Joint Action for Humans and Industrial Robots

Progress Report

Prof. A. Knoll Informatics VI: Robotics and Embedded Systems
Prof. G. Rigoll, Dr. F. Wallhoff Institute for Human-Machine Communication
Prof. M. Zäh Institute for Machine Tools and Industrial Management
Technische Universität München

***Abstract*—This progress report summarises the activities and achievements that have been carried out during the first funding period of the CoTeSys project *Joint Action for Humans and Industrial Robots - JAHIR*. As stated in the first project proposal the major goals within *JAHIR* are to investigate the cognitive basis for true joint human-robot interaction, and to use this knowledge to establish a demonstrator platform within the *cognitive factory* which can be used by other projects. This document therefore contains a succinct overview of the projected application scenario together with the layout of the assembly cell. A general idea of the software architecture is presented and used input devices are listed. Finally current cross activities within the cluster and other project relevant activities are described.**

I. Introduction

The project *Joint Action for Humans and Industrial Robots* aims to integrate industrial robots in a human-dominated working area, so that humans and robots can cooperate on a true peer level. *JAHIR* focuses on the information gained from the intuitive non-verbal communication channels based on advanced computer vision algorithms due to the noisy environment in the *cognitive factory*, which makes speech recognition unreliable therefore rules speech out as input modality in the current project phase. Unlike in fixed set-ups, the environmental conditions in production areas are heavily unconstrained, with crossing workers, moving machines, and changing lighting conditions. This requires stable preprocessing of the visual input signals which selects only objects or image apertures which are "interesting" and reasonable for joint action. A solution is to detect these regions of interest using techniques including color segmentation, motion estimation, background-subtraction or the use of a human-inspired visual attention system.

The integration of various sensors, such as cameras and force/torque sensors is needed for interaction scenarios such as the handing over of various parts during the assembly. In our setting, the robot and the human worker share the same workspace. The robot needs to react to sensor input in real time in order to avoid obstacles, which means that robot movements cannot be programmed offline. Thus the suggested workspace sharing will require the use of more advanced solutions in the area of online motion planning. Additionally, an intelligent safety system is necessary to ensure the physical safety of the human worker, which should go beyond the state-of-the-art systems which slow down or stop the robot if the human worker comes close.

JAHIR builds up a demonstrator platform that other projects from RA A – E can use for their research. In return, the results of projects from other research areas can be used to improve the cognitive system, which is the main element in *JAHIR*. The *JAHIR* demonstrator is part of the *cognitive factory* [41], which is a demonstrator for cognitive systems in industrial settings in the German cluster of excellence *Cognition for Technical Systems (CoTeSys)* [7].

II. Use Case

The long-time vision of this project is that humans and industrial robots can work together in the same workspace using multiple input modalities to ensure the cooperation and the safety of the human worker. To reach this vision, we will engineer and implement step by step solutions that extend and improve the system continuously.

The first step is the hand-over of tools to the human worker including finding the tool on the desk, grasping it and passing it on. Addressing this task requires knowing the positions of the robot, the human's hand, and the relevant tool. The next steps include the use of collision avoidance methods and taking the attention of the human worker into account.

The project *JAHIR* can be a role model for future industrial production lines, that use the both the cognitive power and the sensibility of the human worker and the precision and physical power of an industrial robot supported by a cognitive architecture that can react to unforeseen events.

To show the leading-edge impact of the project issues, requires an adequate scenario. The two projects in the *cognitive factory JAHIR* (hybrid assembly) and *ACIPE* (manual assembly) decided to assembly slightly modified LEGO Mindstorms robots. The robot assembles the feet of the LEGO robot in a fast and precise way, and then passes them to the human worker who will mount them on the rest of the LEGO robot body with screws and install the cables. Because of the delicate operations, the body is assembled by the human worker with the support of the robot, which can place parts and hand over tools or without physical support (as in *ACIPE*). To integrate with the other project embedded in the *cogintive factory* (*CogMaSh, CoDeFS*), the feet can also be produced with the FSM System of Festo using the milling cutter and the conveyer belt to pass the parts to the hybrid and manual work places.

