-scale vision-based metrology system that relies on features rather than targets is described in [25]. This system is utilized to determine the position of rivets on the skin of fuselage shells, enabling shape and position adjustment for the assembly of fuselage sections. While these techniques work well with texture-rich objects, they may struggle with texture-poor items commonly found in industrial environments.

When it comes to using point clouds, the Iterative Closest Point (ICP) algorithm [3] is a trusted and effective method for aligning 3D models by minimizing the distance between corresponding points. However, it requires good initial pose estimates or otherwise can get stuck in local minima. In the context of additive manufacturing, a real-time object pose tracking system using ICP has been showcased by Liu et al. [14], aimed at automating the depowdering process post 3D printing. This system accurately determines the pose of 3D-printed parts, even when they are partially obscured. Utilizing the continuously updated 6D pose data, the system dynamically plans and modifies the robot's trajectory, enabling efficient powder removal through methods such as vacuuming or air blasting.

Recent single-stage methods for estimating an object's 6D pose have been developed using deep learning techniques. These approaches include 6D pose estimation from a single RGB image [12, 19, 5] and multiple RGB images [13]. Deep learning has also been implemented for RGB-D image-based pose estimation approaches [11, 26]. For instance, Xiang et al. [28] introduced a convolutional neural network that estimates a 6D pose from an RGB image and then refines the pose with ICP using the depth map. While point cloud-based methods [9, 21, 22] for 6D pose estimation have been proposed, they are less common. This is primarily due to the unstructured nature of point clouds, which poses challenges for conventional convolutional architectures [8, 10].

Deep learning-based methods, when trained on realistic training images, generally exhibit greater robustness to scene attributes like illumination and contrast and lead to better results [6]. However, these approaches have been mainly evaluated for computer vision datasets and lack results for industrial components in production cells. Integrating deep learning with more conventional methods can leverage both texture and geometry-based features, enhancing the overall performance by combining the initial estimate from the deep neural network with the precision of the ICP algorithm. For example, the robotic spray painting application presented in the work of Wang et al. [27] uses such a combined approach. The target object, a chair, is classified into four basic orientations using deep neural networks. After that, ICP registration is used to refine the rough estimate. In this approach, classifying into four orientations may not always suffice, as there is a high risk of obtaining a local minimum solution with ICP, particularly for symmetric and near-symmetric objects. Another drawback of the proposed approach is that it was only demonstrated on a single object, and hence versatility cannot be guaranteed.

AI-driven approaches are recognized globally as important catalysts for factory transformation towards digital evolution [2]. Industrial robots with smart sensors are becoming increasingly popular. However, the advanced systems required for these robots are often not readily available from manufacturers or are only partially provided. Many of these systems are still in the experimental phase or undergoing trials before commercial deployment or customization for specific applications. The transition from a proven concept in the laboratory to a fully operational system in real-world settings is often a time-consuming and intricate process [1].

3 Pipeline-based concept

After carefully considering the limitations and difficulties associated with various technologies, a novel approach has been developed for extracting a 6D pose of a component in robotic manufacturing applications. This approach is based on a multi-stage pipeline concept, representing a network of pre-programmed service tasks with predefined capabilities. The key advantage of this concept is that it eliminates the need for reprogramming. The same pipeline can be utilized for different components or different production environments by simply adjusting the parameters of the service tasks to accommodate various CAD models. Figure 1 showcases the pipeline and illustrates the overall working concept to provide a visual representation of the concept.

The pipeline network consists of two main flows. The first flow, known as the training flow, focuses on synthetic data generation and AI training for estimating the 6D pose of the component. The primary objective in this stage is to generate high-quality training data efficiently. This data is then utilized in subsequent stages to train multiple AI networks. Modifications are required in the AI data-loading scripts to ensure compatibility with the data generated in the previous step. It is important to note that most existing AI implementations are designed and benchmarked for general datasets that typically contain simple objects. However, in the manufacturing sector, components exhibit diverse geometries and materials. Therefore, a thorough examination and careful selection of data generation methods are necessary. The advantage of this approach is that the training data generated once can be leveraged to train multiple AI networks for various tasks like object detection and segmentation. The implementation details of this flow are provided in Section 4, where a com-

