Technology-Enhanced Process Elicitation of Worker Activities in Manufacturing

Sönke Knoch, Shreeraman Ponpathirkoottam, Peter Fettke, and Peter Loos

German Research Center for Artificial Intelligence (DFKI) and Saarland University
Stuhlsatzenhausweg 3, 66123 Saarbrücken, Germany

Abstract. The analysis of manufacturing processes through process mining requires meaningful log data. Regarding worker activities, this data is either sparse or costly to gather. The primary objective of this paper is the implementation and evaluation of a system that detects, monitors and logs such worker activities and generates meaningful event logs. The system is light-weight regarding its setup and convenient for instrumenting assembly workstations in job shop manufacturing for temporary observations. In a study, twelve participants assembled two different product variants in a laboratory setting. The sensor events were compared to video annotations. The optical detection of grasping material by RGB cameras delivered a Median F-score of 0.83. The RGB+D depth camera delivered only a Median F-score of 0.56 due to occlusion. The implemented activity detection proofs the concept of process elicitation and prepares process mining. In future studies we will optimize the sensor setting and focus on anomaly detection.

Keywords: Process elicitation, activity recognition, manufacturing.

1 Introduction

Workers retain flexibility in semi-automated assembly systems becoming more and more complex. Especially during unforeseen occurrences or incidents a large degree of flexibility during assembly is required, e.g., for small and medium-sized companies [5]. Production planners need accurate information about the assembly process to make realistic assumptions during the plant design or in the operational phase balancing work load between production lines. The manual work in manufacturing processes is analyzed using predetermined motion time systems, such as Methods-Time Measurement (MTM), cf. mtm-international.org, and REFA, cf. refa.de. In MTM, a person documents all motions in assembly tasks under different plant settings and looks up the standard time for relevant motions in the MTM catalogue. In REFA, organization-specific catalogues are created timing each motion with a stop watch.

Process mining from Business Process Management (BPM) bridges the gap between data and process science [1]. Process mining is grounded on meaningful log data. Especially regarding manual activities in BPM this data is sparse. Thus, we suggest a process elicitation system that tracks and pre-processes manual activities during assembly processes at workstations in job shop manufacturing. Today, the elicitation of worker motions in manufacturing requires an expensive manual procedure. Activity recognition in combination with existing sensors has the potential to increase efficiency and effectiveness during this process by delivering detailed information about manual activities. Sensor data is pre-processed and filtered by applying Complex Event Processing (CEP), efficiently handling huge amounts of heterogeneous data. Operative work steps are integrated into process models in compliance with the standards of formal modeling and enabling process monitoring during execution time. This new knowledge helps to understand how the work plans, process models defining assembly processes created during the product's industrial engineering, are executed on the shop floor. It delivers input for process discovery, conformance checking and enhancement to foster the detection of anomalies and to uncover optimization capabilities.

Following the design-science research approach, a light-weight artifact was developed addressing the problem of process elicitation in manual assembly facing a heterogeneous sensor setting, which has not been addressed by current work that is mainly focused on log data of existing software systems. The design's efficacy was evaluated, which proofs that an easy and fast instrumentation of common assembly workstations can be applied to enable temporary monitoring tasks connecting the business process to an Internet-of-Things in Industrie 4.0 factories. To integrate the suggested sensors with existing software, machines and sensors, an event-driven architecture couples state-of-the-art modules loosely keeping the system extremely flexible. Raw events of multiple sensors are aggregated and combined in event rules to detect patterns using CEP. The system was set up in a laboratory setting and enables activity tracking during the assembly of a realistic artificial product by fusing information from four sensor applications together with events from a worker guidance system and mapping them to basic motions to break down the underlying process activity. In an experiment with 12 participants the optical detection of grasping material was analyzed. Methods and metrics from the activity recognition domain were applied to measure the artifact's performance proving the concept of sensor-driven process elicitation.