III. Assembly Systems

Single work pieces with a high number of variations are normally produced by a human worker, whose sensitive abilities permit a flexible and fast reaction to outside influences. Opposite an automatic assembly system can produce high quantities on a fixed schedule. Due to the rigid activities of the machine manufacturing plants, only small number of variations are possible. Since the production process is as far as possible automated, production methods with hazardous goods or in environments dangerous for humans can be realized. Hybrid assembly (Fig. 1) represents an intersection of manual assembly and automated manufacturing [30]. This unites the advantages of the two assembly concepts, which makes it possible to produce both unique pieces and mass-produced goods [28]. For *IAHIP* a robot cell for hybrid assembly was plan[illegible] ces

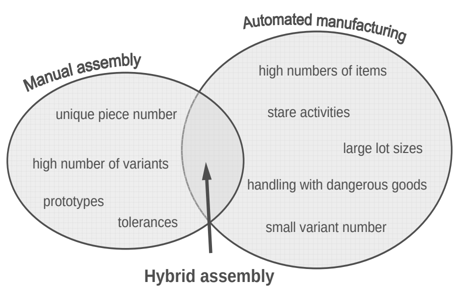

Fig. 1. Hybrid assembly as intersection of the manual assembly and the automated manufacturing

or prototypes, the hybrid mounting system acts as a handling assistant for example a hand-over of tools or the commission of work pieces. In mass production, the hybrid mounting system can take over repetitive and monotonous tasks. The worker then deals with sensory and fine-motor work, for example inserting work pieces.

A special form of hybrid assembly is direct human robot co-operation, in which the strict separation of robot and worker jobs is waived and thus manual and automatic capacities in a job are bundled. The co-operation cell is an autonomous unit for the assembly of building groups and complete products [5].

Direct human robot co-operation makes the flexible division of work between humans and robots possible, to use their specific abilities optimally. Since the robot deals only with the activities that can easily be automated, this leads to simplified process tools, simple robot programming and material allocation [33].

In direct human robot co-operation, the advantages of a human worker and an industrial robot can be combined (Fig. 2).

IV. Hardware Setup

A. Layout of the Joint Action Robot Cell

According to the nine valuation criteria presented in [17] (integration, worker in the safety area, accessibility, space

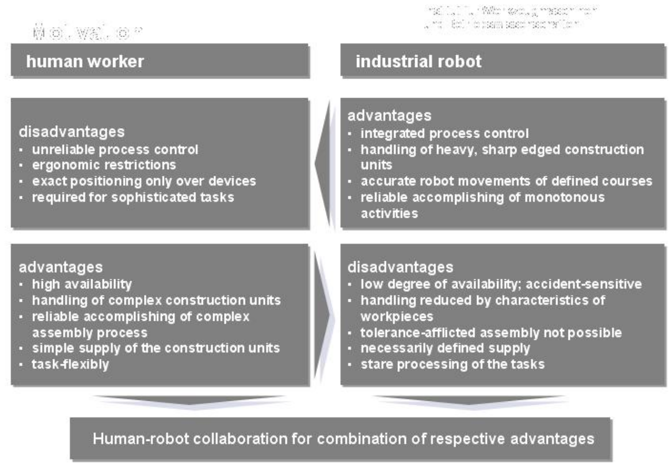

Fig. 2. Comparison human worker vs. industrial robot

requirement, space utilization, ergonomics, access to the conveyor, innovation human-robot cooperation, and flexibility) a total layout of the robot cell is generated. For the development of a fully planned layout some aspects had to be discussed with more details: An arrangement of robot, work bench and human was checked. A topology-research with a L-layout and an arrangement in which one robot is in opposite to another robot was made. In two other studies the staging of materials both to the robot and the worker and the access to the conveyor belt were checked.

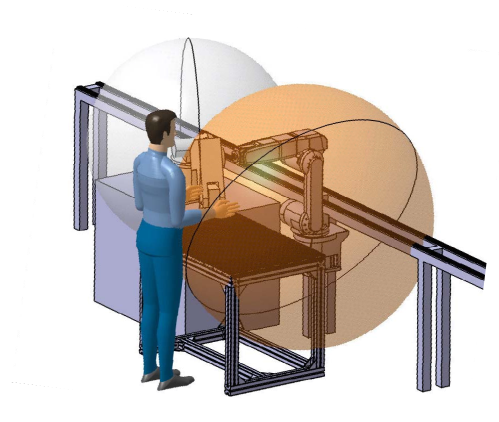

Fig. 3. 3D-scene of the joint action robot cell

Prof. Knoll initially provided two KUKA robots KR 6-2, but because of technical problems with the real-time interfaces (RSI-XML-connection between robot control and PC) we then decided to buy a KUKA KR 16. In a late-term planning phase, the KR 16 was replaced, for technical and financial reasons, by a Mitsubishi robot RV-6SL. This has a maximum carrying capacity of 6 kg and a range of 902 mm. The control of the KR 16 takes place in the interpolation clock (IPO) with 12 ms; the RV-6SL can be addressed with 7.1 ms.