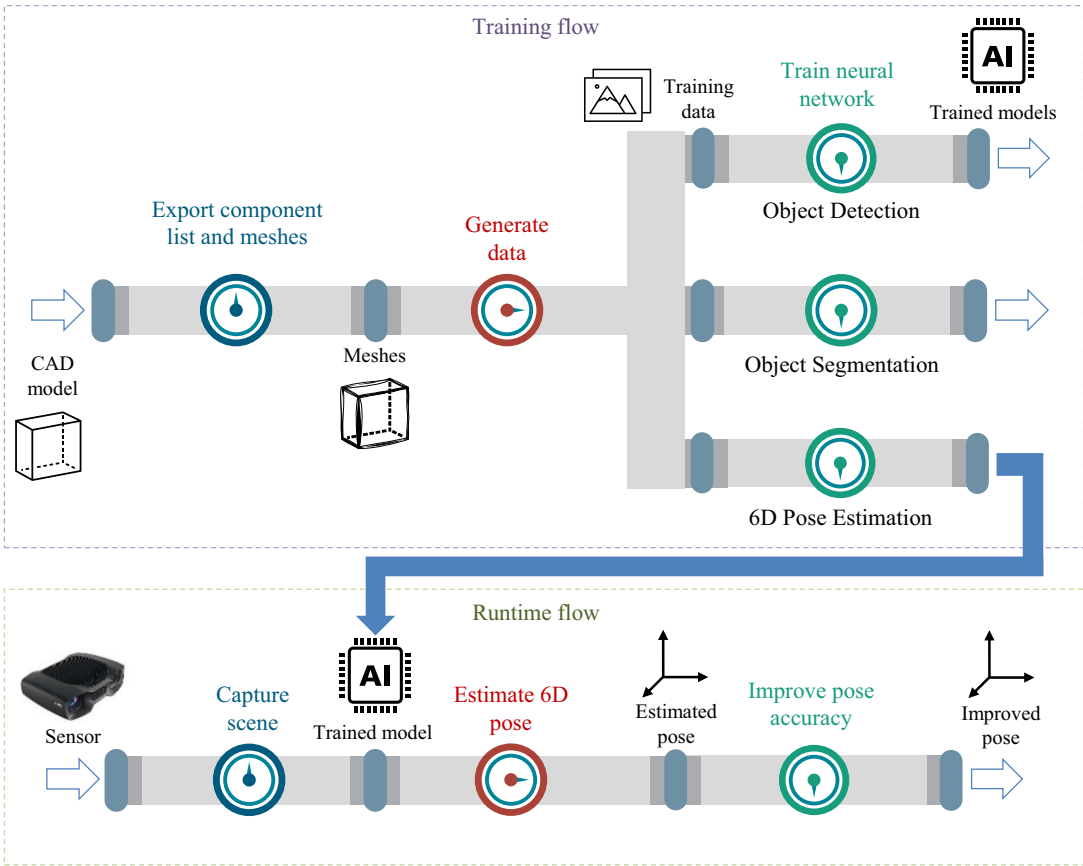

Figure 1. Pipeline for extracting a 6D pose of a component

prehensive explanation is provided. The runtime flow is designed to obtain the final 6D pose of a known component from a scene. It is essential that the training flow is executed beforehand, as the trained models are utilized on captured sensor data to obtain an initial estimate of the component's position. Subsequently, ICP-based pose refinement method is employed to improve the AI-estimated pose, thereby enhancing the accuracy. Detailed explanations of this method is provided in Section 5.

This flow-based approach ensures that the pipeline network is reusable and versatile. Even if synthetic data generation and training need to be repeated for different components and use cases, they can be accomplished with minimal human effort. Thanks to the advancements in graphics cards, the entire training flow can be completed within a few hours. Additionally, the runtime flow is completely reusable as it relies on the mesh data of the known component. The initial pose predicted by AI serves as a safeguard against local minima during the fine adjustment stage.

To test the robustness of the pipeline in different production environments, a special demonstrator was used. This demonstrator consists of two components mounted on an industrial hand cart, as shown in Figure 2. One component, *ManholeBox*, has distinctive features on all sides, while another component, *GeometricPlate*, is a planar object with significant features only on its top and bottom sides. The overall setup is modular, which means parts can be translated and rotated in different directions. This flexibility enables thorough testing to ensure the pipeline can handle changes in the production environment effectively.

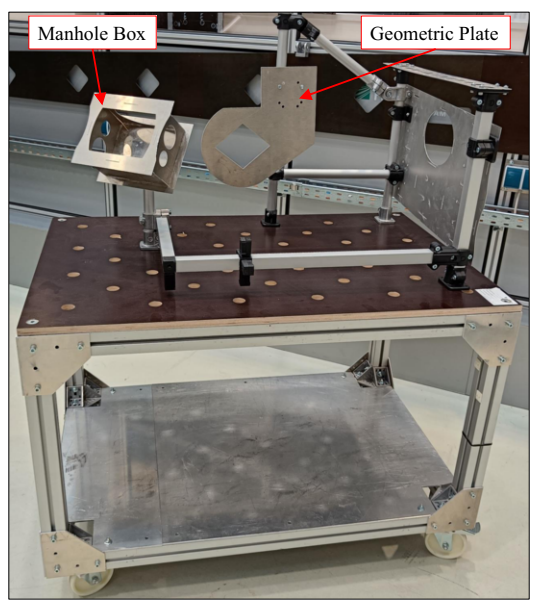

Figure 2. Demonstrator with two components used for validation

4 Synthetic data generation and 6D pose estimation

Getting realistic labeled data for training AI is one of the most critical aspects of achieving high AI performance. An AI that predicts

an object's 6D pose needs an accurate description of the object's position and orientation in 3D space for training. Gathering this data is a highly laborious and time-consuming process. However, several script-based synthetic data generation tools have emerged in recent years, enabling a scalable generation of training data [6, 17].

