

Visual Guidance for Assembly Processes

Julian Kreimeier*

Hannah Schieber[†]

Shiyu Li^{||}

Alejandro Martin-Gomez[§]

Daniel Roth[†]

Technical University of Munich
Clinic for Orthopedics and Sports
Orthopedics
TUM University Hospital
Munich, Germany^{*,†,||}

Johns Hopkins University
Computer Science Department
Baltimore, USA[§]

University of Arkansas
Department of Electrical Engineering
and Computer Science
Arkansas, USA[§]

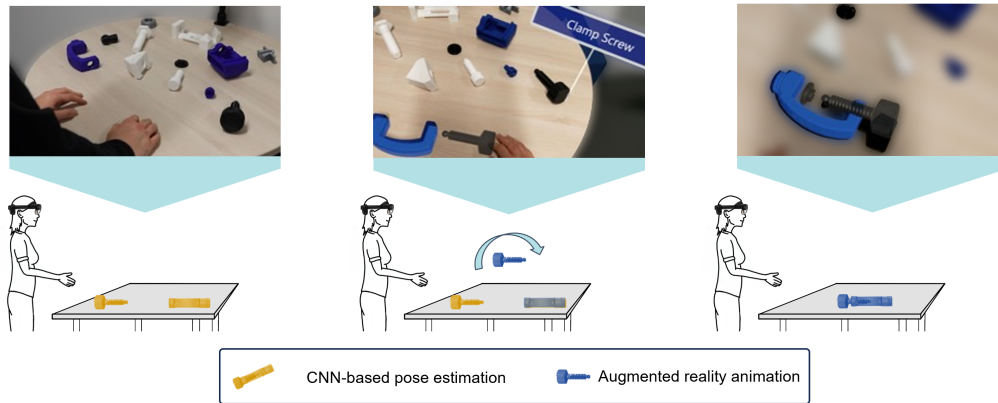


Figure 1: The Augmented Reality (AR) guidance application to support instrument assemblies.

ABSTRACT

Augmented reality (AR) can improve users' efficiency in various tasks. AR allows to guide a user with superimposed information, for example, during an assembly process. While paper-based assembly instructions are cumbersome and time-consuming, deep-learning driven AR-based instructions can dynamically adapt to the assembly scene and augmented 3D information over the physical objects (in-situ). We present KARVIMIO, an AR assembly guidance application for instruments based on 3D printed parts as reproducible testbed. Our approach utilizes purely synthetic training data for pose estimation to allow an easy generalization of the approach to new assembly groups and other areas of use.

Index Terms: Augmented reality, guidance, pose estimation, synthetic data

1 INTRODUCTION

Assembling complex objects in manufacturing, maintenance, or medical scenarios is challenging due to time pressure, high accuracy demands, and the need for prior device knowledge. For example, in medical scenarios, surgeons rely on a wide range of various surgical instruments. Despite a comprehensive medical education, knowing every assembly step of manifold and complex instrument sets is extremely challenging for surgical staff. AR guidance could be helpful in such situations, but requires markerless tracking for sterility [1]. Markers are robust and often used in AR applications, but not applicable in many industry or medical use cases due to a narrow field of view, sterile environment or changing tasks.

*e-mail: julian.kreimeier@tum.de

[†]e-mail: daniel.roth@tum.de

Previous research [1] has identified multiple requirements for i) the design (e.g., clear in presentation, visualizations in field of view), ii) the practicability (e.g., user autonomy, supporting effect, comfort), iii) the usage (e.g., ease of use, simplicity, no hand or speech interactions), iv) the situational fit (decision making-autonomy, limited cabling, responsibilities).

To address these requirements, our project builds upon the following key elements: a) synthetic data generation for training and generalization, b) markerless pose-estimation and/or tracking, c) status aware AR assembly guidance, d) gaze-based interaction.

