CoTeSys Progress Report 2008
ACIPE – Adaptive Cognitive Interaction in Production Environments

Prof. Dr. rer. nat H. Bubb, Prof. Dr. H. Müller, Dr. A. Schubö
Prof. Dr.-Ing. habil. G. Rigoll, Dr.-Ing. F. Wallhoff and Prof. Dr.-Ing. M. F. Zäh

I. Project overview

Focusing on human manual workplaces in the Cognitive Factory, the ACIPE project aims at creating an assistive system, which is able to give workers required information in intuitive ways at the right time and enables an ergonomic interaction of a worker and her/his environment.

Traditional systems for digital assistance in manual assembly are inherently suboptimal for providing ergonomic worker guidance as part of an efficient assembly process. The display of sequential instructions does not offer an increase in productivity beyond a certain degree. Little situational support and the resulting deterministic guidance lead to a reduced acceptance by the worker.

A solution is seen in the development of concepts and technical implementations, which will allow for the adaptive generation of assembly instructions. Adaptive in this context means the integration of factors of the production environment as well as factors regarding the human worker. Therefore, algorithms for dynamic work plan generation are developed. Furthermore, sensing technologies for online observation of the human are investigated and experiments for gaining knowledge about human cognitive processes are conducted.

II. Completed work

In the progress report of 2007, a first draft of the architecture was presented. Based on information flow analysis, this has been refined and a system architecture was deducted. It is introduced in Section II-A. Technical realizations are presented in Section II-B, followed by performed Experiments in Section III-C. The Section closes with a listing of work fulfilled in cooperations with other CoTeSys-Projects.

A. System Analysis

Figure 1 shows the layout of the system architecture. On one side, the process model defines tasks to be completed, and on the other side, the worker performs activities, aiming at task completion. The overall system supplies the worker with situation- and worker-adaptive assistive information, aiming at ergonomic worker integration and increase of efficiency, effectiveness and quality of the assembly process.

The system consists of seven components, described in the following:

1) Personal Cognitive Assistant (PCA): Core of the system is the Personal Cognitive Assistant (PCA). It receives feedback from and about the worker from the input layer and selects assistive information to be presented to the worker by the output layer. It provides the process model with situational feedback, e.g. the desired step-size, and in return receives abstract up-to-date assembly paths. These are turned into actual assistive information by text, images, video or audio from the multi-modal instruction database. In order to adapt this output to the worker, a cognitive worker model is maintained, the worker's assembly history is recorded, and the worker's preferences are derived and stored. To estimate the workers cognitive and physiological state, his/her performance is compared with experimental data contained in the generic database of human cognition and behavior. The PCA will be developed on base of the approaches described in [3], [4].

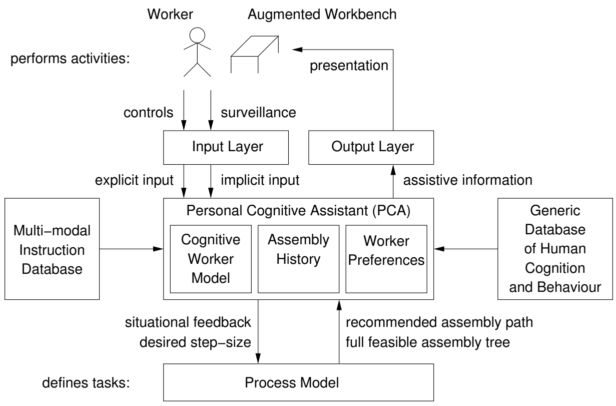

Fig. 1. System Architecture

2) Process Model: The process model performs a mapping of the product state and dynamically determines the full feasible assembly tree and recommended assembly path for the desired step-size (Section II-B.1).

3) Generic Database of Human Cognition and Behaviour: The generic database of human cognition and behaviour contains results from experiments, as described in Section II-C: e.g. error rates, dwell and completion times, MTM, motion and eye tracking data.

4) Multi-modal Instruction Database: The multi-modal instruction database contains actual instructions in form of text, images, videos and audio, which are used to enhance the abstract instructions of the process model.