2 Problem Statement

In process engines, manual activities are usually black boxes documenting human tasks which are beyond the process engine's reach. A system that couples sensor data with such activities would enable *process monitoring* and the following three types of *process mining*. To position them, we use the *BPM life cycle* by [10] who introduced a new and comprehensive life cycle concept partitioned into the phases (1) strategy development, (2) definition/modeling, (3) implementation, (4) execution, (5) monitoring/controlling, and (6) optimization/improvement.

Process discovery can be applied when no a-priori information is available, typically during work plan creation in the definition and modeling phase (2). Production engineers and workers test best practices to assemble a new product variant to create the work plan. An instrumented prototypical assembly workstation, cf. Sect. 3, can deliver event logs during that test phase and supports the discovery of detailed work plans/assembly processes. In manufacturing, discovery would be applied during the elicitation of the process while conducting observations, measurements and workshops with workers, team leaders and engineers. Since the main high-level process is already known during the execution phase (4), discovery could reveal hidden sub-processes.

Conformance checking compares an existing work plan/process model against the generated event logs (5). It addresses the detection of anomalies and deviations by matching reality against existing models. In manufacturing, the detection of assembly faults is one of the major applications. Assembly faults lead to longer process execution times compared to the scheduled times or reduce the number of accurate pieces compared to the output that was planned.

Process enhancement can be used to improve a work plan/process model (6) using knowledge contained in event logs. If an optimization potential (best practices or anti pattern) is mined, the work plan is adapted. For example, the order or duration of assembly tasks is adjusted. In the worst case, a complete redesign of the process model could become necessary.

To enable these three types of process mining the generation of meaningful log data is essential. Thus, a system is needed that tracks progress of process instances consisting of activity instances that occur during a concrete case. The case id is equivalent to the product id that is read through automatic identification and data capture. If no case id is available activities associated to resources such as assembly workstations, tools, and materials are tracked. During the product assembly a set of `Activities` is performed which are associated with certain `Resources` observed by `Sensors`. Activities can be composed of multiple sub-activities, e.g., start/stop and have additional properties, e.g., duration or performed left/right handed. The sensing of worker activities, its domain specific requirements and the filtering of events with convenient rules enhances the potential of process mining in manufacturing benefitting from new Internet-of-Things infrastructures in Industrie 4.0 factories.

3 Solution within a Laboratory Setting

Within the BPM life cycle, our system is intended to run during execution (4) as a monitoring application (5), when work plans are executed in assembly lines connecting multiple assembly workstations on the shop floor. Additionally, it can be applied in early prototypes during the design of assembly workstations and work plans (2). During the monitoring of a work plan execution, the assembly of a concrete product, the process instance, is observed. The outcome delivers input for the design, optimization and improvement (6) of assembly workstations, work plans and assembly processes in further life cycle iterations.

The suggested activity detection and event pre-processing is an event log generating software system in process mining, cf. Fig. 1. It observes and senses worker activities filtered by a CEP engine and generates new event logs. These

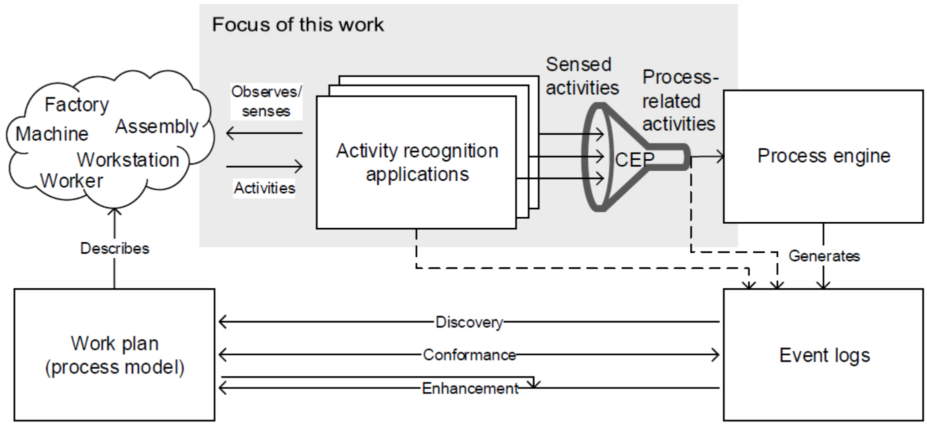


Fig. 1. Activity recognition in the context of process mining (extension of [1]).

events can be integrated in business processes executed in an engine running processes defined a-priori and formally described in a standardized representation such as the Business Process Model and Notation (BPMN).

