Guided assembly using deep learning-based object recognition and augmented reality instructions

Mariana Zaninelo Reis, Prof. Dr. José Mario De Martino

{m223752@g.unicamp.br, martino@dca.fee.unicamp.br}

Departamento de Engenharia de Computação e Automação (DCA)
Faculdade de Engenharia Elétrica e de Computação (FEEC)
Universidade Estadual de Campinas (Unicamp)
Campinas, SP, Brasil

Abstract – Efficient assembly is vital to guarantee quality and celerity in manufacturing processes. Usually, the assembly tasks in a non-robotic process are expressed by assembly instructions that provide text and drawing information, both in 2D and 3D, about the form and order of the execution of tasks. However, these instructions are rarely interactive and, depending on their complexity, hard to follow. The main objective of this master's work is to develop an approach capable of providing the user with interactive step-by-step guidance for assembly tasks. Deep learning techniques will be applied to recognize, in real-time, equipment parts, indicate the assembly order, and the coupling to other parts. The assembly instructions will be provided exploring augmented reality. We plan to implement the solution on a mobile device and assess and compare our solution with more traditional ones, considering assembly speed-up and perceptual feedback from users.

Keywords – Augmented Reality, Deep Learning, Guided Assembly, Object Recognition.

1. Introduction

Industrial production lines are usually automated for most of their activities. However, some specific assembly processes are still predominantly performed manually. For these tasks, execution guides are used that can be printed or made available on support computers.

These assembly guides have technical drawings, assembly instructions, as well as related processes that must be executed, such as measurements and verification of aspects that guarantee the quality of the product.

Depending on the level of complexity of the assembly, these guides are not always enough for the assemblers to perform their tasks in a simple way. The main reason for this is the fact that they are not interactive and do not allow any kind of verification of the activities they are carrying out.

New ways of providing these assembly instructions have been proposed recently. Among the main motivations are the intention to improve the user experience and the performance indices in the manufacturing lines, such as task execution time and assertiveness level, in addition to guaranteeing the quality of the operations being performed.

The use of augmented reality combined with real-time object recognition is an area of study that can help improve assembly instructions.

Augmented reality offers an interactive experience of a real-world environment enhanced by computer-generated images and information. In an augmented reality environment, the user does not lose his natural perceptions, which is essential in its use in an industrial setting. Augmented reality applications are already widely applied or studied in some areas, such as games, medical applications, industrial applications, product design and development, assembly lines and navigation.

The application of augmented reality to promote guided assembly can be beneficial in terms of enabling user interactions. Superimposing synthetic images on real objects and providing assembly instructions that promote greater interaction has proven to be favorable in carrying out complex tasks.

There are already studies showing gains over conventional methods of assembly instructions [2]. Benefits such as reduced assembly time and reduced level of difficulty in performing tasks can be observed. However, some aspects need to be analyzed in the use of augmented reality, such as the level of familiarization and acceptance of the technology by users, the type of device that will be used and its implications, in addition to the costs and resources for implementation.

There are two ways to launch augmented reality experiences, with or without the use of markers. Applications that use markers, usually QR codes or some image to help position the synthetic imagery, are more common. However, their use is not always viable, depending on the scenario where the montage is carried out and the availability of space to adequately place the markers.

Markerless augmented reality experiences have been more discussed lately. However, depending on the application, it is necessary to reference or superimpose computer-generated images on real objects [1]. The combination of object recognition techniques has been explored in conjunction with augmented reality applications without the use of markers so that computer-generated images and information can be referenced.

Object recognition can improve the positioning of synthetic images [4], making the augmented reality experience more realistic. In guided assembly, object recognition can improve the interactive experience and help to guide the activities performed.

The evolution of algorithms based on deep learning for object recognition tasks made it possible to apply them in more complex activities and to obtain greater levels of assertiveness [5, 6]. Currently, there are models based on deep learning that can be used on mobile devices, covering their application area.

The use of neural networks, specifically deep learning, for object recognition allows state-of-art recognition of complex assembly parts.

2. Proposal

The main objective of this work is to develop a guided assembly solution that provides the user with instructions to perform tasks in an interactive way. Algorithms based on deep learning will be used to recognize equipment parts to be assembled, providing the correct assembly sequence in real-time. Augmented reality is the technology that will be responsible for relaying assembly instructions to the user.

A mini bench vise will be used as a case study. The bench vise, printed in a 3D printer, is shown in Figure 1. The reason for choosing this object is that it is composed of some parts, which will allow the study of the use of the proposed solution in tasks that require the execution of some steps.

To use algorithms based on deep learning for object recognition, it is necessary to create a dataset that will be used for training the neural network. The dataset will consist of photos of each piece, varying their dispositions and the background in which they are found.

Building a dataset with many samples, specifically in this case, where it is necessary to have photos of the parts, is not an easy task. Neural networks, in general, require a significant amount of data for their training. Therefore, we intend to use image augmentation techniques to improve the performance of the neural network.