For large assembly CAD models in production, the training flow described in Figure 1 can be split into three parts. The first part of the flow exports the list of components and meshes from the CAD model, whose 6D pose needs to be extracted later. The second part of the flow generates thousands of synthetic images and ground truth required for training AI. This data is later used in the final part of the flow to train a suitable AI for estimating the 6D pose of the components. The synthetic data generation approach used here to generate high-quality data using CAD models for training has been addressed separately in the work of Rawal et al. [23]. For this, different data generation procedures are formulated and used in various combinations to generate a scene for rendering. The scene is defined by randomizing several entities, such as realistic textured planes, lights, camera perspective, object materials, etc. At the same time, the target domain knowledge from the assembly CAD model, such as physical constraints between different components in an assembly, is preserved. This allows a good mix of training data that can be used for a use case. For a different use case, the data can be regenerated from its CAD model with minimal manual effort. Using this approach, 15K images of the demonstrator were generated in 12 hours. Figure 3 shows some of the sample images generated using the CAD model of the demonstrator.

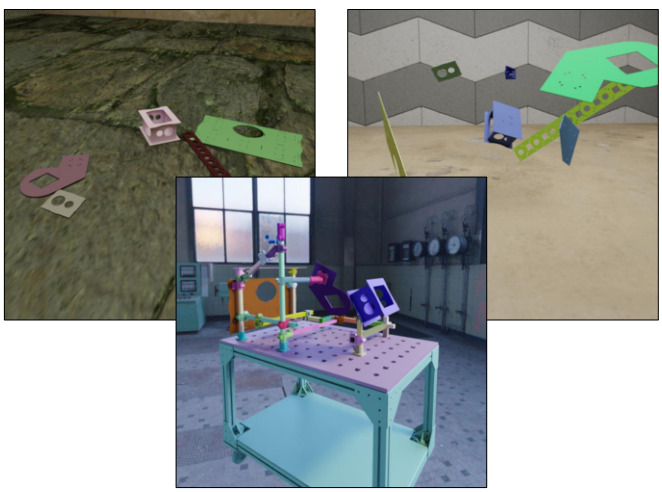

Figure 3. Synthetic image samples generated from demonstrator components

The generated dataset is then used to train CNN-based deep neural networks. Several implementations exist for estimating an object's 6D pose in computer vision. However, the implementations that estimate 6D pose from a single RGB image are more popular because they can be used with low-cost sensors such as webcams. In this work, two state-of-the-art proven networks, *DOPE* [12] and *EfficientPose* [5], were chosen and adapted so that they could be trained with the custom-generated dataset. The *DOPE* network is based on first detecting 2D keypoints in images and later estimating the 6D pose using the Perspective-n-Point (PnP) algorithm. *EfficientPose* network, on the other hand, is regression-based and can directly predict the position and orientation of an object.

The training was started for the generated dataset with standard parameters provided by the frameworks. The training curves for both networks converged well after 100 epochs. For *DOPE*, every object had to be individually trained, which increased the training time. This was not the case for *EfficientPose*. *DOPE* often failed to detect all keypoints, resulting in missing prediction instances, and is therefore unable to estimate the pose. On the other hand, *EfficientPose* demonstrated greater robustness with consistent predictions. Hence, in this paper, results only from *EfficientPose* neural network are presented. For *EfficientPose*, the ADD(-S) metric after training reached almost 83%, comparable to the metric achieved for multiple objects on the *Occlusion* dataset [5]. The average error in translation was limited to 22 mm and in rotation to 21°.

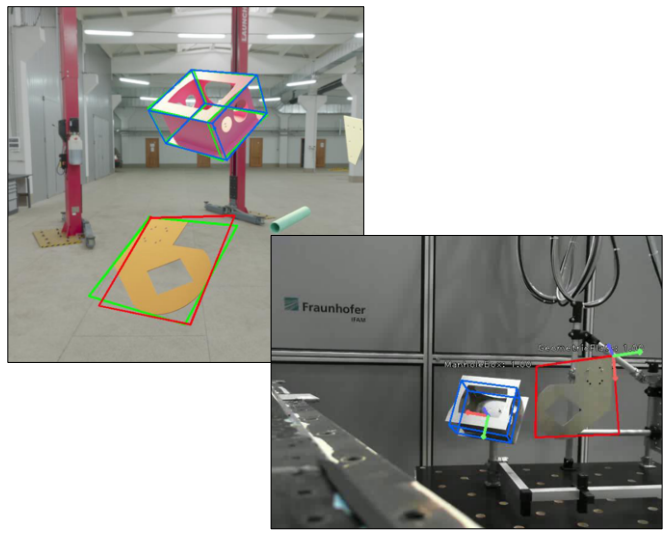

Figure 4. Ground truth and predictions with *EfficientPose* on synthetic and real images. Green bounding boxes represent ground truth poses, whereas bounding boxes with other colors represent the predictions.