3.1 Requirements and Setup

For the artifact, as stated in [11], a *fast setup* and the use of *light hardware* was one aspect to enable a fast instrumentation of the assembly workstation for temporary tracking. In addition, the worker's *degree of freedom* during work is not restricted, which excludes heavy sensor technology and on-body motion sensors. To address medium-sized and small companies, the equipment *costs should be limited* acquiring and operating such a system. Finally, only tracking technology *appropriate for manufacturing* environments is considered.

Implementing the artifact, these basic aspects were addressed by using affordable and light sensor hardware. These sensors (see below) do not afford any direct contact with the worker (not restricting the degree of freedom) and work appropriately for manufacturing (usually stable light conditions). Additionally, a flexible event-based architecture was chosen, which allows a flexible integration and removal of sensor components depending on the desired tracking case. Regarding the software that was developed to configure the sensor applications a fast setup to run the optical tracking components even by non-experts is supported. A simple user interface allows the drawing, labeling and configuring of activity zones to observe areas in the view of the respective camera, cf. [11].

The assembly workstation is made from carton prototyping material and instrumented with low-cost sensors from the consumer electronics domain: 3 RGB (top: Logitech C920 HD Pro, bottom: 2x Creative Live! Cam Voice), 1 RGB+Depth (Microsoft Kinect 2), 1 infrared (Leap Motion) cameras and 2 ultrasonic (GHI HC-SR04) sensors (cf. Fig. 2). The event-bus was realized using MQTT. All events are logged to a database for process elicitation and mining.

In the current setting two product variants can be assembled: The bill of material of variant “BG” and “BCD” consists of seven and six different materials, see table in Fig. 2. Three materials are variant-specific resulting in a total number of ten materials available in small load carriers (SLC) at the assembly workstation. Both products consist of two 3D-printed cases filled with three interconnected printed circuit boards fixed with screws while the number of screws varies per variant (“BG”: 8, “BCD”: 4). The only tool available during assembly is a manual torsional screw driver. To assemble a product, the workpiece holder in front of the worker is fed with the variant-specific Top_Casing part (Task 1). Next, the Mainboard is inserted into the Top_Casing (2) and fixed with two Small_Screws (3). Afterwards, the Application_Board is inserted (4) and the Connecting_Board that has to be connected to Mainboard and Application_Board (5). Once that is finished, the Application_Board is fixed with two Small_Screws (6). Finally, the Bottom_Casing is fastened together with the Top_Casing with clips (7) or four Big_Screws in case of variant BG (7+8). The products are artificial, meaning that they have no purpose or function.

3.2 Work Step Events: Composition, Format and Pattern Detection

Work Step Composition. A work step is split up into five fundamental motions that were adapted from the MTM used to analyze the performance of manual operations. MTM is based on empirically gathered data aggregated in time catalogues that focus on activities which are 100% influenceable. The motion cycle covers the five motions *reach* (move an empty hand to a thing), *grasp* (bring a thing under control), *move* (move a thing by hand), *position* (fit things into each other), and *release* (intended loose control of a thing).