The creation of new architectures is not foreseen in this work, the intention is to use existing models that are already known to be efficient for this type of task. Some architectures will be analyzed to define the one that best meets the needs of the project.

The use of augmented reality associated with the work proposal occurs because it allows interaction with the user and facilitates complex work instructions [7, 8].

A mobile device will be used to carry out the proof of concept, a cell phone or tablet. The intention is to place it on a support in front of the activity, so that the user can view the information available and have their hands free to carry out the assembly.

Figure 1: mini bench vise used as study object

As a way of analyzing the results, we intend to compare the assembly performed using conventional tasks guides and the proposed solution. The parameters of comparison are the execution time of the activity and the level of difficulty to perform the assembly. For the analysis of the level of difficulty of the task, perceptive evaluations of the users will be necessary.

If the hypothesis initially raised is true, it is expected that there will be a decrease in the execution time of the activity, as well as in the level of difficulty for execution.

3. Results

Regarding the dataset, the photos that will initially compose it have already been taken. The images were divided into ten classes, nine of which correspond to individual pieces and one that includes multiple pieces in the same images. Each class has three hundred images, totaling a dataset with three thousand samples.

The images have already been separated and stored in folders related to their respective classes, which depending on the type of neural network to be used, is sufficient to be used as a form of labeling. In addition, the entire dataset was labeled using bounding boxes, aiming to test neural networks that use this type of annotation, such as, for example, the YOLO [3, 9, 10] network. An example of samples that are present in each of the ten classes in the dataset is illustrated in Figure 2.

Figure 2: dataset samples

4. Conclusions

No conclusions have yet been reached about the proposed work, however, it is expected that there will be a decrease in the time of the mapped assembly task, in relation to traditional methods of instructions, as well as an improvement in the user experience during execution.

It is also expected that the work will contribute to the development of augmented reality, as it will explore a scenario of interest in several areas, which is the association of real-time object recognition with markerless applications.

References

[1] AKGUL, O.; PENEKLI, H. I.; GENC, Y. **Applying Deep Learning in Augmented Reality Tracking**. 2016 12th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS). **Anais**... Em: 2016 12TH INTERNATIONAL CONFERENCE ON SIGNAL-IMAGE TECHNOLOGY & INTERNET-BASED SYSTEMS (SITIS). nov. 2016.

[2] ALVES, J. et al. **Comparing Spatial and Mobile Augmented Reality for Guiding Assembling Procedures with Task Validation**. 2019 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC). **Anais**... Em: 2019 IEEE INTERNATIONAL CONFERENCE ON AUTONOMOUS ROBOT SYSTEMS AND COMPETITIONS (ICARSC). abr. 2019.

[3] BAHRI, H.; KRČMAŘÍK, D.; KOČÍ, J. **Accurate Object Detection System on HoloLens Using YOLO Algorithm**. . Em: 2019 INTERNATIONAL CONFERENCE ON CONTROL, ARTIFICIAL INTELLIGENCE, ROBOTICS & OPTIMIZATION (ICCAIRO). IEEE Computer Society, 1 maio 2019. Disponível em: <https://www.computer.org/csdl/proceedings-article/iccairo/2019/357200a219/1iQ32yocCDm>. Acesso em: 9 ago. 2022

[4] GARON, M.; LALONDE, J.-F. Deep 6-DOF Tracking. **IEEE Transactions on Visualization and Computer Graphics**, v. 23, n. 11, p. 2410–2418, nov. 2017.

[5] LALONDE, J.-F. **Deep Learning for Augmented Reality**. 2018 17th Workshop on Information Optics (WIO). **Anais**... Em: 2018 17TH WORKSHOP ON INFORMATION OPTICS (WIO). jul. 2018.

[6] LI, W. et al. Integrated Registration and Occlusion Handling Based on Deep Learning for Augmented Reality Assisted Assembly Instruction. **IEEE Transactions on Industrial Informatics**, p. 1–11, 2022.

[7] NISHIHARA, A.; OKAMOTO, J. **Object recognition in assembly assisted by augmented reality system**. 2015 SAI Intelligent Systems Conference (IntelliSys). **Anais**... Em: 2015 SAI INTELLIGENT SYSTEMS CONFERENCE (INTELLISYS). nov. 2015.

[8] PARK, K.-B. et al. Deep learning-based smart task assistance in wearable augmented reality. **Robotics and Computer-Integrated Manufacturing**, v. 63, p. 101887, 1 jun. 2020.

[9] REDMON, J. et al. **You Only Look Once: Unified, Realtime Object Detection**. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). **Anais**... Em: 2016 IEEE CONFERENCE ON COMPUTER VISION AND PATTERN RECOGNITION (CVPR). jun. 2016.

[10] REDMON, J.; FARHADI, A. **YOLOv3: An Incremental Improvement**. arXiv, , 8 abr. 2018. Disponível em: <http://arxiv.org/abs/1804.02767>. Acesso em: 9 ago. 2022