2 RELATED WORK

Acquiring real-world accurate 6D object ground truth data in dynamic scenes is challenging and requires disproportionately extensive training. Kleinbeck et al. [3] proposed step-by-step guidance for building block assemblies while using synthetic data. Building upon 2D object detection, Stanescu et al. [8] implemented an assembly state detection based on YOLO. Schieber et al. [7] and Li et al. [4] addressed more complex and varying 3D printable assembly parts. Li et al. [4] focused on combining object tracking and 6D object pose estimation for re-initializing the tracking when occlusion eliminates the current tracking status. Schieber et al. [7] improved the YOLOv8-based 6D pose estimation by adding a late fusion step incorporating the current assembly state in the 6D pose estimation.

3 PROTOTYPE

Our approach provides in-situ guidance for assembling objects of various types, as illustrated in Fig. 1. Our system uses 6D pose estimation, allowing for more precise visual augmentations directly on the objects. We build on Yolov8Pose [4], an extension of Yolov8 [2] that integrates keypoints for 6D pose estimation. Yolov8Pose [4] utilizes the keypoints and employs the RANSAC PnP to determine the 6D pose of each object with an average runtime of 63.60 ms.

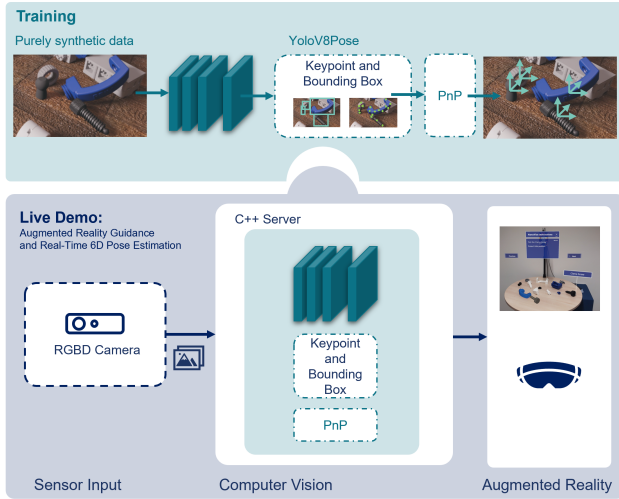


Figure 2: System setup. Our system is divided in a pure pretraining phase and the live augmented reality approach.

The network was trained using domain randomization [6] with synthetic data [4] from assembly sequences. Despite the system’s high precision and robustness for objects’ poses estimation, manual adjustments were made to align visually augmented content as accurately as possible with the actual position of the physical objects.

We use a Microsoft HoloLens 2 HMD and an AR application build with the Mixed Reality Toolkit in Unity. The assembly parts are tracked using a Microsoft Azure Kinect RGB-D camera which is connected to a laptop or workstation. We tested a workstation with an Intel i9-10980XE CPU, NVIDIA GeForce RTX 3090 GPU, and 128GB RAM and a Razer Laptop with a NVIDIA GeForce RTX 3070 Ti GPU. The YoloV8Pose model is executed on the GPU to enable markerless, real-time capable RGB-based 6D pose estimation. The HoloLens is calibrated with an optical marker relative to the Azure Kinect camera. Pose estimation data is transmitted to the HoloLens via WiFi, allowing it to render visual augmentations directly onto the real-world position of objects, as seen in Fig. 2.

4 DISCUSSION AND FUTURE WORK

Our approach utilizes synthetic data to train the 6D pose estimation, while tracking can also be a valuable approach for assembly groups [4]. Tracking is usually challenged by occlusion and requires a frequent re-initialization. Incorporating hand pose estimation with 6D pose estimation could significantly mitigate issues related to occlusion.

To enhance smart manuals we are constantly working on improving the pose estimation. Other aspects like state detection can further improve the approach [8, 7].

Additionally, developing automatic assembly state detection could facilitate context-aware visualization adjustments. For instance, visual elements might alter their transparency or rendering style (such as wireframe or silhouette) based on factors like the user’s distance from the object, hand occlusion, or the complexity of the current assembly step. Expanding multimodal interaction capabilities could further improve the efficiency of visualization approaches, providing additional perspectives to support the alignment task [5].