The layout of the hybrid assembly cell was designed in a

3D simulation to ensure that the robot has full access to the conveyor belt and the whole desk. In the future a co-operation with the robot of the *CogMaSh* project is possible. Figure 3 shows the layout in the simulation and Figure 4 the real setup.

B. Construction of the Robot Cell

In July and August 2007 the robot cell was built up. A workbench was designed and built, the robot was ordered and was put into operation. Some sensors like the markerbased tracking system *smARTtrack* from the A.R.T.-Tracking GmbH [1] and a video camera based tracking system were checked. A pneumatic gripper was mounted to the robot flange.

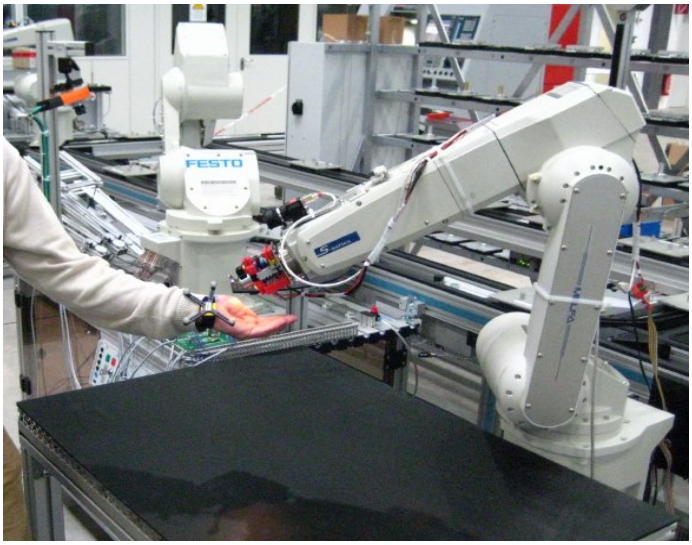

Fig. 4. Mitsubishi robot RV6SL with workbench

In order to guarantee the safety of the worker in the future, measures must be taken to ensure that no collision between worker and robot is possible(DIN EN ISO 10218). For this some sensors (e.g. safety laser scanner, adjustable footstep mats, light curtains, emergency stop tracers) must be integrated in the joint action robot cell.

V. Software Architecture

The high-level software architecture of *JAHIR* was defined according to research results of cognitive neuroscientists covering the aspects of successful joint action: joint attention, action observation, task sharing and action coordination [31]. In the following sections we will explain in detail the software modules and how the defined software architecture can cover these aspects.

Figure 5 shows the modules of the architectures and their connections to one another.

A. Joint Attention

For cooperation between robot and human, it is important that both partners coincide on the same objects and topics and create a perceptual common ground and shared representation [31]. That means the system has to be able to know where the human's focus of attention is. One method of detecting

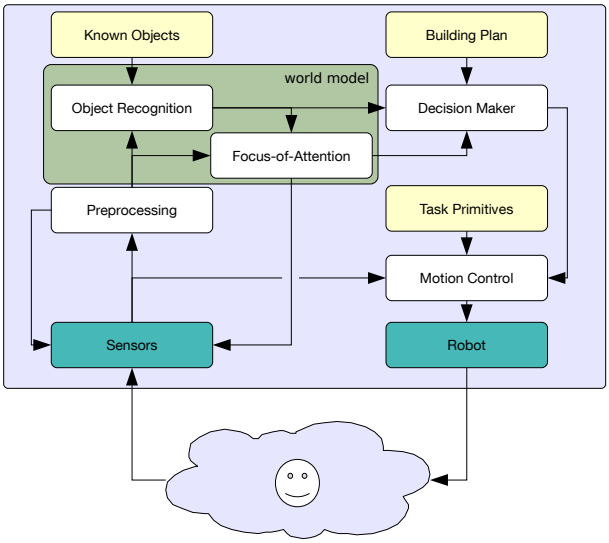

Fig. 5. High-level scheme of the cognitive system architecture of *JAHIR*. The yellow boxes are part of the knowledge database. The blue boxes are the connection to the *real world*. The white boxes are the software modules of the system.

