Research and Implementation on Key Technologies of Assembly Assistance System based on Augmented Reality

Jianhui Zhao¹, Xiaolin Mei², Jinsheng Ma³, Rui Wang²*, Ruyi Zhang³

¹ School of Instrumentation and Optoelectronic Engineering, Beihang University, Beijing, Beijing, 100191, China
² School of Mechanical Engineering and Automation, Beihang University, Beijing, Beijing, 100191, China
³ China Institute of Aeronautical Integrated Technology, Beijing, Beijing, 100028, China

*Corresponding author’s e-mail: tch_wangr@buaa.edu.cn

Abstract. Aiming at the demand of assembly assistance for complex industrial products, this paper studies the structure and key links of augmented reality aided assembly system, and takes a gear pump as an example to realize the parts detection function based on yolo neural network algorithm. On this basis, the assembly assistance system is built based on Vuforia, and the user interface is designed, realizing the functions of assembly progress identification and assembly information prompt.

1. Introduction
Augmented reality (AR) technology seamlessly integrates virtual scenes rendered by computer with those in the real world, and presents them to users through display devices. The emergence of AR technology makes human-computer interaction more natural, and has become a research hotspot in recent years[1].

Assembly is an important application field of AR technology. In the process of assembly or maintenance of industrial products, problems such as numerous assembly steps, complex assembly processes, and similar assembly parts are common, which not only affect the progress of disassembly and assembly, but also seriously affect product reliability due to operational errors. To solve the problems, AR aided assembly technology, or Augmented Assembly technology, has been developed, which uses augmented reality technology to assist the assembly process, that is, to superimpose virtual prompts such as 3D models, texts, and animations in the real assembly environment to help the operator complete the assembly operation. Moreover, it can improve assembly efficiency, assembly quality, and reduce assembly training costs. An augmented reality aided assembly system can provide workers with a virtual and real assembly environment, allowing operators to quickly and accurately obtain part information, assembly steps, and current assembly status, and timely know assembly skills, precautions, and whether the installation is correct. Compared with traditional assembly instruction manual, the system greatly reduces the cognitive burden and memory burden of operators, thereby significantly improving assembly efficiency and reducing assembly errors.

In recent years, there have been many domestic researches aimed at augmented reality assistance assembly. Liu Ran from Shanghai Jiaotong University took the assembly of automobile door parts as
an example to study the pose estimation in augmented reality assembly, and adopts different pose estimation methods for the different characteristics of the assembly base and parts during the assembly process[2]; Yin Xuyue et al. developed an integrated training system that includes assembly operation guidance and key component inspection and recording functions for the assembly operation scene of aerospace products[3]; Aiming at the maintenance operation of EMU, Li Hualing from Beijing Jiaotong University designed an augmented reality system based on TLD algorithm, which superimposed the maintenance operation information and three-dimensional model into the real operation scene for on-site operation assistance and maintenance personnel training[4]; Yang Kangkang et al. studied the key technologies of augmented reality for complex assembly, realized tracking registration based on LK optical flow method and ICP algorithm, and built an augmented reality guidance prototype system[5]; Yang Qing et al. studied the composition of augmented reality aided assembly system for large-scale complex equipment, and the registration technology of augmented reality in complex mobile scenes, to provide assembly guidance and operation training for large-scale complex equipment under real-time interaction[6].

Figure 1. The model of the pump and its parts

This paper takes a gear pump as an example to study the key process of the assembly assistance system based on augmented reality, and implements object detection based on deep learning algorithms. On this basis, a prototype of augmented reality aided assembly system is built based on Vuforia. The size of gear pump is: length=1180mm, height=923mm, width=850mm. The model of the pump and its parts are shown in Figure 1.

2. Assembly assistance system based on augmented reality
In order to build an augmented reality aided assembly system, the following functional modules need to be completed.

2.1. Image acquisition module
Both the collection of scene information and the superimposed display of prompt information are based on the acquisition of on-site images through certain equipment, which can be a camera fixed on the ground or that on a mobile phone or AR glasses. Meanwhile, it is also considerable to use infrared, laser, sound wave and other technologies to obtain depth information to improve the accuracy of subsequent object recognition. In this paper, the Intel Realsense D435i deep-sensing camera is chosen, which can output color maps or depth maps as needed. At present, only color maps are analysed.

2.2. Object detection module
In order to assist the product assembly process, the first step is to obtain the current assembly progress of the product, which can be judged by detecting the classes and locations of existing parts in the current scene, or identifying the subassembly under each installation step as an entirety. For systems that only uses RGB information, the object category and its pose information are generally identified by feature matching, template matching or deep-learning technology, while for those with additional depth information, the relevant information can also be obtained by point cloud recognition, which is more accurate.
2.3. Information prompt module

After obtaining the current assembly status, appropriate prompt information should be provided to the operator to complete the assembly operation. For example, the parts to be installed or disassembled in the next step, whether there are any part installation errors, the operation skills of assembling and disassembling, etc. This requires pre-arranging the spatial position of the prompt information relative to the assembly, calculating its projection on the screen and the occlusion relationship with the real object, etc., and finally putting it on the display after rendering.

