Using Activity Recognition for the Tracking of Assembly Processes: Challenges and Requirements

Mario Aehnelt[1], Enrico Gutzeit[2], Bodo Urban[3]

Fraunhofer IGD
Joachim-Jungius-Str. 11
18059 Rostock, Germany
{[1]mario.aehnelt, [2]enrico.gutzeit, [3]bodo.urban} @igd-r.fraunhofer.de

Abstract: Activity recognition helps to improve the quality of assistance applications by enabling adaptive and purposeful user support. However, industrial applications require a far more robust recognition in order to ensure process reliable operations. The paper highlights specific challenges and requirements with respect to activity recognition from the perspective of assembly workplaces in manufacturing. A task-oriented assembly model is used to derive human activities based on sensorial observations of work equipment and material.

1 Introduction

In manufacturing up to 40 percent of costs and even 70 percent of production time fall upon the assembly of intermediary components and final products [LW12]. Especially when producing in small lot sizes or in case of individualized items, we predominantly find manual work operations and a small degree of automation. Work assistance will help significantly under these circumstances to reduce both production time and costs. However, assistance applications as we can find them in several different work domains require in manufacturing a very robust and reliable recognition of physical work activities as well as situations. Recognition failures would lead to wrong assistance and thus to assembly mistakes, which finally compromise quality and functionality of the product to be assembled.

Known approaches of activity recognition work well under laboratory conditions. But they normally find their limits in rough industrial environments, as we face them in manufacturing. Additionally, industrial safety and privacy concerns do not allow the usage of body-worn sensors or similar technologies which enable the recognition of human activities. For this reason we need to investigate and apply alternative approaches.

This paper introduces the assembly of complex machines and technical systems as challenging application domain which demands situation aware information assistance for the worker. We consider work context and assembly tasks in more detail in order to derive a model of human activities and their relationship with work equipment and material. Finally, we formulate requirements for a robust and reliable activity recognition under manufacturing work conditions.

2 The assembly application scenario

The assembly of machines and technical systems is an essential part of production. It joins single machine parts to first-order assemblies (*pre-assembly*), them to assemblies of higher order (*intermediate assembly*) and finally to an end product (*final assembly*). The German VDI guideline [VDI90] further differentiates between *primary assembly*, which includes the main joining operations as defined with the DIN 8580 [DIN03], and *secondary assembly*, including all additional assembly activities like handling, adjustment and control.

Although, the assembly work process is principally characterized by physical work activities, it also requires a not unimportant amount of cognitive work in order to interpret and understand work orders as well as work detail information [AU14]. These cognitive processing takes place during the work preparation and again after executing required work steps in order to review work results in comparison to plans and expectations.

Assembly workplaces can be situated within manufacturing buildings, outside or at distant sites, where the product is assembled and built into larger machines or technical systems (e.g. cooling units into a ship section). In all cases we are facing changing and rough industrial work conditions with respect to light, noise, temperature, dirt or vibrations, especially in extreme work situations.

3 Chances and limitations of activity recognition

Human activity recognition allows the purposeful support of work activities. The better the quality as well as granularity of identified activities, the more precise assistance applications can provide required help and information. In our assembly application scenario activity recognition helps to:

- *identify* changing work situations which vice versa require adequate assistance functionalities,
- *differentiate* single assembly steps in order to provide only relevant information details for a specific step,
- *track* the overall assembly progress to allow estimation and control of time figures for manufacturing resource planning, and to
- *observe* the compliance with process oriented work regulations.

These examples also show how normal human work activities relate to an implicit interaction with manufacturing support systems in general and assembly assistance applications in particular. Interacting implicitly through the automated recognition of activities and their meaning for the work process is more natural and less interruptive in comparison to an explicit interaction with graphical user interfaces [Sc00]. A growing amount of research addresses the challenge of finding new technologies, methodologies and approaches to enable such ways of implicit human-computer interaction. The research can be divided into computer vision-based approaches or approaches based on alternative sensorial input.