To promote broader application and innovation, we wish to empower the release of more open-source datasets (e.g., [7, 4]), which can serve as valuable testbeds for optimizing assembly guidance solutions in various research contexts.

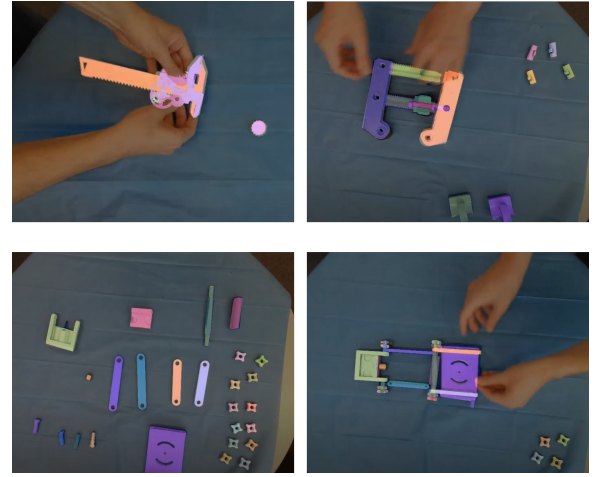


Figure 3: 6D pose estimation, each tracked part is highlighted by another color.

5 CONCLUSION

We propose an approach for AR assembly guidance utilizing synthetic data generation and 6D pose estimation to drive visual augmentations. Our approach allows for an efficient extension and adaption of the guidance system to future assembly parts purely based on their 3D models.

ACKNOWLEDGMENTS

This work was funded by the German Federal Ministry of Education and Research (BMBF) program ”AI-based assistance systems for in-process health applications”, grant number 16SV8973. We thank Medability GmbH for providing the 3D prints.

REFERENCES

- [1] E. Cramer, A. Kucharski, J. Kreimeier, S. Andreß, S. Li, C. Walk, F. Merkl, J. Högl, P. Wucherer, P. Stefan, et al. Requirement analysis for an ai-based ar assistance system for surgical tools in the operating room: stakeholder requirements and technical perspectives. *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–10, 2024. 1
- [2] G. Jocher, A. Chaurasia, and J. Qiu. YOLO by Ultralytics, 2023. Version Number: 8.0.0. 1
- [3] C. Kleinbeck, H. Schieber, S. Andress, C. Krautz, and D. Roth. Artfm: Augmented reality visualization of tool functionality manuals in operating rooms. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2022. 1
- [4] S. Li, H. Schieber, N. Corell, B. Egger, J. Kreimeier, and D. Roth. Gbot: graph-based 3d object tracking for augmented reality-assisted assembly guidance. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pp. 513–523. IEEE, 2024. 1, 2
- [5] A. Martin-Gomez, A. Winkler, K. Yu, D. Roth, U. Eck, and N. Navab. Augmented Mirrors. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 217–226, Nov. 2020. ISSN: 1554-7868. doi: 10.1109/ISMAR50242.2020.00045 2
- [6] H. Schieber, K. C. Demir, C. Kleinbeck, S. H. Yang, and D. Roth. Indoor synthetic data generation: A systematic review. *Computer Vision and Image Understanding*, p. 103907, 2024. 2
- [7] H. Schieber, S. Li, N. Corell, P. Beckerle, J. Kreimeier, and D. Roth. Asdf: Assembly state detection utilizing late fusion by integrating 6d pose estimation. *arXiv preprint arXiv:2403.16400*, 2024. 1, 2
- [8] A. Stanescu, P. Mohr, M. Kozinski, S. Mori, D. Schmalstieg, and D. Kalkofen. State-aware configuration detection for augmented reality step-by-step tutorials. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 157–166. IEEE, 2023. 1, 2