5) Input Layer: The input layer generates events for explicit and implicit user input. Explicit input is generated by controls operated by the worker (done/next, previous, postpone, refuse). Implicit input is based on surveillance of the worker as described in Section II-B.2 (e.g. handover, grasp, laydown).

6) Output Layer: The output layer receives assistive information from the PCA and performs hardware-specific adaptation and preparation for the actual presentation through the augmented workbench.

7) Augmented Workbench: The augmented workbench presents information through a projection system. It is equipped with controls for explicit worker feedback and sensors for implicit worker feedback.

B. Technical realization

1) Process Model: Several authors have modeled the entirety of possible assembly sequences for a single product in graph-based structures. These approaches have in common, that the main intention lies in reducing the combinatorial entirety of assembly sequences to the ones deemed as feasible. An assembly task is said to be geometrically feasible, if there is a collision-free path to bring the two subassemblies into contact from a situation in which they are far apart. And an assembly task is said to be mechanically feasible if it is practicable to establish the attachments that act on the contacts between the two subassemblies that correspond to a state (not necessarily stable). All operations of an assembly sequence have to fulfill both conditions strictly. However, none of the existing representations allow for the selection of assembly tasks in real-time. This hinders the delivery of situationally adapted instruction to the worker. A key to solving these issue is seen in the environmentally-dependent and situation-dependently triggered paths on state-based graphs. A mapping of the products processing states and a dynamic determination of the otherwise sequential tasks is achieved by the graph-based structure shown in Figure 2. The product-specific graphs are deducted from construction and assembly-related information (i.e. precedence relations). In accordance with Figure 2, the source vertex represents the initial state of the product to be assembled (state z_{start}) and the sink vertex represents the target state of the fully assembled product (state z_{target}). The edges of the graph (e.g. $\text{A}_{\text{start,I}}$) symbolize assembly task instructions. The execution of such by the worker will transfer the work piece in focus from one state to another (e.g. z_{start} to z_{I}). The path from the initial state z_{start} to the target state z_{target} leads across the respective major intermediate states z_{start} to z_{target} (e.g. z_{II}). The minor intermediate states shown in Figure 2 ($z_{\text{I},1,0}$, $z_{\text{I},1,1}$, ...) are reachable via an increased degree of detail of the assembly task instructions. If a vertex has more than one outgoing edge $\text{A}_{\text{i,j}}$, then alternative assembly sequences or alternative parts can be disposed of. The graph will not allow for cycles, under the presupposition that a disassembly of products to be built shall not be possible. This simplification entails, that the target state of the fully assembled product (state z_{target}) is reachable on a path from every state of the graph. Therefore, there are no states which would hinder a further assembly towards the target state. A future extension of the presented concept will account for disassembly tasks and is in the scope of the current research activities.

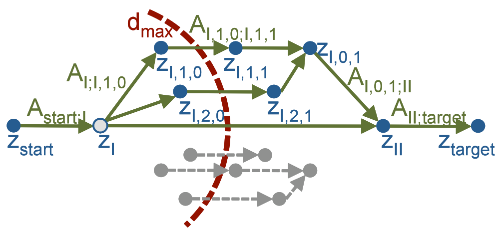

Fig. 2. Schematics of a state-based assembly graph

As the above knowledge representation is mainly derived from the precedence graph relations, it may not contain all possible assembly steps. This posed the necessity for a methodology for the derivation of assembly primitives to transform non-defined states into known ones. Based on a camera system (to be set up), the assembly scene is recognized and analyzed (see Section II-B.2). According to an appropriate distance metric, the subsequent selection of a reachable known state and the derivation of the (dis-)assembly tasks is executed.

Within the process of deriving the work step to display next, a (quasi-) standardized means of representing manual tasks is lacking. Existing methods can be distinguished into analytical and representational. The former proved to be neither applicable nor feasible in this scenario. Current representational methods however do not share a common structure and scope of tasks (in neither vertical or horizontal aspect). Regarding these shortcomings, a common, generally accepted representation of work steps based on the VDI-Richtlinie 2860 (assembly and handling) and relevant chapters of the DIN 8593 (manufacturing processes) was developed (see Figure 3). The representation is based on an XML-structure.