Figure 4 illustrates the deviations between the ground truth and the predicted 6D poses for a synthetic image of the demonstrator objects. For a real image, it presents the inference results using the trained *EfficientPose* model. The image showcases the 3D bounding boxes and local frames of the components. As the model is trained on synthetic data, the simulation-to-reality gap also contributes to slight prediction inaccuracies. This discrepancy in accuracy can be attributed to the direct regression of the pose from the image without learning local 2D features.

5 Pose refinement and results

The accuracy of 6D pose estimation using deep learning approaches can be limited for high-precision manufacturing applications. However, this can be further improved by incorporating additional scene data, such as point clouds. By capturing point clouds of the objects, registration-based methods like ICP can be employed to reduce misalignment iteratively. This strategy complements the previous RGB image-based 6D pose estimation approach because a good initial transform in the form of a rough estimated pose can be provided as input to the ICP method. This helps to avoid getting stuck in local minimum solutions and improves the overall accuracy of the pose estimation. Furthermore, the image and depth channel of the data can

be utilized separately to extract more information and enhance the accuracy of the results. By leveraging both the RGB image and depth data, it is possible to improve the precision of the pose estimation process.

While performing ICP, the maximum correspondence distance parameter is a critical factor in conjunction with the initial transformation. Adjusting this parameter according to the initial transformation can achieve a significant number of inlier correspondences between the source and target point clouds. Ideally, an accurate initial transformation and a small value of maximum correspondence distance are desired for optimal results. However, if the initial transformation is inaccurate, increasing the maximum correspondence distance may be necessary to find more correspondences. However, this increase may introduce a potential risk of correspondence mismatch, which can result in a loose alignment.

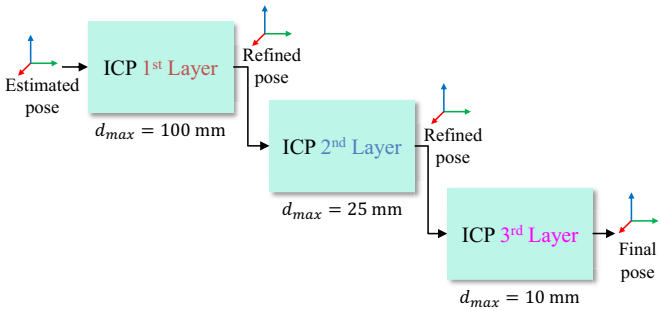

Figure 5. ICP block layers for pose refinement

To address this issue, the ICP method can be repeated multiple times with decreasing maximum correspondence distance. Figure 5 illustrates this concept with three layers of ICP blocks, where the variable d_{max} represents the maximum correspondence distance. The specific values shown in Figure 5 are tuned for the *ManholeBox* and *GeometricPlate* components used in the demonstration. However, these values may need to be adjusted depending on the scale of the components and the accuracy of the initial transformation. In each layer, the source point cloud is created by uniformly sampling the mesh files of the target components and the captured scene is chosen as the target point cloud for all ICP blocks. The AI-predicted pose is selected as the initial transformation for the first layer, and the refined pose obtained after alignment serves as the initial transformation for the subsequent layers.

For testing the overall pose extraction concept on components, sensors capable of capturing 3D scene information, such as color images, depth images, and point clouds, are needed. Here, *Zivid Two M70* [29] structured light sensor is used. The captured 3D data is provided in the form of an ordered point cloud, where each pixel in the corresponding image has a corresponding point in the point cloud. In Figure 6, an image and a colored point cloud captured by the sensor are shown. Both the image and the point cloud are saved in the sensor coordinate frame. The image has color values for all the pixels, while the point cloud only shows the triangulated points that are within the measuring range of the sensor.

Figure 7 shows the 6D pose prediction result on an image captured by the sensor. The estimated 6D poses are represented by plotting the 3D bounding boxes on the images. However, visualizing the translation error along the Z-axis of the sensor frame is extremely difficult because the Z-axis of the sensor frame is perpendicular to the image plane. Figure 8 visualizes these 6D poses by plotting them on

