# Multimodal Support for Industrial Assembly

Henrik Björklund[1], Johanna Björklund[1,2], and Mona Forsman[2]

[1] Umeå University, {henrikb,johanna}@cs.umu.se
[2] Codemill AB, mona.forsman@codemill.se

**Abstract.** We report on the on-going development of a support system that will assist assembly workers in the manufacturing industry. The system will guide users through the various assembly steps, verify correct execution, and curate data for use in production management.

## 1 Introduction

Process automation is a dominating trend in industrial assembly, where increasingly sophisticated robots take on previously manual tasks. It is particularly relevant for manufacturers that produce large series of uniform goods, such as makers of passenger vehicles and home appliances. On the opposite end of the scale, we find manufacturers of small series of highly customized products. Here, manual assembly is expected to remain more cost efficient than its robotic counterpart for many years to come. Process automation can still bring value, but the developed technology must be designed to assist human assembly workers, rather than replace them.

This abstract describes an on-going project together with Komatsu Forest[3] to develop an assembly support (AS) system for manual assembly. Komatsu Forest is the world's second largest supplier of forestry machines such as harvesters, loggers, and planters. Their products are bespoke, allowing the customer to decide on virtually every aspect of the machine. The total number of components in a single harvester typically exceeds 10 000. The large number of components in combination with the great variation in the construction means that the likelyhood of assembly errors is high. If an error is discovered after the product has been deployed, it can cause long down times and great expense. The AS system is intended to reduce the risk for erroneous mountings, and increase the quality of the produced machines.

The system uses a range of multimodal analysis techniques to guide human operators through the assembly steps, detect deviations and errors, and collect and analyse data that can help improve the product design and the over-arching manufacturing process. The AS system is built on the subsystems listed below. After each, we specify its its current technological readiness level (TRL) [1].

- A computer vision (CV) system that uses a camera rig to analyse the intermediate product (TRL 5)

[3] https://www.komatsuforest.com

– An augmented-reality (AR) system that shows the assembly worker a live video of the intermediate product, onto which a CAD model is projected, together with instructions for the next set of assembly steps (TRL 4)
– A speech transcription system (TRL 5), that is on the one hand used to direct the system, and on the other to enter spoken deviation reports.
– An ML/CV analytics system (TRL 3) that identifies components and verifies that the right component is attached to each mounting point. The correct set of matchings is taken from the product specification, and the system can be trained to visually recognize new components.
– a data aggregation system that indexes and anonymises data (TRL 2), and includes it in the product's digital twin. A digital twin is essentially a record of key aspects that is updated throughout the products life-time, and that assists in everything from maintenance to product refinement [3].

In summary, the CV and AR systems are furthest developed. Less has been done on the language processing system, but this is also the area where the team has greatest prior experience. The anonymisation and data aggregation systems are still in the conceptualisation and design phase.

There has been substantial interest in support systems for use in the manufacturing industry during the last decade. For two recent surveys, see [5, 2]. The use of multimodal analysis has recently been investigate in the case of human-guided robot assembly by Wan et al. [4]. The authors consider a two-phase solution. In the first phase, the human operator demonstrates the assembly, and in the second phase, the robot detects geometrical objects and manipulates these to follow the operator's instruction. To realise accurate 3D vision, the system uses AR markers in the training phase, and point clouds in the execution phase.

In our work, the roles are reversed, and it is the autonomous system that guides the human assembly worker. This means that the system must be able to match the geometry of the target structure against the partially assembled product, to detect possible errors and misconfigurations, and to help the worker get back on track when this happens. As the system is meant to guide human workers and not robots, it is essential that the workers are in control of the system: The system should adapt to their preferred style of working, rather than the other way around.

## 2 Solution approach

Figure 1 sketches the practical realisation of the AS system. A camera rig mounted over the work station provides a live stream of the assembly process. The video feed is displayed with an overlay showing a 3D model of the product on a wall-mounted video screen. Computer vision is used to detect the product and calculate the projection of the 3D model; see Figure 2. As the worker selects a component to assemble, the system shows possible positions on the product in the video, together with assembly instructions. Assembled components are recorded, together with images of the final product for later quality assessment and documentation.