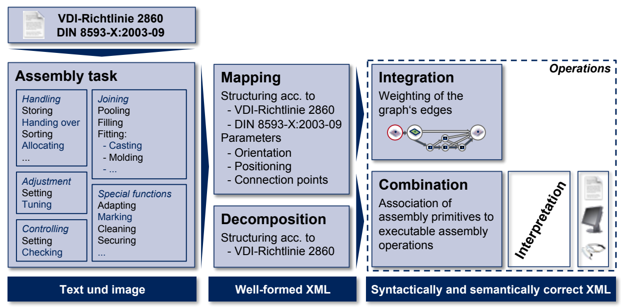

Fig. 3. Derivation of assembly primitives

2) Inputs/Outputs: The communication between worker and system has to be realized. Therefore, implementations of in- and outputs have been performed.

a) Explicit Inputs: Direct interaction with systems can be performed with standardized interfaces, e.g. with buttons or keyboards (HIDs). Those interfaces can be used e.g. to browse through the instruction history of the assistive system. Another possibility would be to choose system options. In the system architecture (Figure 1), these are called *explicit inputs*, because the worker feeds her/his input directly into the system.

b) Implicit inputs: In contrast to explicit inputs, implicit inputs are understood as context-dependent interactions between the worker and the system. These are used to track events in the background based on observations from the workspace during an assembly process.

To be able to monitor human actions, the workbench features vision-based sensors. Observations generate events belonging to the actual work-piece status. This is done with a global top-down-view camera mounted above the workbench. With that device, it is possible to watch the actions on the workbench and locate objects on the surface.

Needed parts are currently stored in boxes. A vision-based box-detector has been implemented to detect the locations of these storage boxes. The boxes used vary in color and size. To detect their positions in pixel coordinates, a color-based image segmentation is performed. Relevant areas in the image-plain are extracted from background information using thresholding filters in the HSV-color-space. A classification is performed to determine whether an area is a box or not. The center of gravity for each box is then calculated and stored in an scene representation database. This modality allows a free distribution of the storage boxes on the workbench and the worker can choose where he wants to place them.

To decide which part has been taken out of which box, the location of the worker's hand has to be known as well. This is done by automated restriction of the search region in the image followed by detection of human skin. Assuming that the worker's hands are moving while he is performing his task, only those areas are of interest, where changes in the image occurred during an analysis-period. A motion-detector has been implemented for cropping regions containing movements. The next step is to find human skin in the image plane. Therefore, a skin color filter operation is applied to those regions with detected movements. After classifying regions with hands (*hand-blobs*), the actual position of the hand is approximated by the center of gravity of these hand-blobs. The combination of motion-detection and skin-color filtering improved the recognition results of the actual hand position.

The previously shown processing steps are used to detect objects and generate raw data – e.g. position data for the worker's hands. For creating *implicit inputs*, these data are now fused to get a *hand-over-box-event*. This event is the basis for detecting, if the worker has grasped a part. Under the assumption that the worker will move his hands during the process mainly above the center of the workbench – the area with no boxes on it – one can conclude, that if the hand rests over a box, a part will probably have been taken out. Thus, the system triggers the *grasp-event*.

c) Table Projection Unit: With a table projection unit, instructions can be displayed onto the whole workspace. Using a calibrated surface of the workbench in relation to the camera position, the free space for displaying those instructions is known to the system. The top-down-view camera delivers the information about areas which are already taken and blocked with assembly parts. This feature enables the system to dynamically assign the position of display areas to the instruction system.

C. Experiments

The generic data base of human cognition and behaviour within the system architecture includes relevant information for adapting instruction presentation to the worker. Controlled experiments for feeding this database have been conducted.