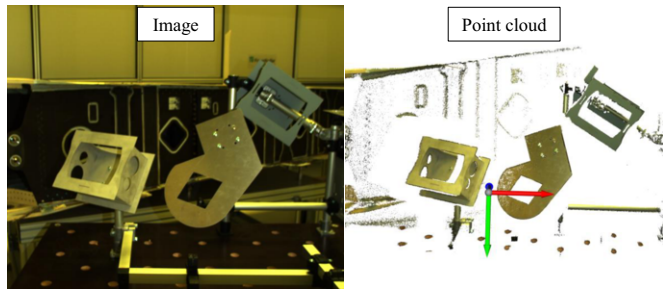

Figure 6. 3D scene data captured using *Zivid Two M70* sensor

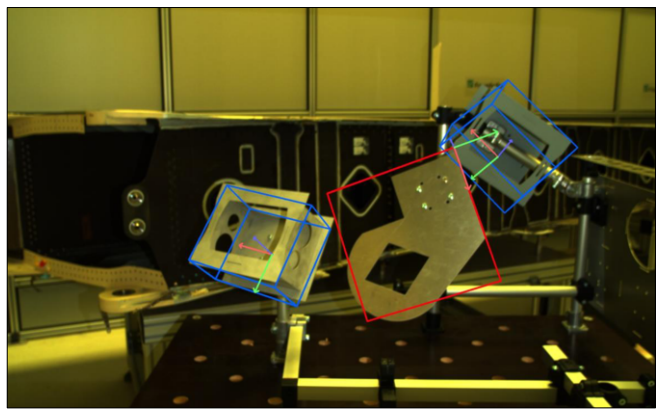

Figure 7. Estimated poses of demonstrator components using *EfficientPose* neural network

3D point clouds. Here, the estimated 6D poses are depicted as orange meshes, while the improved poses obtained after applying ICP layers are shown as green meshes. The point cloud is plotted in the sensor coordinate frame, which is also visible. It is worth noting that the green meshes align perfectly with the components present in the scene, indicating a successful convergence in the final registration result. The translation and rotation error can be easily spotted by looking at the offsets between the orange and the green meshes. These errors in the estimated 6D pose are effectively corrected using ICP.

For studying the pose estimation accuracy in detail, nine data frames of demonstrator components from different perspectives were captured and evaluated. Surprisingly, during the 6D pose estimation, the predicted poses consistently showed a tendency to be further away from the sensor compared to their actual positions in 3D space. This indicates a consistent error in the Z coordinate of the translation. To further investigate the impact of translation in each direction, statistical analysis of the results was conducted. Table 1 provides a summary of the results obtained from all instances. The table includes mean and standard deviation values for both components separately. The mean values indicate the accuracy of the final pose, assuming that the final pose after pose improvement resulted in an ideal fit. On the other hand, the standard deviation values reflect the precision of the estimation. The table also presents the correction values for individual translation elements in all three directions. Additionally, it shows the absolute correction values for translation and rotation in all directions.

Upon analyzing the individual translation correction values, it is evident that there is a significant correction along the Z-axis (Δ_{t_z}) of the sensor frame for both components. This observation confirms

Figure 8. Refined pose achieved using ICP. The coordinate frame represents the origin of the target point cloud. Meshes in orange show the estimated 6D pose, and meshes in green show the improved results.

Table 1. Pose correction results using ICP

Correction	*ManholeBox*		*GeometricPlate*	
	Mean	Std. Dev	Mean	Std. Dev
Δ_{t_x} (mm)	16.67	13.19	22.68	12.47
Δ_{t_y} (mm)	4.63	3.62	10.95	3.23
Δ_{t_z} (mm)	118.29	33.10	77.25	46.33
Δ_t (mm)	120.21	33.54	83.59	43.90
Δ_θ (°)	10.41	5.39	20.82	13.73

the earlier finding that the predicted poses consistently appeared further away from the sensor than their actual positions. In contrast, the pose correction values in the other directions are relatively low. Additionally, when considering the angular correction values, it is observed that the *ManholeBox* component has smaller corrections compared to the *GeometricPlate* component. This suggests that the 6D pose estimation for the *ManholeBox* component exhibits better accuracy in terms of orientation. This is because *ManholeBox* has geometric features on all the sides, whereas *GeometricPlate* is planar. Overall, the pose improvement steps successfully corrected approximately 120 mm of translation error and 20° of rotation error.

6 Use case: Robotic spray coating

After validating the pipeline network on the demonstrator components, it was integrated into an industrial robotic spray coating application. Figure 9 shows the setup of the robotic spray coating use case. A lightweight robot equipped with spray-coating end effector deposits corrosion-inhibiting compounds on the manufactured passenger aircraft door. The aircraft door is mounted vertically in the spraying chamber. The *Zivid* sensor is attached to the end effector to localize the aircraft door with respect to the robot. The sensor captures the component features, and from a known CAD model of the component, its accurate 6D pose can be extracted. The extracted pose is then transformed from the sensor to the robot base frame. For this,