attention is to recognize pointing gestures based on data gained by the visual sensor devices (cameras) using tracking techniques or from data gloves [26]. Therefore, we need to recognize a gesture as pointing gesture, compute where the human worker directs, and then transform this information to the focus of attention. Thus an internal system representation of the environment is needed that is always kept up to date with information from the sensors.

In addition to the pointing gestures the head orientation of the human worker will be used to infer his focus of attention. According to behavioral studies [9], the head orientation is directly connected to the human's focus of attention. Getting the information about the head orientation can be done e.g. with tracking techniques using mutual information presented in [22]. This algorithm will be integrated in the tracking library presented in Section VI. The information about the focus of attention can also be used in our cognitive system to search for needed objects in the focus of attention. In our system this is done in the *focus-of-attention* module. The *focus-of-attention* modules loops back to the sensors to directly control a camera mounted on a pan-tilt unit to follow the human's focus of attention.

Unlike in fixed set ups, the environmental conditions in production areas are heavily unconstrained, with crossing workers, moving machines, or changing lighting conditions. This requires stable *preprocessing* of the visual input signals, which selects only objects or image apertures that are interesting and reasonable for joint action and object recognition. A solution is to detect these regions of interest using techniques including color segmentation (e.g. skin colour), motion estimation (e.g. moving parts), background-subtraction or the use of a human-inspired visual attention system that creates a saliency map (e.g. the *Neuromorphic Vision Toolkit* [14] which is steadily

kept up to date [20]). The connection to the sensors allows to follow objects in a higher resolution using a camera mounted on a pan tilt unit.

Related work using attention has been done in [13], [8].

B. Task Sharing

To manage a predefined goal, i.e. the assembly of a product, the robot-system needs to know about the task in the production process as well as the human worker. Therefore the representation of the tasks should be as generic as possible to be able to change the role allocation even during the production. The knowledge of the building steps, plans, and skill primitives is represented by the *Building Plan* module which is depicted in Figure 5.

C. Action Observation and Coordination

All information perceived by the *sensors* builds up an internal representation of the working environment, the *world model*. In addition to the perceived data, the *world model* may carry information of the inventory (e.g. how much parts are stocked), the assembly line (e.g. which part is on the assembly line), and the worktable (e.g. information about areas where parts are stocked, or working areas where parts that are already in use can be placed).

These information is used in the decision making process (*Decision Maker*). To decide the next working step, we also need knowledge about the task from the *Building Plan* module. All these inputs to the *Decision Maker* lead to the next goal to solve, e.g. where to move next and what to grab. We call this kind of motion *long term motion*.

Defining the *Decision Maker* is one of the main challenges of *JAHIR*. Many conditions have to be considered, e.g. the next working step in the building plan (*Building Plan*), the needs of the human worker (another tool, part), and the human's intention (e.g. focus of attention, gestures, ...). We will review tools for task planning using *hierachical task networks* like O-Plan [32] and SHOP2 [19], or planners which make use of pruning rules like TALplanner [16] and TLPlan [3], or find other of new techniques for the selection of the next actions.

To control the movement of the robot (*Motion Control*), real time control of the robot is needed. This is especially required for the integration of various sensors, e.g. cameras and force/torque sensors for interaction scenarios such as the handing over of various parts during assembly. Because the robot and the human worker share the same work space, the movements of the robot cannot be programmed off-line. The robot needs to react in real time and plan the motion to avoid obstacles. Because the system has to plan motions online regarding safety aspects, we call this kind of motion *short term motion*. The system needs to keep the real task in mind while avoiding collisions. The kind of workspace sharing used in *JAHIR* will require the use of reactive collision avoidance [6] and more advanced solutions in the area of online motion planning [39].