Computer vision based approaches propose the recognition of human activities and gestures by using 2D and/or 3D image input devices. This is a large and still growing research field with many subfields [MA07]. Depending on the input device (single-, stereo or multi camera, time of flight (ToF), structured light), the location (indoor, outdoor), the kind of gestures (head, hand and/or body, static and/or dynamic) the concepts and algorithms are very different. In general a typical gesture recognition approach consists of a data preparation, segmentation/object detection, feature extraction and classification part. Additionally, in some cases a tracking [YJS06] part is needed to permanently know the object of interest location (e.g. hand or head). Tracking can be also included for other purposes, e.g. for tracking skin cluster in color space instead of objects in image space [GVZ11]. The data preparation part involves image correction, registration and/or 3D reconstruction. An example for pose estimation of faces in 3D can be found in [GI07] and for body gestures with ToF in [GP11]. Due to the difference and special conditions for head, face and body recognition a vision based interaction framework including all types in a stable way is seldom addressed. Instead scientific papers mainly addressing special gestures e.g. hand gestures [AJK13], [GVZ11] or head pose estimation [MT09].

Alternative approaches [BHC06], [KHS10] specifically focus on embedding the interaction unobtrusively in people's environments experimenting with situation and context-aware devices such as kitchen utilities, wearable sensors, capacitive touch devices on clothes, or RFID enhanced objects. However, leaving the controlled environment as we find it in a laboratory and heading towards an industrial application will dramatically reduce the recognition accuracy. Here it is still common practice to work with explicit interactions and feedback.

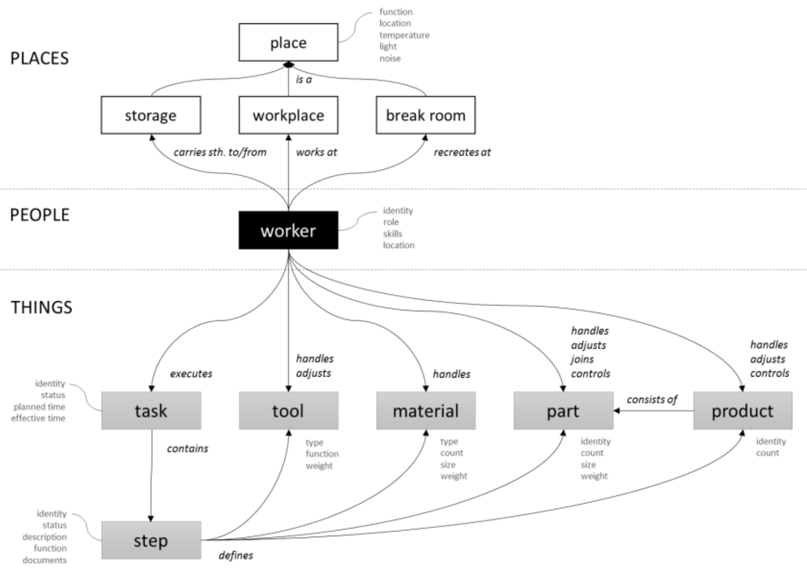

Figure 1: Assembly context model with relevant places, people and things

4 Assembly context and activities

In section 2 we introduced our assembly application scenario. But it requires a far more detailed consideration of context and activities at assembly workplaces to understand the relationship between work environment, equipment and material with human activities. Following the suggestion of Dey in [DAS01] we divide the assembly context into *places*, *people*, and *things*. Each category characterizes associated contextual attributes and behavior which have an influence on the work situation (see Figure 1). Thus, we can separate typical work environments such as the assembly workplace, a storage, or the break room for example. Primarily their function but also their location hints to activities the worker is involved in. We further can see a very close relationship between a specific work task (e.g. assembly of machine case) and materials (screws), tools (screwdriver), as well as parts (pre-assembled case parts) to be used. The sequence of work steps defines not only this relationship. It also describes assembly activities which are required to join parts by using tools and material. Vice versa the tools and material used by the worker give us a direction on which assembly activity is executed next. In [BA14] we showed how probabilistic models can be used to reason assembly sequences from the material taken by the worker.