1) Experimental setup: The experimental setup, consisting of a standard workbench equipped with a Polhemus motion tracker, beamer, monitor and DV camera, was supplemeted by an remote eye tracker and by two foot pedals as an user interface. The eye tracker can be used for the investigation of eye movement behaviour giving insights to e.g. search strategies and dwell times on instruction details. The eyetracker camera was placed under the plate of the work bench and rotated for pointing upwards. This enables recording the eye positions of a worker looking down to the working area.

2) Results: So far, results revealed parameters for the prediction of task performance. Moreover, different modes for instruction presentation and the value of certain user interfaces have been analysed and evaluated.

a) Task Parameters: The effect of task complexity as a manipulator of working performance was demonstrated in the dependent time for one work step times within an assembly task. Certain assembly primitives (i.e. three-dimesional fitting) lead to longer mean step times than others (i.e. two-dimensional fitting) with the same amount of parts, in accordance with difficulty ratings of the subjects. Moreover, step times varied within one assembly primitive as a linear function of the amount of parts. Therefore, the observed mean assembly times with a certain step classes can be used together with the amount of parts for the prediction of completion times and mental workload during ongoing assembly.

b) Presentation Modes: Results demonstrate an advantage of contact analog instruction presentation. With complex tasks (i.e. three-dimensional fitting), the assembly times per part was about 10 seconds shorter with contact analog instruction mode in comparison to monitor presentations. Motion tracker data deliver more detailed information on the subprocesses involved and enable to segment the work process in meaningful subunits (see Figure 4).

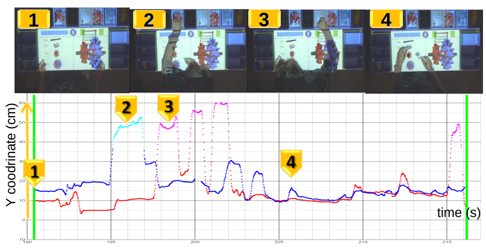

Fig. 4. Trajectories of hand movements and respective bird-view pictures in a manual assembly task.

The onset latency of the first movement to a box was shorter with the contact analog mode. Moreover, peak velocity and peak acceleration were higher under this condition, possibly because participants were more confident concerning the relevant box position. Therefore, it seems that the highlighting of boxes enabled to select the relevant part position and to shift attention faster resulting in an overall decreased movement onset time and higher movement velocities.

c) Interfaces: An analysis of pedal presses, motion tracker data together with video records shows that the foot pedals can be used as a user interface without the disturbance of the ongoing assembly process. Moreover, the interface conveys parallel working processes like e.g. quick checking of correct part assembly in previous substeps during the assembly procedure. It enables the observation of the exact point in time when new information is needed and of working strategies.

3) Further Paradigms: The 'contextual cueing' pradigm from fundamental psychology, which is used in the cooperation project #148 (see section coperations), was applied to the working scenario in order to investigate the localization of relevant objects in a specific task configurations. Implicit learning mechanism depending on context configurations and search strategies can be inferred with this method.

D. Cooperation between projects

Cooperations between the projects ACIPE (#159), JAHIR (#328) and CogMaSh (#339) have been established to implement a common interface for fulfilling the challenges of the Cognitive Factory. Software has been developed to gain access to the control of the common conveyor belt. The existing approaches for the product-based assembly assistance will be further enhanced and integrated to the complete demonstration platform of the Cognitive Factory (including the CoTeSys projects CogMaSh and JAHIR). This comprises the further development of the concept for the standardized product-resource communication, the adequate integration of the product in the assembly process from the data processing point of view (product state-based control) and the development of algorithms for the validation of the assembly feasibility of a product. Hence, the basics for the integrity of the approach for the adaptive assistance of throughout the assembly control system are achieved. The scheduling methods of CogMaSh include the abilities of the different workstations and the distribution of workloads between the different manufacturing areas. This calls for a tight integration of the ACIPE processes.

For being able to use shared-memory access to multiple cameras and sensor modules, the real-time database (RTDB) has been tested an verified in cooperation with project JAHIR. The RTDB also can be used as a short-term memory buffer – a sensor buffer – in future real-time applications.