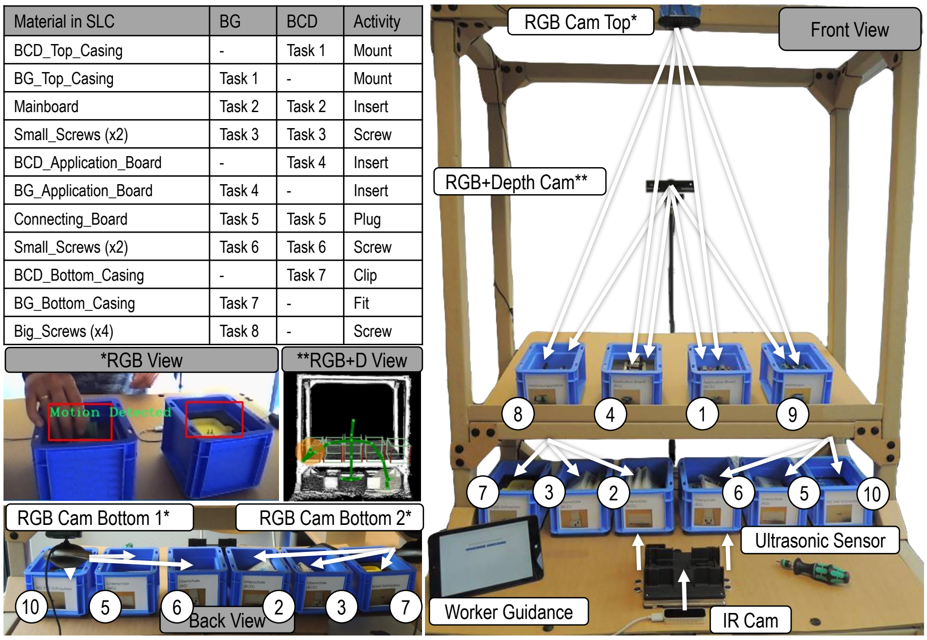
Message Format. The event generating sensor applications are identified by topics that clients can subscribe to communicating over a publish/subscribe protocol, such as MQTT. Topics are simple text strings hierarchically structured with forward slashes, e.g., `Sensor/Resource/Activity`. The `Sensor` contains the name of the data delivering sensor application, e.g., RGB+D, and the `Resource` denotes the observed area at the assembly workstation containing a tool or material, e.g., screws. The `Activity` indicates the actual action, e.g., In or Out, if someone reached in or out of the SLC with screws. This allows a device independent subscription to sensor events using the single- or multi-level wildcard character (+/#), e.g., *subscribeTo(*`#/Screws`*)*. The event’s payload is en-/decoded to/from an internal format within the processing engine.

Event Pattern. Two types of rules were applied to match event patterns: patterns consisting of bounding events and patterns consisting of multiple events from the same type. For the first, a rule waits for the start and end event limiting, e.g. the grasping of the material screws. It checks if two events `e1` and `e2` arrive in a successive order without any event `e3` between and within a time window of carefully defined `T` seconds (20 in our current setting). The exclusion of `e3`, having `e3.topic=e2.topic`, ensures accurate detection when the worker grasps twice successively in the same SLC within `T`. The positioning of a material in the assembly area represents the second type of patterns, since its boundaries are not strict. An aggregation query processes a series of events and fires every time an additional event from the same type occurs. Once a pattern has been detected, each query calculates the time difference between relevant events and sends a high-level event about the detected operation to the event bus.

3.3 Process Elicitation through Activity Detection

Four different sensor applications were implemented. The *Local Activity Detection* with RGB web cameras detects activities within activity zones defined by the user. The competing *Local Activity Detection and Skeleton Tracking* with RGB+D camera tracks the skeleton and activities in three-dimensional rectangular areas wrapping SLCs on a digital representation of the assembly workstation. It delivers information about the body posture, location and the hand that is used. The *Hand Detection* with IR camera senses whether a hand entered or left an a-priori defined area. The *Motion Detection* recognizes motions based on an ultrasonic distance measurement, similar to a light barrier. In the following, we will focus on the Local Activity modules optically sensing material grasps.