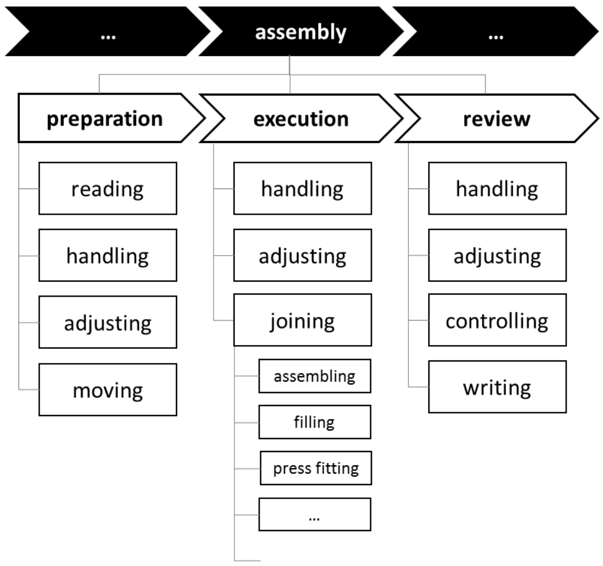

Figure 2: Assembly activities based on VDI 2860 and DIN 8580

In addition to our previously described context model, we require an activity model which simplifies occurring human activities during an assembly (see Figure 2). Here we can work with the assembly method definitions from [VDI90] and [DIN03].

As the result assembly methods directly refer to physical activities of the worker. Thus, a typical joining method is *press fitting* which further includes screwing, clamping, clipping, nailing, bolting, wedging or brace tensing to join two machine parts. Using an activity recognition which differentiates between these very fine activities would allow a close tracking of assembly processes. [HLH12] showed for example the tracking of screw driving using acceleration sensors in wearable wrist watches. However, industrial safety can avoid the usage of similar sensors at assembly workplaces which leads us back to the instrumentation of work equipment and environment instead of tracking activities with body-worn sensors.

5 Industrial requirements for activity recognition

As shown in the previous sections, the industrial application of activity recognition, for example at assembly workplaces, requires robust and reliable technologies which identify human activities even at small granularities. Summarizing our observations in this specific application domain, we come to following general requirements:

- *Modelling and recognition of different activity granularities* – Human activities in manufacturing are interesting for the tracking of single work processes to provide required assistance, or for monitoring the overall work performance for example. Depending on such goals the activity granularities to recognize can differ dramatically. In one case the step by step observation and guidance requires a very fine recognition of single movements or environment changes, in other cases it is sufficient to track the beginning and ending of work tasks only. This leads to hierarchical activity granularities as shown in section 4 which need to be modeled and finally identified by the activity recognition.

- *Plausibility of recognized activities* – Activity recognition in manufacturing has the chance to integrate with manufacturing planning and execution systems, including systems which collect and interpret a rich set of sensorial and other data from workplaces, machines and logistical events. This allows a close comparison of observed activities with planned work tasks as well as provided work progress information and thus an increase of the data plausibility. Additionally, it is important to keep a required degree of freedom for plan variations which still result in valid work processes.

- *Reliable recognition and fallback strategies* – Industrial environments do not ease the application of technologies which strongly depend on the uniformity of environmental conditions. The activity recognition needs to work stable even in extreme work situation. Further, it needs to identify failures and inaccuracy in time. Then feedback from the worker needs to be requested in order to inform about as well as to avoid manufacturing failures in consequence of possibly critical or even wrong recognitions.

- *Consideration of industrial safety and privacy* – Observing the workers' activities makes possible estimations not only on his work progress but also his performance, and possible issues related to it. It can be understood as tool to quantify and compare the value of single workers which normally not complies with legal regulations. In a similar way, the activity recognition needs to consider issues related to the industrial safety of manufacturing operations. Thus, it is often not possible to directly equip the worker with sensors. It should be proved which opportunities the work environment and equipment enables here. It can make more sense, to use sensors on work tools and material that give more specific hints on work activities.

In general, industrial environments require similar to medical applications a far more robust implementation of activity recognition, but in parallel there are already rich data sources and manufacturing systems which help to improve the recognition quality.

6 Conclusions and future work

The manufacturing application domain is a demanding environment for activity recognition. However, using it can enable novel assistance opportunities which in consequence help to improve work performances and reduce production times and costs. Especially, the assembly part of manufacturing offers a wide field for research and a reality testbed for laboratory algorithms and technologies. Although, work conditions can vary there very much, there is the chance to integrate and work with already available data collecting and interpreting systems in order to recognize human activities on different granularity levels.