D. Distributed Programming

In *JAHIR* we use a distributed software architecture with more than one computer, because modules (e.g. the tracking systems, the robot controller) need a lot of computational power, work on different operating systems and are implemented in different programming languages. To solve these problems, we use the plattform and language independent middleware *Internet Communications Engine (ICE)* [40] for distributed programming.

VI. Tracking and Sensing

As already mentioned in section V, it is important for a system to perceive its environment in order to realize *Joint-Action*. Especially for the hand-over in section II, the position of the human hand is needed to give or take objects. Further, it has to be considered that the worker just wants to grasp something on the table and doesn't want to interact with the system. These and other problems of intention recognition are being discussed throughout the cluster in *CA4 Nonverbal Communication*.

In the future an unintrusive remote visual tracking system should deliver the desired information about position and orientation of objects (see section VI-B) and the whole human (project MeMoMan, see [34]).

A. First input attempts

In order to get a demonstrator up and running, the following input modalities and devices have been evaluated for their usability in later experiments:

- **Cyber Glove**: This data glove provides 22 high-accuracy joint-angle measurements. It uses proprietary resistive bend-sensing technology to accurately transform hand and finger motions into real-time digital joint-angle data. It can be used with the Virtual Hand SDK to visualize a virtual hand and recognize gestures. In combination with Polhemus, delivering the global hand position, a user can non-verbally communicate and control the system even in noisy environments.
- **smARTtrack**: This is a stereo-camera system from the A.R.T.-Tracking GmbH [1]. These two cameras are synchronized and emit infrared flashes at a rate of 60 Hertz which are reflected by passive spherical markers. A target consits of several markers. If the geometry of the target is known, 6 DoF tracking is possible. This system can cover a workspace of up to 3 m^2, which is enough for the targeted handover scenario. To enlarge the tracked area and reduce possible occlusion of targets by obstacles, more cameras can be used. This system is currently used to obtain the global hand position (see Figure 6). Supplement research on gesture recognition within this project has been done in a similar environment at the institute for Human-Machine Communication [25].
- **PMD based image segmentation**: Aiming to improve image segmentation and extending regular computer vision algorithms, an image sensor is deployed to acquire additional depth information. By using the novel Photonic

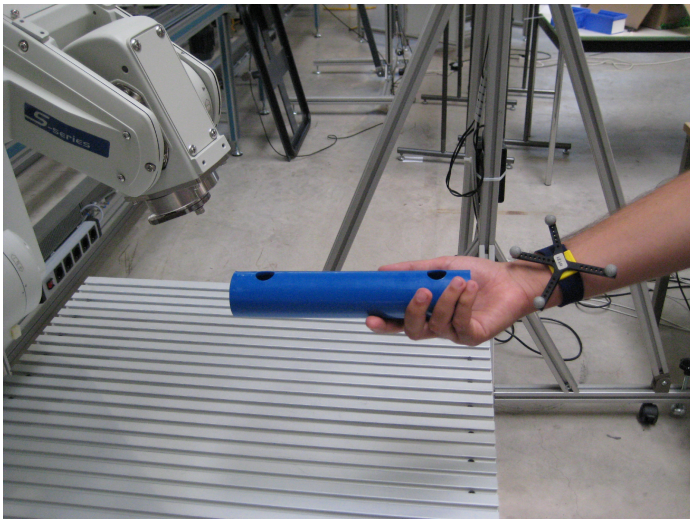

Fig. 6. Hand Tracking with markers

Mixer Device (PMD), it becomes possible to measure the observed object's distance to the camera. Motivated by expanding and making existing image processing and segmentation algorithms more robust, the additional depth information is overlapped with the image map gained by a regular camera system. The combination of these two inputs can be used to make applications like people counting, gesture recognition, and barrier detection more effective. A more detailed overview over the reseach activities is given in [36].

- **ARToolkit**: The ARToolKit [15] is a software library for building Augmented Reality applications. The Software is able to detect the size and rotation of a square marker with a black pattern on a white background in a camera image. Based on knowledge about the real size and camera calibration, the orientation and position of the marker in camera coordinates can be calculated. If the location of the camera in the world coordinate system is known, it can be used to compute the marker world-coordinates. Objects and fingers were successfully tracked using this Toolkit. However, the recognition of the patterns is very sensitive to lighting conditions and increasing the number of markers led to mismatching.