Among them, the object detection module is the most critical and complex part to realize the assembly assistance system, which is studied and practiced below. The accuracy and speed of component detection in the scene determines whether the assembly prompt information can be provided to the operator correctly and timely.

3. Object detection technology

Traditional image-based target detection methods mainly include feature recognition and template matching. Among them, the object detection module is the most critical and complex part to realize the assembly assistance system, which is studied and practiced below. The accuracy and speed of component detection in the scene determines whether the assembly prompt information can be provided to the operator correctly and timely.

Traditional image-based target detection methods mainly include feature recognition and template matching. The feature recognition method uses feature extraction algorithms to extract and describe the feature points of the image to be recognized. After extracting the feature vector of the collected image, compare it with those of the prepared images in the database. If the feature similarity is higher than the predetermined threshold, it is determined that the matching is successful, and the best one is selected as the matching result. Commonly used algorithms include SIFT (Scale-invariant feature transform), SURF (Speeded Up Robust Features), FAST (Features From Accelerated Segment Test), Brief (Binary Robust Independent Elementary Features), ORB (Oriented Fast and Rotated Brief), etc[7]. Many researchers have also made some improvements and optimizations, or algorithm fusion, on the basis of these algorithms to improve the efficiency and accuracy in specific scenarios [8-10].

The template matching method is based on the grey intensity value of the image. The principle is to use the template image to slide on the input image, and use a specific matching algorithm to compare the feature points of the image to be recognized against the sub-region of the input image, to find out the image area that matches the template image at each position. This method requires the preparation of a large number of template images, and is often used in situations where the target pose is limited or known[11-12]. As for the template matching method in three-dimensional space, the feature sequence needs to be extracted with depth information, such as the Linemod algorithm for extracting the normal features of the surface of the object for matching[2], and the point cloud registration algorithm based on rough recognition[13].

In addition, deep learning algorithms are also used in target detection, which can be divided into two categories according to the detection process. One is to divide the entire detection process into two steps, firstly generate candidate frames, and then identify the objects in the frames, such as Faster-RCNN, Mask R-CNN and other algorithms; the other is to take the entire detection process as one step, directly giving the detection results including the object category and location, such as SSD, YOLO and other algorithms. In contrast, the former has higher accuracy and the latter has better real-time performance, but both are sufficient enough to meet the needs of industrial target detection. Based on these algorithms, some people made improvements on it to meet the needs of specific scenarios [14-16]. Recently, some researches have extended them to three-dimensional deep learning algorithms, such as SSD-6D[17], etc., to achieve end-to-end output of the object's spatial position and posture information.

The following takes the YOLOv5 algorithm as an example to realize parts detection.
4. Part inspection based on YOLOv5

YOLO algorithm, also called ‘You Only Look Once: Unified, Real-Time Object Detection’, means only one CNN operation is required to give a unified, real-time target detection result. The algorithm uses sliding windows of different sizes and aspect ratios, to slide across the entire picture in a certain step, and then the regions corresponding to these windows are classified to realize the detection of the entire picture.

In the following, the gear pump aforementioned is used as an example to build the data set and train the YOLOv5 model. There are 7 main parts of the assembly, which are the rear end cover, the pedestal, the keyless gear, the support shaft, the keyway gear, the drive shaft, and the front end cover, which are labelled from part1 to part7 respectively, and 3D printed. Use the camera to collect the images of each part in each angle of view, to total amount of about 300. Use the software Labelimg to label the data set, as shown in Figure 2, where deferent parts are framed with different colors.

![Figure 2. Data annotation](image)

The network structure model is yolov5-l, the number of training rounds is 300, and the Realsense camera calling program is written to achieve real-time images. It takes several hours to complete the training. The results of each training can be observed in the TensorBoard tool. Figure 3 shows the changes in indicators such as prediction loss, precision rate, and recall rate.

![Figure 3. Model indicators change while training](image)

As can be seen from the figure, the model converges continuously, and the precision and recall rates continue to rise. After 300 rounds of training, mAP0.5 has reached 0.96, achieving a high prediction accuracy. Then run the inspection program and aim the camera at the assembly scene. Figure 4 shows part of the detection results.

![Figure 4. Detection results (640x480, 30fps)](image)
It can be seen that the program has realized the detection and recognition of assembly parts well, but there are still some misrecognitions. It is mainly due to the similar shapes and textures of some parts, as the prints are all white coloured. Expanding training samples, increasing training rounds, and increasing model complexity are considerable methods to improve detection accuracy.

5. Augmented reality system based on Vuforia
Vuforia is a mainstream augmented reality SDK with multiple recognition functions, high recognition quality, and cross-platform compatibility, which supports most mobile phones, tablets and glasses, and is widely used in the development of augmented reality systems. In the following, an augmented reality aided assembly assistance system is built based on Vuforia and Unity.