More research needs to be invested into analyzing the relationships between manufacturing workflows, environmental events and real human activities at the workplace. First steps are done with modeling and detecting valid assembly sequences from product models [BA14]. It now requires the recognition of single work activities from events of the work environment and equipment which allows a very close observation and guidance of work activities on work step level. Here will help us the clear assembly activity definitions available by industrial standards and the adequate instrumentation of related work tools and material. Together with the domain dependent background knowledge it will be used to stabilize the recognition.

References

[AJK13] Ankit Chaudhary; Jagdish Lal Raheja; Karen Das; Sonia Raheja: Intelligent Approaches to interact with Machines using Hand Gesture Recognition in Natural way: A Survey. In CoRR, 2013, abs/1303.2292.

[AU14] Aehnelt, M.; Urban, B.: Follow-me: Smartwatch Assistance on the Shop Floor: Proceedings of the 16th International Conference on Human-Computer Interaction (HCII). 22 - 27 June 2014, Crete, Greece. Springer-Verlag, 2014.

[BA14] Bader, S.; Aehnelt, M.: Tracking Assembly Processes and Providing Assistance in Smart Factories: Proceedings of the 6th International Conference on Agents and Artificial Intelligence (ICAART). 6 - 8 March 2014, Anger, France, 2014.

[BHC06] Bravo, J.; Hervás, R.; Chavira, G.; Nava, S.: Mosaics of Visualization: An Approach to Embedded Interaction Through Identification Process. In (Hutchison, D. et al. Eds.): Cooperative Design, Visualization, and Engineering.

Springer Berlin Heidelberg, Berlin, Heidelberg, 2006; pp. 41–48.

[DAS01] Dey, A. K.; Abowd, G. D.; Salber, D.: A Conceptual Framework and a Toolkit for Supporting the Rapid Prototyping of Context-Aware Applications. In Human-Computer Interaction, 2001, 16; pp. 97–166.

[DIN03] DIN Deutsches Institut für Normung e.V.: DIN 8580 Manufacturing processes - Terms and definitions, division. Beuth-Verlag, Berlin, 2003.

[GI07] Gurbuz, S.; Inoue, N.: Real-Time Head Pose Estimation using Reconstructed 3D Face Data from Stereo Image Pair: Acoustics, Speech and Signal Processing, 2007. ICASSP 2007. IEEE International Conference on, 2007; pp. II-785-II-788.

[GP11] Gonzalez-Sanchez, T.; Puig, D.: Real-time body gesture recognition using depth camera. In Electronics Letters, 2011, 47; pp. 697–698.

[GVZ11] Gutzeit, E.; Vahl, M.; Zhou, Z.; Lukas, U. von: Skin cluster tracking and verification for hand gesture recognition. In ISPA 2011 - 7th International Symposium on Image and Signal Processing and Analysis, 2011.

[HLH12] Hein, A.; Low, T.; Hensch, M.; Kirste, T.; Nürnberger, A.: Gesture Spotting for Controlling a Mobile Assistance System for Service and Maintenance. In (Goltz, U. et al. Eds.): GI-Jahrestagung. GI, 2012; pp. 549–560.

[KHS10] Kranz, M.; Holleis, P.; Schmidt, A.: Embedded Interaction: Interacting with the Internet of Things. In IEEE Internet Computing, 2010, 14; pp. 46–53.

[LW12] Lotter, B.; Wiendahl, H.-P.: Montage in der industriellen Produktion. Ein Handbuch für die Praxis. In Montage in der industriellen Produktion, 2012.

[MA07] Mitra, S.; Acharya, T.: Gesture Recognition: A Survey. In Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on, 2007, 37; pp. 311–324.

[MT09] Murphy-Chutorian, E.; Trivedi, M. M.: Head Pose Estimation in Computer Vision: A Survey. In Pattern Analysis and

Machine Intelligence, IEEE Transactions on, 2009, 31; pp. 607–626.

[Sc00] Schmidt, A.: Implicit human computer interaction through context. In Personal Technologies, 2000, 4; pp. 191–199.

[VDI90] German Engineers' Association: VDI 2860:1990-05 Assembly and handling; handling functions, handling units; terminology, definitions and symbols. Beuth-Verlag, Berlin, 1990.

[YJS06] Yilmaz, A.; Javed, O.; Shah, M.: Object tracking: A survey. In ACM Comput. Surv., 2006, 38.