B. Structure of the new tracking library

As described in Section V and already mentioned in the section before, most of the input devices of the system depend on tracking algorithms. That is why one of the main research topics since May 2007 has been the definition of a general purpose software library with which one can handle complex visual tracking tasks easily.

The goal of this research is to have a *SDK* capable of solving many tracking problems with only few changes in the parameter settings and not by rewriting specific source code. For this reason a modular and multi-layered design of steps towards visual tracking applications was developed (see Figure 7). The structure of the tracking library is a result of the experience from previous work (e.g. [23], [21]). After the proper set up of the general structure, various sample applications will be used to extend and develop new functions, classes, specialisations, or generalisations.

For our cognitive system it is essential to know where objects are positioned in space in order to react according to the situation. Possible applications that can be used in *JAHIR* are person tracking, tracking of tools and parts, hand tracking, three dimensional face tracking and gaze tracking. Until the library is in a usable state, marker based tracking systems will be used.

The layers of the library is explained in short from top to bottom:

- **Layer 1 Math & Data**: The layer *Math & Data* contains in principle the most general parts of the library: the *Matrix*-class, contains definitions of matrix-related types such as vectors, square matrices, transformation matrices and mathematical operations such as matrix addition, subtraction, transpose and matrix-multiplication. The *Data* object is the *physical* representation of all kinds of data used during the tracking process. Examples for Data are: images (e.g. binary, RGB), feature points (e.g. calculated with the scale invariant feature transform approach (SIFT) [18]), colour histogram, edges or the object-state.
- **Layer 2 Util**: The second layer contains various utilities. It provides static functions allowing other classes to use *simple* and *basic* methods directly. The *Util* section is divided into *Geometry* (basic transformation methods), *ComputerVision* (basic cv methods), and *Drawing and Visibility test* (methods used for visibility testing and rendering).
- **Layer 3 Object**: The tracking library uses a 3D CAD model with textures for solving the tracking task. Out of this model *Shape & Appearance* features can be extracted and saved in the class *Visual features*. To be more robust, more than one feature type can be used for tracking. For the Bayesian tracking process, models for the object's motion (*prediction step*) and the likelihood computation (*correction step*) are needed.
- **Layer 4 Agents**: The classes in the agents layer control and use the classes from Layer 1 to 3 and save their results in the semantical data structures in Layer 5. Layers 4 and 5 together represent the overall flow of the tracking process. The sensor (e.g. a RGB-camera, a laser scanner) perceives and saves the unprocessed raw data into a data structure (e.g. an image) - the *SensorData*. After processing, this data (the measurement) is used for tracking (e.g. SIFT features of an image). The tracking either uses one of the available trackers (Kalman Filter or Particle Filter). The result of the tracking loop is the position, rotation, velocity and acceleration of the object (the state). This state can be drawn into an output image or used as input to another application (e.g. as input in an cognitive architecture to plan the next step).
- **Layer 5 TrackingData**: This layer represents the data used in the tracking process. The classes in this layer are the semantical representations of the physical data types