Create a recognition target in Unity with reasonable prompts, configure the parameters, and run the project. As is shown in Figure 5, the prompt information is appropriately displayed, which means the status and position of the assembly is correctly identified.

![Figure 5. Test result](image)

The user interface is also designed in order to help the operator with the assembly and disassembly operations, as is shown in Figure 6. The current and expected assembly forms are displayed on the left side of the interface, and the progress of the total installation is displayed on the top, where the green background indicates the parts that have been installed, while the white background indicates those not installed yet.

![Figure 6. Running result](image)

6. Conclusion
Taking a gear pump as an example, this article carried out researches on assembly assistance system based on Augmented Reality for maintenance and assembly of complex industrial products. To build a practical assembly assistance system, the timeliness and accuracy of the object detection function must be ensured. The YOLO algorithm has been proved to have a good performance for the detection task, with mAP0.5 reaching 0.96 after 300 training rounds and calculation time per frame less than 0.04 second. Vuforia platform enables us to conveniently build an AR system. And to achieve the goal of assembly assistance, the user interface and prompt information must be reasonable designed. In practical use, the system can be installed on mobile devices such as AR glasses, to conveniently help the workers to complete the assembly operation efficiently.
However, the problem of confusion of similar parts still exists. By expanding the sample size, increasing the algorithm complexity or pasting additional markers on the parts, the recognition accuracy may be improved.

References
[1] Ying, H., Weiwei, X. (2017) Overview of Augmented Reality Technology. Computer Measurement and Control., 25(2):1 -7.
[2] Ran, L., Xiumin, F., Xuyue, Y., Wang Jikai, Xu, X. (2018) Combination estimation method of matrix and part pose in AR-assisted assembly. Mechanical Design and Research., 34(06): 119-125+137.
[3] Xuyue, Y., Xiumin, F., Lei, W., Kai, Q., Rui, R. (2018) Augmented reality guidance training system and application for aerospace product assembly operations. Aviation Manufacturing Technology., 61(Z1): 48-53.
[4] Hualing, L. (2017) Research on target tracking algorithm of EMU auxiliary maintenance system based on AR. (Doctoral dissertation, Beijing Jiaotong University).
[5] Kangkang, Y. (2019) Research and application of the key technology of augmented reality for complex product assembly. (Doctoral dissertation, Nanjing University of Aeronautics and Astronautics).
[6] Qing, Y., Xiaobin, W., Ali, W., Jiande, Z., Lizhong, T., Zhong, L. (2017) Research on the Intelligent Assembly Assistant System of Complex Products Based on Augmented Reality. Mechanical Design and Manufacturing Engineering., 46(11): 33-37.
[7] Wei, W., Pengbo, F., Zhengting, C., Zhaojuan, C. (2021) Overview of Augmented Reality Aided Assembly Technology. Packaging Engineering., 42(14): 108-118.
[8] Jianhui, X., Li, T. (2020) Workpiece recognition based on SURF improved algorithm. Manufacturing Technology and Machine Tool., 701(11):12-16.
[9] Radkowski R, Oliver J H. (2013) Natural feature tracking augmented reality for on-site assembly assistance systems. International Conference on Virtual, Augmented and Mixed Reality.
[10] Jizhuang, H., Li, L., Yongkui, Y., et al. (2016) Research on target recognition matching and localization method based on SURF-BRISK. Journal of Chang'an University: Natural Science Edition., 36(3): 93-101.
[11] Chao, C., Wangqiang, N., Peng, D. (2019) Container number recognition algorithm based on template matching and feature matching. Journal of Shanghai Maritime University., 40(01):68-73.
[12] Zhuping, f., Bo, z., Xi, w., et al. (2019) Multi-object visual recognition and classification system applied to machine assembly. Manufacturing Automation, 43(7): 5.
[13] Qi, C., R., Wei. L., Wu. C., et al. (2017) Frustum PointNets for 3D Object Detection from RGB-D Data. 2017.
[14] Kai. Z. (2021) Research on Recognition and Positioning of Industrial Parts Based on Artificial Neural Network. (Doctoral dissertation, Shanghai Normal University).
[15] Pengfei, Z., Zhiliang, S., Xiaoyao, L., Xiangbo, O. (2021) Classification and recognition algorithm of main bearing cover based on deep learning. Journal of Graphics., 42(04): 572-580.
[16] Shuanjun, S., Zhongyuan, H., Qiyu, W., Yiqi, N., Qianwei, H. (2021) Application of Improved YOLOV3 Algorithm in Part Recognition. Mechanical Science and Technology., 1-9.
[17] Kehl, W., Manhardt, F., Tombari, F., et al. (2017) SSD-6D: Making RGB-Based 3D Detection and 6D Pose Estimation Great Again. 2017 IEEE International Conference on Computer Vision (ICCV). IEEE.