Local Activity Detection with RGB Cameras. Activity detection can be achieved by performing background subtraction and foreground extraction using RGB web cameras. Therefore, the cameras are arranged so that their field of view observes the SLCs. Usually, there is minimal/no motion in the observation regions other than motion of hands during the assembly process. However, the system is highly sensitive to changing background intensity and effects of unavoidable artifacts like shadows and repeated adding/removing of objects (material). The system should be robust against such effects and able to adapt continuously and learn the presence or absence of objects in the scene as a change in the background and not indicate activities in these regions. Here we adopt a time tested method involving fast adapting mixture of Gaussians by [14]. The values that each pixel can take is modelled as a mixture of Gaussians. An advantage of the method is that it chooses the number of Gaussians for each pixel automatically and independently of other pixels not wasting memory as in the case of a pre-defined number of Gaussians. For each pixel, weights are assigned to the Gaussians modelling it. In a particular pixel, if the weights are concentrated in a few Gaussians, a stable state is achieved thereby indicating inactivity. If the weights are distributed over a lot of Gaussians this indicates activity in the specific pixel. The parameters of the algorithm are adjusted in a way that the Gaussians modelling the background adapt quickly enough to ignore the added/removed tool but slowly enough to detect the movement of the hand in and out of the activity regions. The activity regions are rectangles defined by the user. If more than half the pixels in this region has been identified as foreground by the algorithm, an activity is detected.



Material in SLC	BG	BCD	Activity
BCD_Top_Casing	-	Task 1	Mount
BG_Top_Casing	Task 1	-	Mount
Mainboard	Task 2	Task 2	Insert
Small_Screws (x2)	Task 3	Task 3	Screw
BCD_Application_Board	-	Task 4	Insert
BG_Application_Board	Task 4	-	Insert
Connecting_Board	Task 5	Task 5	Plug
Small_Screws (x2)	Task 6	Task 6	Screw
BCD_Bottom_Casing	-	Task 7	Clip
BG_Bottom_Casing	Task 7	-	Fit
Big_Screws (x4)	Task 8	-	Screw

Fig. 2. The instrumented assembly workstation made from carton prototyping material in front and back view. Arrows indicate the sensor perspectives. Camera perspectives are shown on the mid left. The numbers from 1 to 10 refer to the IDs of Tab. 1. On the top left, the assembly tasks of both product variants are provided.

Local Activity Detection with RGB+D Cameras. The Kinect sensor performs skeleton tracking as well as activity detection by forming a two and a half dimensional (2.5D) model of the assembly workstation. The 2.5D model is obtained by using the Kinect fusion functionality, which is a part of the Kinect SDK. The model is shown from the camera's perspective. Since a user's perspective is more intuitive, during setup Kinect is positioned on the user side in front of the workstation and moved and rotated through 180 degrees around the workstation to its final position. This process is the geometric transformation between the two diagonally opposite positions of the Kinect sensor (in front of the assembly workstation and behind). Once the final position has been fixed, the vectors normal to the worktop surface and two mutually orthogonal vectors along the worktop surface are determined using three user-defined points on the surface in the model. It is assumed that the Kinect's X axis and the table surface are in parallel. Kinect's skeletal tracking provides a total of 25 joints which delivers the necessary data for visualizing and analyzing motions of the worker. The hand joints are used for testing the overlap with activity regions. To allow tolerance to joint position accuracy as given by the Kinect, a sphere of a 25cm radius (identified empirically) surrounds the hand position. The activity in an activity region (cuboid in 3D space) is triggered when the sphere around a given hand joint overlaps with its cuboid, see Kinect view in Fig. 2. The condition for overlap: the projection of the vector joining the centre of the sphere and the centre of the cuboid along any of the three cuboid dimensions is less than the sum of the radius of the sphere and half of the respective dimension.

4 Evaluation

In the following, we provide insights into a first study analyzing the performance of the optical sensor applications using RGB and RGB+D cameras. The events generated by these applications enable the use of CEP and the elicitation of operative work steps mapped to activities in formal process models.

4.1 Experimental Setup and Data Analysis

The suggested system was evaluated with 12 participants consisting of students and campus staff. To evaluate the system, an artificial but realistic chain of assembly tasks had to be solved, cf. Sect. 3.1. Each participant had to assemble four artificial products, two of each variant. The participants were split into two counter-balanced groups. One group started with the variant "BG", the other with "BCD". The four products were assembled in an alternating order. The instructions about the assembly process were provided on a simple worker guidance system (WGS), a web page shown on an eight inch tablet. Text, images and videos were used to explain the different assembly steps. Difficulties in understanding were answered by a supervisor who was present during the whole experiment. In addition, the WGS provided temporal boundaries for relevant sensor events by delivering an event each time a new assembly step was started. Therefore, the user had to confirm each step which is common practice in industrial manufacturing. Regarding the assembly, apart from a correctly assembled product, no limitations were given which lead to a large degree of freedom making the recognition of activities more realistic while exacerbating it.