An implementation of the display techniques with a table projection unit has been shown at the trade fair AUTOMATICA 2008 in Munich. A human-robot assembly station was equipped with a basic assembly guidance system to demonstrate the principle procedure of implicit and explicit interactions with great success.

In close cooperation with project(#148) methods for the investigation of context effects on search performance have been applied to the working scenario.

III. Next steps

The next steps are concerned with further integration with the Cognitive Factory, the implementation and extension of concepts developed so far, and the continuation of subject experiments to fill the *database of human cognition and behaviour*.

Within ACIPE a formal representation for assembly instructions has been developed, which will be synchronized with CogMaSh's knowledge representation, and opened to all Cognitive Factory projects. Based on information flow analysis the ACIPE work place will be integrated into the overall factory work processes and demonstrated by example of the common product scenario. Therefore, the components and concepts developed so far will be integrated, leading to a running system suitable for demonstration.

ACIPE specific scenarios currently used in subject experiments will be refined to demonstrate the capabilities of the ACIPE work place beyond the common product scenario. Learning algorithms will be integrated to increase worker-adaptiveness. The augmented workbench will be extended to increase the variety of explicit and implicit input from the worker leading to a higher level of control of the system in addition to the support provided through it.

Experiments conducted in cooperation with project(#148) will be continued. Here, the influence of objects identity as property of context configurations will be analyzed in order to determine which factors constitute the search of task relevant objects. Object dimensions will be varied for analyzing, which object properties are most important for attentional selection within the working scanario. The eye tracker data enable to infer attentional mechanism and search behaviour within this paradigm. Further experiments will combine different measurement methods like motion and eye tracking within the same experiments. This will give more detailed insights in the interplay of related subprocess like attentional selection and action execution. On the long run, active EEG electrodes will supplement the experimetal setup for the recording of event related brain potentials.

IV. Conclusion

Interactive multimodal guidance and ergonomic support of the worker in manual and semi-automatic assembly is achieved by a framework comprised of the respective models for work plan, scene representation and human cognition. Appropriate instructions in manual assembly scenarios are derived from known product states and unknown product states are coped with. Furthermore this includes the interpretation of events by classification and data-fusion of multi-modal observations. In this course, specific cognitive processes and their involvement in single task aspects are recognized.

So far, experiments for analyzing human cognitive processes regarding manual assembly have been conducted and eye and hand movements were recorded. Basic mental models have been derived and algorithms for state-based mapping of the product's processing states and a dynamic determination of otherwise sequential steps were developed. Further integration of these will provide an adaptive assistance system for assembly that allows for naturalistic interaction.

REFERENCES

[1] A. Bannat, J. Gast, G. Rigoll, and F. Wallhoff, "Event Analysis and Interpretation of Human Activity for Augmented Reality-based Assistant Systems," in *IEEE Proceeding ICCP 2008*, Cluj-Napoca, Romania, August 28-30 2008.

[2] A. Bannat, F. Wallhoff, G. Rigoll, F. Friesdorf, H. Bubb, S. Stork, H. J. Müller, A. Schubö, M. Wiesbeck, and M. F. Zäh, "Towards Optimal Worker Assistance: A Framework for Adaptive Selection and Presentation of Assembly Instructions," in *1st International Workshop on Cognition for Technical Systems*. Technische Universität München, Munich, Germany, October 2008.

[3] F. Friesdorf, "Complex Systems' Workflowmanagement," in *2nd International Conference on Applied Human Factors and Ergonomics*. Las Vegas, Nevada, USA: USA Publishing, 14-17 July 2008.

[4] ——, "Methodik zur systemergonomischen Entwicklung kognitiver Assistenz für komplexe Arbeitssysteme," in *54. Kongress der Gesellschaft für Arbeitswissenschaft: Produkt- und Produktions-Ergonomie - Aufgabe für Entwickler und Planer*, April 2008.

[5] F. Friesdorf, M. Plavšić, and H. Bubb, "An Integral System-Ergonomic Approach for IT-based Process Management in Complex Work Environments by Example of Manufacturing and Health Care," *Human Factors and Ergonomics in Manufacturing*, 2008.