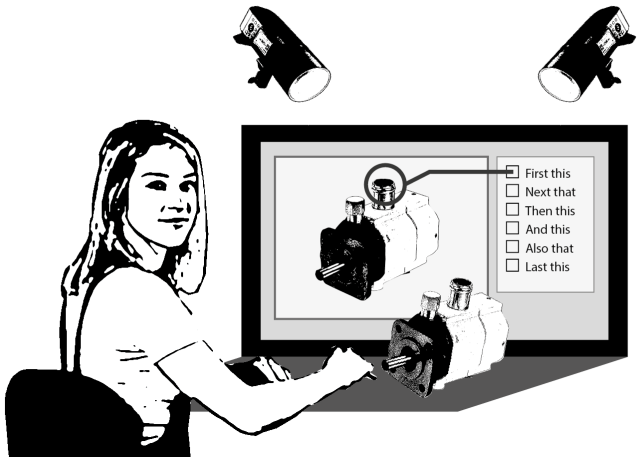


**Fig. 1.** A sketch of a work station with the AS system. A camera rig streams video of the product to a screen mounted behind the work station. In the video, a 3D model of the product is projected, with markings of where to attach a selected component.

An important part of the project concerns improving the quality through the analysis of problems in the assembly process. The most readily available and useful data on such problems are the *deviation reports* filed by the assembly workers. Such reports can be filed for a number of different reasons, ranging from faulty or missing components to mistakes made earlier in the assembly process. To date, filing a report requires a certain amount of effort. The worker has to put down what he or she is doing and move to a computer or tablet to fill out the report form. The result is that only serious deviations are reported, not minor problems and suggestions for improvements. In order to encourage the workers to file more such reports, the project develops a voice controlled interface, including automatic speech recognition and transliteration.

A second source for data on problems is the vision system itself. How often does it, for example, have to notify an assembly worker about incorrect mounting of a specific component? The purpose of collecting data is of course to analyse it to come to conclusions about possible improvements. In the project, we aim at analysis that can be used both as feedback to the assembly workers and as support for the production management in their ongoing work with the streamlining of the assembly process. We use machine learning techniques to develop analysis models. As a basis for training, we take the about 45 000 deviation reports Komatsu has collected over the last few years. Each report contains a, usually brief, natural language comment by the assembly worker, as well as some metadata. The latter indicates, among other things, which machine was being assembled, what specific component the deviation concerns, date, time, and so forth.

The collected reports have been manually annotated and separated into classes by production management staff. We can thus use supervised learning

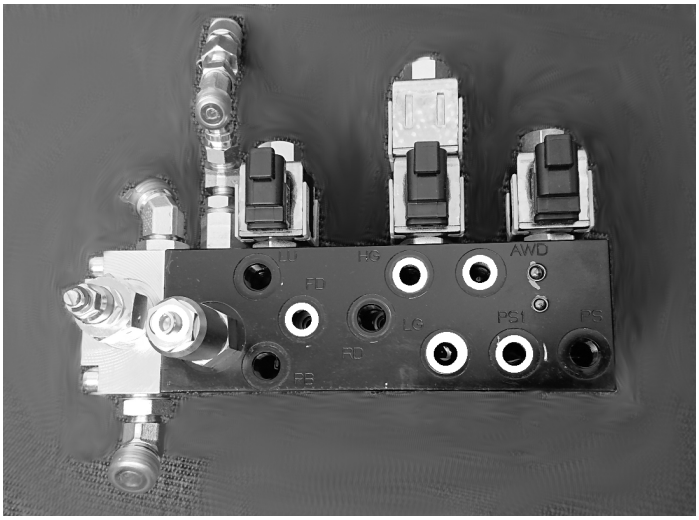


**Fig. 2.** An image of an intermediate product with automatically detected fixation points marked out. These marks are also used for orientation of the 3D model of the object with the video view.

to train classification models, such as support vector machines and neural networks. These models will then be used to classify future reports, initially under the supervision of human staff. In addition to this traditional classification, we also use unsupervised learning techniques such as topic modelling and clustering on the deviation data. The purpose of this is to try to find new classes that have not been considered in the manual identification of deviation types. If significant classes are found in this manner, they can point production management to bottlenecks or systematic problems in the production process.

The goal is not to analyse employees, so all data is anonymised before it is used, and the trade unions are consulted on all data management issues.

## References

1. Annex G of the Horizon 2020 work programme. Technical report, The European Commission, 2016.
2. E. Bottani and G. Vignali. Augmented reality technology in the manufacturing industry: A review of the last decade. *IISE Transactions*, 51(3):284–310, 2019.
3. F. Tao, J. Cheng, Q. Qi, M. Zhang, H. Zhang, and F. Sui. Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 94(9-12):3563–3576, 2018.
4. W. Wan, F. Lu, Z. Wu, and K. Harada. Teaching robots to do object assembly using multi-modal 3D vision. *Neurocomputing*, 259:85 – 93, 2017.
5. X. Wang, S. K. Ong, and A. Y. C. Nee. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing*, 4(1):1–22, Mar 2016.