To measure the system's precision, the ground truth (GT) was captured filming the assembly process with the consent of the participant. Three persons manually annotated all videos with a given set of tags. One tag represented one GT activity and consisted of the name of the observed material in the SLC and the start and stop time taken from the video. The timestamps of video and sensor events had to be synchronized to allow a comparison of the GT against the system's event log. Each system event between two instructions from the WGS was matched to a corresponding GT entry based on the material referenced in event and GT. This lead to a bipartite graph with two independent sets of vertices (GT entries and system events) such that every edge connects a vertex in the one set (GT) to one in the other (system). To find the best matching system event for a given GT entry, the optimum matching had to be calculated. This represents the assignment problem of finding the minimum weight matching in a weighted bipartite graph. The weight of an edge connecting a vertex representing a GT entry and a vertex representing a system event is

Table 1. Results of activity detection per SLC containing a certain material.

ID	SLC/material (anomalies)	Sensor	TP	FN	FP	Precision	Recall	F-score	Clustered
1	BCD_Application_Board	RGB	32	4	10	0.76	0.89	0.82	-
	(50%)	RGB+D	28	8	40	0.41	0.78	0.54	7
2	BCD_Bottom_Casing(37%)	RGB	26	9	16	0.62	0.74	0.68	-
3	BCD_Top_Casing (39%)	RGB	32	1	6	0.84	0.97	0.90	-
4	BG_Application_Board	RGB	34	7	2	0.94	0.83	0.88	-
	(54%)	RGB+D	26	15	23	0.53	0.63	0.58	7
5	BG_Bottom_Casing (35%)	RGB	24	2	9	0.72	0.92	0.81	-
6	BG_Top_Casing (29%)	RGB	27	1	10	0.73	0.96	0.83	-
7	Big_Screws (14%)	RGB	131	40	24	0.85	0.77	0.80	-
8	Connecting_Board (20%)	RGB	46	4	10	0.82	0.92	0.87	-
		RGB+D	30	20	39	0.44	0.6	0.50	8
9	Mainbaord (15%)	RGB	54	0	26	0.68	1.00	0.81	-
		RGB+D	45	9	14	0.76	0.83	0.80	27
10	Small_Scres (12%)	RGB	65	2	8	0.89	0.97	0.93	-

the time distance. After construction, we received a simple undirected weighted bipartite graph. JGraphT, cf. jgrapht.org, was used to compute the maximum weight matching in $O(V|E|^2)$, where E is the set of edges and V the set of vertices.

4.2 Results

Tab. 1 shows the results of the experiment in numbers. For each SLC carrying a certain material the results of the three RGB cameras observing all SLCs and the results of the RGB+D camera observing the four upper SLCs true positives (TP), false negatives (FN) and false positives (FP) in absolute numbers such as precision, recall and the F-score are provided. In the last column the number of events clustered when the material was equal within a time distance of T reducing the number of insertions (FP) are listed. Anomalies (percentage provided in brackets under the respective SLC/material name) represent a deviation from the instructions, such as picking into the wrong SLC or moving the hand over several SLCs while searching the correct material, e.g., when they look alike.

The RGB cam application delivered satisfying results with an F-score higher than 0.8 for seven SLCs and higher than 0.9 for two. Only the observation of the SLC containing the material "BCD_Bottom_Casing" was with F=0.68 low. Having a look at the physical setup during the experiment as shown in Fig. 2, it can be seen that the angle of the camera to the SLC is adverse. Another restriction evolved by the chosen setup is the shock resistance. Physical shocks caused by participants by bumping against a material box or the whole assembly workstation led to FPs. A similar effect was discovered by arm/hand shadows generating activities in adjacent SLCs. Although both did not occur very often, it can be addressed in industrial settings by using a solid assembly workstation, putting SLCs on tracks and illuminating the workplace from top. Finally, the precision can be improved by using industrial cameras where the area of interest is observed with a fixed focus setting delivering a higher resolution.