[6] C. Stoessel, M. Wiesbeck, S. Stork, M. F. Zaeh, and A. Schuboe, "Towards Optimal Worker Assistance: Investigating Cognitive Processes in Manual Assembly," in *Proceedings of the 41st CIRP Conference on Manufacturing Systems*, 2008.

[7] S. Stork, C. Stößel, H. J. Müller, M. Wiesbeck, M. F. Zäh, and A. Schubö, "A Neuroergonomic Approach for the Investigation of Cognitive Processes in Interactive Assembly Environments," in *Proc. 16th IEEE International Symposium on Robot and Human interactive Communication RO-MAN 2007*, 26–29 Aug. 2007, pp. 750–755.

[8] S. Stork, C. Stößel, and A. Schubö, "The Influence of Instruction Mode on Reaching Movements during Manual Assembly," submitted.

[9] ——, "Optimizing Human-Machine Interaction in Manual Assembly," in *Proceedings of 17th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)*, 2008.

[10] F. Wallhoff, "Entwicklung und Evaluierung neuartiger Verfahren zur automatischen Gesichtsdetektion, Identifikation und Emotionserkennung," Ph.D. dissertation, Technische Universität München, 2006.

[11] F. Wallhoff, M. Ablaßmeier, A. Bannat, S. Buchta, A. Rauschert, G. Rigoll, and M. Wiesbeck, "Adaptive Human-Machine Interfaces in Cognitive Production Environments," in *IEEE Proceeding ICME 2007*, Beijing, China, July 2-5 2007, pp. 2246–2249.

[12] F. Wallhoff, M. Ruß, G. Rigoll, J. Göbel, and H. Diehl, "Improved Image Segmentation Using Photonic Mixer Devices," in *IEEE Proceeding ICIP 2007*, vol. VI, San Antonio, Texas, USA, 16.-19.09. 2007, pp. 53–56.

[13] M. Wiesbeck and M. F. Zaeh, "A Model for Adaptively Generating Assembly Instructions Using State-based Graphs," in *The 41st CIRP Conference on Manufacturing Systems*, M. Mitsuishi, K. Ueda, and F. Kimura, Eds. Tokyo, Japan: Springer, 2008, pp. 195–198.

[14] M. F. Zaeh, C. Lau, M. Wiesbeck, M. Ostgathe, and W. Vogl, "Towards the Cognitive Factory," in *Proceedings of the 2nd International Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV)*, Toronto, Canada, July 2007.

[15] M. F. Zäh, M. Wiesbeck, F. Engstler, F. Friesdorf, A. Schubö, S. Stork, A. Bannat, and F. Wallhoff, "Kognitive Assistenzsysteme in der Manuellen Montage," in *wt Werkstattstechnik online*, vol. 97, 9. Springer-VDI-Verlag, 2007, pp. 644–650.

[16] M. F. Zäh, M. Wiesbeck, H. Rudolf, and W. Vogl, "Virtual and Augmented Reality," in *Proceedings of Virtual Concept*, 2006.

[17] M. F. Zäh, M. Beetz, K. Shea, G. Reinhart, K. Bender, C. Lau, M. Ostgathe, W. Vogl, M. Wiesbeck, M. Engelhard, C. Ertelt, T. Ruehr, M. Friedrich, and S. Herle, "The cognitive factory," in *Changeable and Reconfigurable Manufacturing Systems*, H. A. ElMaraghy, Ed. Berlin: Springer, 2008, p. in publication.

[18] M. F. Zäh and M. Wiesbeck, "Ein modell zur zustandsbasierten erzeugung von montageanweisungen," in *Bericht zum 54. Arbeitswissenschaftlichen Kongress vom 9. - 11. 4. 2008 an der Technischen Universität München*. Dortmund: GfA-Press, 2008, pp. 219–222.

[19] C. Vesper, S. Stork, M. Wiesbeck, and A. Schubö, "Intra- and interpersonal coordination of goal-oriented movements in a working scenario," in *In: 1st International Conference on Cognitive Neurodynamics (ICCN'07)*, 2007.