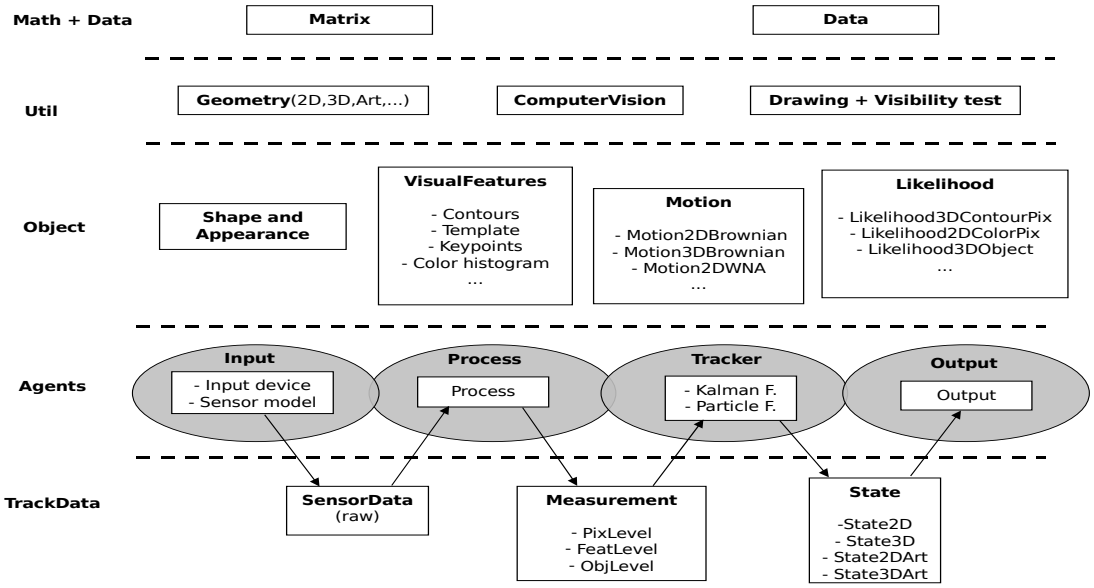

Fig. 7. Structure of the new general purpose library

Fig. 8. Tracking of a book cover

of the *Math & Data* layer.

C. Some Tracking Examples using the library

- The first sample application using the new tracking library was the tracking of a book cover using scale invariant feature points [18]. We use an implementation of the algorithm which applies the power of the Graphics Processing Unit (GPU) for the Gaussian pyramid construction, keypoint detection, feature list generation, orientation computation and descriptor generation (see [38]). The matching between reference features (*visual features*) and the features of the current frame is done on the CPU in an efficient way with a *KDTree* using a *Best-Bin-First* approach [4] and *RANSAC* [10] to discard outlying features. After the matching we know probable corresponding features of the current frame of the real scene and the reference frame (image of the book cover). Using least square estimation (*LSE*), the pose of the book cover can be estimated. This pose is the measurement which we use in an *Extended Kalman Filter* [37] for tracking. Because of the high computational needs of the SIFT algorithm, in a second step the tracking is done using the contracting curve density algorithm (*CCD*) presented in [12] and improved in [23]. SIFT was only used for initialising the object position. The main benefit of this approach is that the tracking takes place in real time. The output of the tracking is the pose of the book cover in 6 dimensions (3 translation, 3 rotation). Now that the book's position in the room is known, we can use this knowledge as input to solve other tasks in connected applications.

  Our aim is that other models of objects (e.g. parts, tools) can be loaded and used for tracking and 3D/6D pose estimation.
- The second application using the tracking library is person tracking using colour statistics. We build up a joint probability histogram from one or multiple reference image/s of a human worker which is used as visual feature for tracking. Using a particle filter, the human worker is tracked in real time and his position is precisely estimated. The system is now working in a stable way, and it will be used in the next funding period to know where the human worker is standing in the production cell. Figure 9 show the resulting estimation of the head position using this tracking attempt.
- Another application in the pipeline is detailed and high resolution face tracking using a pan tilt camera. The application performs planar face detection and controls the pan tilt camera to track the face of the user. We are working on extending this application to perform 3D face tracking with a pan tilt zoom camera which will provide not only the pose of the face but also a high resolution image.

Fig. 9. Tracking of the face using joint probability histograms

VII. Dissemination

A. Publications

Several scientific publications have resulted from the activities within the *JAHIR* project, which are listed below:

- Surveillance and Activity Recognition with Depth Information, ICME'07 [35]
- Robust Multi-Modal Group Action Recognition in Meetings from Disturbed Videos with the Asynchronous Hidden Markov Model, ICIP'07 [2]
- A Unifying Software Architecture for Model-based Visual Tracking, Technical Report 2007 (forthcoming) [24]
- Mutual Information-based 3D Object Tracking, accepted by International Journal of Computer Vision (IJCV) [22]
- Integrating Language, Vision and Action for Human Robot Dialog Systems, Proceedings of the International Conference on Human-Computer Interaction [29]
- Human-Robot dialogue for joint construction tasks, Proceedings of the 8th international conference on Multimodal interfaces (ICMI'06) [11]
- Improved Image Segmentation using Photonic Mixer Devices, ICIP'07 [36]
- Static and Dynamic Hand-Gesture Recognition for Augmented Reality Applications, HCI'07 [26]
- Robotergestützte kognitive Montagesysteme - Montagesysteme der Zukunft wt-online [28]