Analyzing the results of the RGB+D camera, it was discovered that the optical sensor delivered a large amount of insertions (FP). Compared to the web cam which generates exactly two events per activity (hand in and hand out), a strong flickering was discovered regarding the event generation of the RGB+D sensor. Thus, we clustered sensor events by removing all events with a time difference of T compared to a matching event of the same type (RGB+D sensor and material) from the set of insertions (FP). We set $T = D(G^{v,a})$ where D is the duration of the a^{th} ground truth action in the v^{th} video. Nevertheless, it can be seen that the F-score of the RGB+D sensor is low. One reason is the high number of anomalies for the SLCs "BCD_" (50%) and "BG_Application_Board" (54%). The other reason is the position of the RGB+D sensor behind the assembly workstation. Large parts of the worker's body are covered by the workstation which leads to a less accurate skeleton tracking that we are using to estimate the hand position. This fact is generating a lot of flickering which is the reason for a high number of insertions (FP) and failures (FN) during activity detection.

5 Related Work

In the pervasive computing community Funk et al. [7] suggest a cognitive assistance system that aims to increase efficiency and assistance in manufacturing processes based on motion detection with an RGB+D camera. A permanent and calibration intensive integration of the sensor is necessary to provide feedback in contrast to our approach that aims to provide a temporary and light-weight sensor setup. Instead of using additional instrumentation, Bader et al. [3] use existing sensors (RFIDs) to provide assistance on a display. RFID could be integrated in our system through CEP but is alone insufficient to provide details on worker activities. Quint et al. [13] suggest a system architecture for assistance in manual tasks with the aim to combine components, such as visualization techniques, interaction modalities and sensor technologies. Sensor events are matched from RGB and RGB+D cameras to states in state machines. Compared to our approach the emphasis lies on the information model and the architecture more than on studying the artifact's accuracy and application in BPM.

In the augmented reality domain several systems realize activity tracking to provide assistance during assembly tasks in the psychomotor phase of a workflow. Henderson and Feiner's prototype [8] provides several forms of assistance visualizing arrows, labels, highlighting effects and motion paths using markers attached to all tracked objects. In contrast, Peterson et al. [12] focus on the technical realization of markerless spatiotemporal tracking with uncalibrated cameras to automate the creation of augmented reality video manuals from a single first-person view video example. Both papers feature a strong emphasis on augmented reality and do not touch the BPM world. Nevertheless, the suggested technologies provide the potential of analyzing worker activities.

Data fusion, which is necessary to integrate heterogeneous sensor information and application events, matches the concept of CEP and is supported by our even-driven architecture, similar to Bruns et al. [4], who integrate CEP and finite state machines fusing sensor data to determine the actual states of ambulance vehicles and not the overall process. The interconnection between BPM and CEP was introduced as event-driven BPM by von Ammon et al. [2] and demonstrated, e.g., by Estruch et al. [6] who suggest an approach for CEP with BPMN 2.0 in the manufacturing domain. They focus on the modelling aspects in a top-down manner and leave out the sensor-level. Not tackling the manufacturing domain, Herzberg et al. [9] provide a framework to correlate specific process instances by enriching events already recorded with context data resulting in process events and show its feasibility applying process mining in the logistics domain. They focus on the event correlation and leave out the data elicitation.

Summarized, work on activity tracking is mainly focused on assistance using existing or additional sensor technology. Work interconnecting BPM and CEP is sparse in the manufacturing domain. The technology-enhanced process elicitation suggested here is related to both worlds and aims to deliver detailed information about worker activities, which has not been addressed so far.