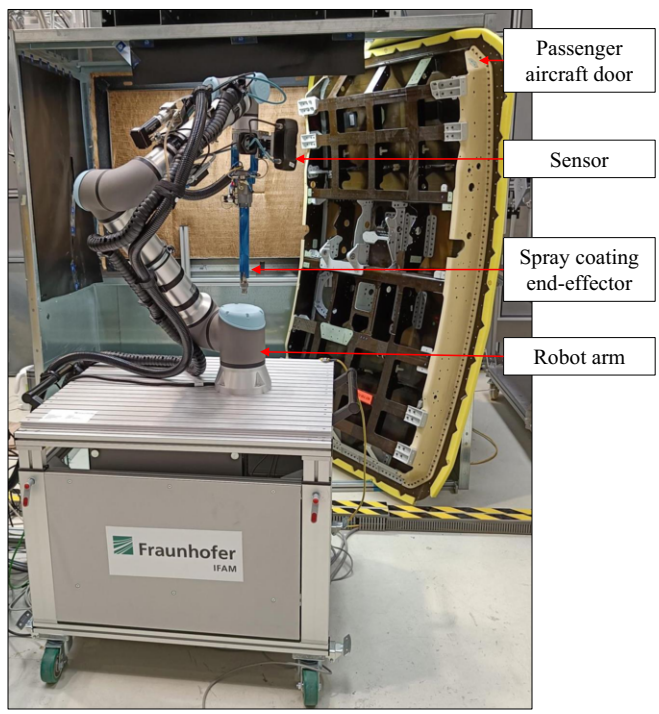

Figure 9. Robotic spray coating setup

the hand-eye calibration between the sensor and the robot flange is required in prior. The transformation between the robot base and the robot flange is known from the robot pose.

The pipeline-based approach developed here using the demonstrator could be easily adapted for the spray coating use case for the aircraft door in one week. This time was needed mainly for generating synthetic data and training neural networks. Figure 10 compares the estimated and final pose of the aircraft door. The captured point

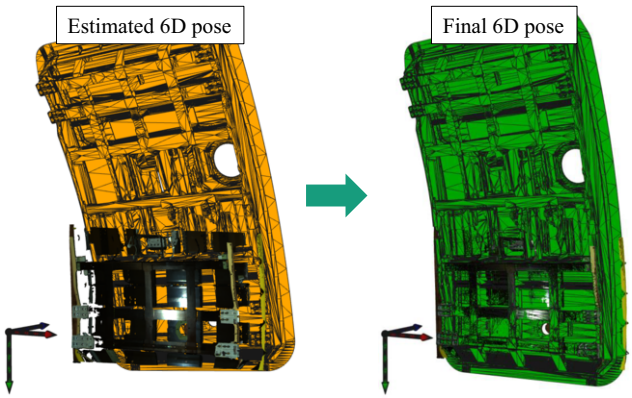

Figure 10. 6D pose estimation and pose correction of the passenger aircraft door. Mesh in orange show the estimated 6D pose, and mesh in green show the final result.

cloud covers the bottom part of the door. The final pose after refinement shows an excellent alignment. A translation correction of 72 mm was applied to the estimated pose to achieve the final pose result.

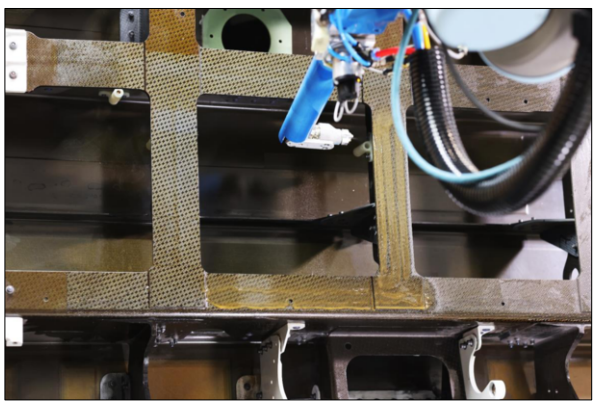

Figure 11. Robot spraying the designated areas on the passenger aircraft door

Once the 6D pose of a component is known accurately, a predefined robot program is adjusted to match the current pose of the component. After checking for collisions, the trajectories defined in the robot program are executed to deposit the corrosion-inhibiting compound on the component's surface. An accurate solution will result in less overspray of the corrosion-inhibiting compound. Figure 11 shows an image of the robot spraying the compound along the adapted path. The outcome of using this approach is not merely the automation of a manufacturing process, but also the creation of a versatile and reusable solution that fits multiple use cases and components. If the use case or components within a use case are modified, the training flow, as depicted in Figure 1, must be executed again after updating the CAD model of the production environment. The trained model can be ready for deployment within as little as two days, depending on the performance of the GPUs used for training. Compared to traditional computer vision approaches that do not utilize deep neural networks, this method eliminates the need for extensive programming and testing efforts for each use case. This not only reduces development-related costs but also lowers the skill level required to develop such a solution. Implementing this approach requires a moderate investment, such as setting up a high-performance AI server and a compatible robotic system. If required, these manufacturing applications can be scaled up for multiple use cases without significantly increasing costs. In this use case, the data acquisition and processing parts are digitalized so that a human operator can handle the spray coating application with a few clicks, focusing more on other essential aspects of the process. This is one example of how AI-based digital solutions in manufacturing can help human workforces accomplish tasks in an easy and simple way.

7 Conclusion

This work introduces a pipeline network-based approach for localizing a component to achieve a versatile solution for different robotic use cases. The pipeline network contains training and runtime flow. The training flow encompasses a synthetic data generation process and neural network training for estimating rough 6D poses in the scene. On the other hand, the runtime flow involves capturing data from the scene and utilizing the trained neural network model to predict the 6D poses of target components. To enhance the accuracy of these poses, a traditional ICP algorithm is employed for further refinement, resulting in refined poses of the components. The proposed pipeline is based on CAD models of industrial components. This characteristic ensures that the pipeline is reusable and adaptable for various robotic applications with minimal human effort.

The performance of the entire pipeline network was validated on an industrial demonstrator. The *EfficientPose* neural network was chosen to carry out the 6D pose estimation task. The images and point clouds captured by a structured light sensor were used for 6D pose estimation and improvement tasks. Such sensor fusion techniques are helpful as one can benefit from the data differently. The pipeline results show a robust performance and deliver highly accurate 6D poses of target components by simply using their mesh files for alignment. Furthermore, the versatility aspect of the pipeline was also demonstrated for a robotic spray coating use case. Accurate 6D poses extracted from the scene data were used to adapt offline trajectories for depositing corrosion-inhibiting compounds on the component's surface.

The AI tools and technologies used in this work have been developed in the last two to four years. However, because the scope and capabilities of AI technologies are proliferating day by day, the synthetic data generation and 6D pose estimation tools used in this work may become outdated in the near future. The future AI world is expected to be dominated by generative adversarial networks (GAN). New approaches based on GAN could provide newer ways for estimating 6D poses of objects. However, in general, such pipeline-based solutions are the way forward to close skill mismatch gaps in the manufacturing industry. Consequently, AI-based digital solutions should be prioritized to streamline manufacturing tasks and support human workforces with limited robotics expertise.

Acknowledgements

We thank the Lower Saxony Ministry for Economic Affairs, Transport, Housing and Digitalisation and the N-Bank for funding the project "Skotty" under grant ZW 1 80159842. We are also grateful to every project team member for successfully carrying out the work at Fraunhofer IFAM in Stade.

References

[1] J. Arents and M. Greitans. Smart industrial robot control trends, challenges and opportunities within manufacturing. *Applied Sciences*, 12 (2):937, 2022. ISSN 2076-3417. doi: 10.3390/app12020937.

[2] J. Arents, V. Abolins, J. Judvaitis, O. Vismanis, A. Oraby, and K. Ozols. Human–robot collaboration trends and safety aspects: A systematic review. *Journal of Sensor and Actuator Networks*, 10(3):48, 2021. doi: 10.3390/jsan10030048.

[3] P. J. Besl and N. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14 (2):239–256, 1992. ISSN 0162-8828. doi: 10.1109/34.121791.

[4] C. Brillinger, S. K. Kallipalayam Murugesan, C. Moeller, C. Boehlmann, W. Hintze, and D. Niermann. Accuracy analysis for a flow line process using a mobile holding fixture for machining cfrp components. *SAE International Journal of Advances and Current Practices in Mobility*, 4(4):1092–1103, 2022. doi: 10.4271/2022-01-0041.

[5] Y. Bukschat and M. Vetter. Efficientpose: An efficient, accurate and scalable end-to-end 6d multi object pose estimation approach. URL https://arxiv.org/abs/2011.04307.

[6] M. Denninger, M. Sundermeyer, D. Winkelbauer, D. Olefir, T. Hodan, Y. Zidan, M. Elbadrawy, M. Knauer, H. Katam, and A. Lodhi. Blenderproc: Reducing the reality gap with photorealistic rendering. In *International Conference on Robotics: Sciene and Systems, RSS 2020*, 2020. ISBN 2330765X. URL https://elib.dlr.de/139317/.

[7] European Labour Authority. and Fondazione Giacomo Brodolini. *Report on labour shortages and surpluses: 2022*. Publications Office, 2023. doi: 10.2883/50704.

[8] G. Gao. *Learning 6D Object Pose from Point Clouds*. PhD thesis, Staats- und Universitätsbibliothek Hamburg Carl von Ossietzky, 2021. URL https://ediss2.sub.uni-hamburg.de/handle/ediss/9098.

[9] G. Gao, M. Lauri, Y. Wang, X. Hu, J. Zhang, and S. Frintrop. 6d object pose regression via supervised learning on point clouds. URL https://arxiv.org/pdf/2001.08942.

[10] Y. Guo, H. Wang, Q. Hu, H. Liu, L. Liu, and M. Bennamoun. Deep learning for 3d point clouds: A survey. URL https://arxiv.org/pdf/1912.12033.

[11] Y. He, H. Huang, H. Fan, Q. Chen, and J. Sun. Ffb6d: A full flow bidirectional fusion network for 6d pose estimation. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3002–3012, Piscataway, NJ, 2021. IEEE. ISBN 978-1-6654-4509-2. doi: 10.1109/CVPR46437.2021.00302.