6 Conclusion

In this work, an artifact was implemented and evaluated that supports process elicitation tasks in manual assembly systems. The system was set up in a laboratory setting and showed a satisfying performance in activity tracking during the assembly of an artificial product regarding the RGB camera application to proof the concept of process elicitation. The competing RGB+D camera application did not satisfy the performance criteria and will in future only be applied to deliver additional information on the worker (e.g., left or right handed). The experiment showed that the number of false negatives and false positives of the RGB activity detection still needs to be reduced. The evaluation of relevant factors was influenced by the high degree of freedom during product assembly and the laboratory setting within the experiment. Especially carton prototyping allows the construction of a testing environment very fast, but has the drawback of being shock sensitive. Similarly, consumer electronics are low in price but less stable regarding the result. In future, we plan additional experiments with more participants in multiple scenarios controlling and analyzing these side effects to increase the precision. It will be combined with the detection of material parts completing the picture of tracking manual assembly tasks. To support process enhancement and operational support, we will examine the anomalies detected during the study and follow the question how the manufacturing process and its model can be reflected and improved as a basis for a cyber-physical BPM.

Acknowledgments. This research was funded in part by the German Federal Ministry of Education and Research (BMBF) under grant number 01IS16022E (project BaSys4.0). The responsibility for this publication lies with the authors.

References

[1] W. M. P. van der Aalst. *Process Mining: Data Science in Action*. 2nd ed. Heidelberg: Springer, 2016.

[2] R. von Ammon et al. "Integrating Complex Events for Collaborating and Dynamically Changing Business Processes". In: *7th International Conference on Service-oriented Computing*. ICSOC. Stockholm, Sweden: Springer, 2009, pp. 370–384.

[3] S. Bader et al. "Tracking Assembly Processes and Providing Assistance in Smart Factories". In: *6th International Conference on Agents and Artificial Intelligence*. Vol. 1. ICAART. Angers, France: SCITEPRESS, Science and Technology Publications, Lda, 2014, pp. 161–168.

[4] R. Bruns et al. "Using Complex Event Processing to Support Data Fusion for Ambulance Coordination". In: *17th International Conference on Information Fusion*. FUSION. July 2014, pp. 1–7.

[5] K. Dencker et al. "Proactive Assembly Systems - Realising the Potential of Human Collaboration with Automation". In: *Annual Reviews in Control* 33 (2 2009), pp. 230–237.

[6] A. Estruch et al. "Event-Driven Manufacturing Process Management Approach". In: *10th International Conference on Business Process Management*. BPM. Tallinn, Estonia: Springer, 2012, pp. 120–133.

[7] M. Funk et al. "Cognitive Assistance in the Workplace". In: *Pervasive Computing* 14.3 (July 2015), pp. 53–55.

[8] S. J. Henderson et al. "Augmented Reality in the Psychomotor Phase of a Procedural Task". In: *10th International Symposium on Mixed and Augmented Reality*. ISMAR. IEEE, Oct. 2011, pp. 191–200.

[9] N. Herzberg et al. "An Event Processing Platform for Business Process Management". In: *17th International Enterprise Distributed Object Computing Conference*. EDOC. IEEE, Sept. 2013, pp. 107–116.

[10] C. Houy et al. "Empirical Research in Business Process Management - Analysis of an emerging field of research". In: *Business Process Management Journal* 16.4 (2010), pp. 619–661.

[11] S. Knoch et al. "Automatic Capturing and Analysis of Manual Manufacturing Processes with Minimal Setup Effort". In: *International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp. Heidelberg, Germany: ACM, Sept. 2016, pp. 305–308.

[12] N. Petersen et al. "Real-Time Modeling and Tracking Manual Workflows from First-Person Vision". In: *International Symposium on Mixed and Augmented Reality*. ISMAR. IEEE, Oct. 2013, pp. 117–124.

[13] F. Quint et al. "A System Architecture for Assistance in Manual Tasks". In: *Intelligent Environments. IE*. Vol. 21. Ambient Intelligence and Smart Environments. IOS Press, 2016, pp. 43–52.

[14] Z. Zivkovic et al. "Efficient Adaptive Density Estimation Per Image Pixel for the Task of Background Subtraction". In: *Pattern Recognition Letters* 27.7 (May 2006), pp. 773–780.