[12] Jonathan Tremblay, Thang To, Balakumar Sundaralingam, Yu Xiang, Dieter Fox, and Stan Birchfield. Deep object pose estimation for semantic robotic grasping of household objects. *Conference on Robot Learning*, pages 306–316, 2018. ISSN 2640-3498. URL https://proceedings.mlr.press/v87/tremblay18a.html.

[13] Y. Labbé, J. Carpentier, M. Aubry, and J. Sivic. Cosypose: Consistent multi-view multi-object 6d pose estimation. In A. Vedaldi, H. Bischof, T. Brox, and J.-M. Frahm, editors, *Computer Vision – ECCV 2020*, Springer eBook Collection, pages 574–591, Cham, 2020. Springer International Publishing and Imprint Springer. ISBN 978-3-030-58520-4. doi: 10.1007/978-3-030-58520-4{\textunderscore}34. URL https://link.springer.com/chapter/10.1007/978-3-030-58520-4_34.

[14] Z. Liu, J. Geng, X. Dai, T. Swierzewski, and K. Shimada. Robotic depowdering for additive manufacturing via pose tracking. *IEEE Robotics and Automation Letters*, 7(4):10770–10777, 2022. doi: 10.1109/LRA.2022.3195189.

[15] D. G. Lowe. Object recognition from local scale-invariant features. In *The proceedings of the Seventh IEEE International Conference on Computer Vision*, pages 1150–1157 vol.2, Los Alamitos, Calif., 1999. IEEE Computer Society. ISBN 0-7695-0164-8. doi: 10.1109/ICCV.1999.790410.

[16] Mark Purdy and Paul Daugherty. *Why artificial intelligence is the future of growth*. 2016.

[17] N. Morrical, J. Tremblay, Y. Lin, S. Tyree, S. Birchfield, V. Pascucci, and I. Wald. Nvisii: A scriptable tool for photorealistic image generation. URL https://arxiv.org/pdf/2105.13962.

[18] Next Move Strategy Consulting. Artificial intelligence (ai) market (by offering: Hardware, software, services; by technology: Machine learning, natural language processing, context-aware computing, computer vision; by deployment, 2023. URL https://www.nextmsc.com/report/artificial-intelligence-market.

[19] K. Park, T. Patten, and M. Vincze. Pix2pose: Pixel-wise coordinate regression of objects for 6d pose estimation. URL https://arxiv.org/pdf/1908.07433.

[20] A. Peichl, S. Sauer, and K. Wohlrabe. Fachkräftemangel in deutschland und europa – historie, status quo und was getan werden muss. *ifo Schnelldienst*, 75(10):70–75, 2022. ISSN 0018-974X. URL https://www.econstor.eu/handle/10419/272069.

[21] C. R. Qi, H. Su, K. Mo, and L. J. Guibas. Pointnet: Deep learning on point sets for 3d classification and segmentation, . URL https://arxiv.org/pdf/1612.00593.

[22] C. R. Qi, L. Yi, H. Su, and L. J. Guibas. Pointnet++: Deep hierarchical feature learning on point sets in a metric space, . URL https://arxiv.org/pdf/1706.02413.

[23] P. Rawal, M. Sompura, and W. Hintze. Synthetic data generation for bridging sim2real gap in a production environment. URL http://arxiv.org/pdf/2311.11039.

[24] I. Seifert, M. Bürger, L. Wangler, S. Christmann-Budian, M. Rohde, P. Gabriel, and G. Zinke. *Potenziale der Künstlichen Intelligenz im produzierenden Gewerbe in Deutschland*. iit - Institut für Innovation und Technik in der VDI/VDE Innovation und Technik GmbH, 2018. URL https://pure.mpg.de/pubman/faces/ViewItemFullPage.jsp?itemId=item_3248281_1.

[25] D. A. Valencia Zubiaga, J. Wollnack, S. Kamath, and L. Brieskorn. *Multi-camera Metrology System for Shape and Position Correction of Large Fuselage Components in Aircraft Assembly*. 2022. URL https://publica.fraunhofer.de/entities/publication/e1460f99-792f-4b6c-adab-7072d82b2926/details.

[26] C. Wang, D. Xu, Y. Zhu, R. Martín-Martín, C. Lu, L. Fei-Fei, and S. Savarese. Densefusion: 6d object pose estimation by iterative dense fusion. URL https://arxiv.org/pdf/1901.04780.

[27] Z. Wang, J. Fan, F. Jing, Z. Liu, and M. Tan. A pose estimation system based on deep neural network and icp registration for robotic spray painting application. *The International Journal of Advanced Manufacturing Technology*, 104(1-4):285–299, 2019. ISSN 0268-3768. doi: 10.1007/s00170-019-03901-0.

[28] Y. Xiang, T. Schmidt, V. Narayanan, and D. Fox. Posecnn: A convolutional neural network for 6d object pose estimation in cluttered scenes. URL https://arxiv.org/pdf/1711.00199.

[29] Zivid AS. Zivid two m70 technical specification. 04.2023.