B. Cross activities

- **CA2 Cognitive Architectures**: This task force is focussing on the architectures needed for cognitive systems. As described in section V the architecture is a very important topic for *JAHIR* to enable easy integration of future sensors for recognizing the environment. Furthermore, modules for task knowledge, learning and decision-making are mandatory for every system with cognitive abilities.
- **CA4 Nonverbal Communication**: For the Joint-Action in *JAHIR* it is necessary for the system to recognize the intention of the human worker. Points that will be adressed in this workshop are e.g. gesture recognition, eye- and gaze tracking. An understanding of these topics can help to detect regions of interest for the system.
- **CA5 Knowledge and Learning**: This project was presented together with *ACIPE* at the Kick-off Meeting of CA5. Our focus lies on how to represent system knowledge of the perceived environment (obstacles, tools, worker, etc.) and the current working task as well as learning of new assembly plans.
- **CA7 Cognitive Factory**: *JAHIR* is embedded in the cognitive factory together with the projects ACIPE (#159), CogMaSh (#155) and CoDeFS (#161). The current activities and results of these projects will be shown and thereby interesting links for other groups of the cluster shall be identified.
- **CA8 Relation/Transfer between areas A and F**: To connect research areas A with the demonstrator platforms (research area F) *JAHIR* attended the workshop "Relation/Transfer between areas A and F". In the focus thereby the following questions could be located: Which requirements from the demonstrators are important for the basic research? What can the basic research contribute for the configuration of cognitive technical systems?

C. Other activities

- **Participation in CoTeSys Summer School PSSCR'07**: Throughout the cluster Player/Gazebo shall be used for simulation purposes. A *JAHIR* representative participated in the lectures and lab activities. The acquired knowledge will be used to simulate our setting in the cognitive factory. The summer school took place from 13 to 20 August in Garching.
- **Bayern Innovativ**: The project *JAHIR* was presented with a poster and a talk [27] in Augsburg at Bayern Innovativ "Intelligente Sensorik" (21 June) using this opportunity to get in touch with manufacturers of innovative safety sensors.
- **Visit at Transtechnik**: A guided tour to gather information about layout and organization of human dominated workplaces in the factory of Transtechnik was organized by *ACIPE*. Afterwards, the way of presenting complex assembly instructions to their worker was discussed. This trip took place on 25 July in Passau.
- **Visit of Haptic Day**: Several talks about haptic technologies for use in industrial development were given in Garching at the GATE (3 May). Followed by a demonstration session. Among other devices a finger tracking system with tactile feedback was shown. This infrared tracking system with active targets was used for gesture recognition.

D. Future Activities

Regarding the safety aspect, it is planned to establish knowledge transfer between *JAHIR* and the project "Functional Safety of Cognitive Systems" *SafeCoS* (# 200). Further on, innovative safety devices like the SafetyEye from Pilz and other soft sensors shall be integrated in the future. It is planned to get an overview about current use of robots in industry and discuss what assembly tasks are especially suitable for human robot cooperation. Therefore a visit at BMW is planned.

In the next funding period the team of *JAHIR* will implement and realize the chosen scenario described in Section II,

which will show the power of the project issues. Along with that, our new tracking library (described in Section VI-B) will be used for non-inversive tracking of the human worker, his hands and his head-orientation. This will lead to a natural and intuitive way of human robot cooperation using the fusion of multiple input modalities. A detailed description of our plans for the next funding period can be found in the CoTeSys proposal which will be submitted to the *EB* at the latest on 31 October 2007.

REFERENCES

[1] advanced realtime tracking GmbH. Infrared optical tracking systems. http://www.ar-tracking.de/.

[2] M. Al-Hames, C. Lenz, S. Reiter, J. Schenk, F. Wallhoff, and G. Rigoll. Robust multi-modal group action recognition in meetigs from disturbed videos with the asynchronous hidden markov model. In *Proceedings of the International Conference on Image Processing (ICIP)*, 2007.

[3] F. Bacchus and F. Kabanza. Using temporal logics to express search control knowledge for planning. In *Artificial Intelligence*, volume 116(1-2), page 123–191, 2000.

[4] J. Beis and D.G. Lowe. Shape indexing using approximate nearest-neighbour search in high-dimensional spaces. In *Conference on ComputerVision and PatternRecognition*, page 1000–1006, Puerto Rico, 1997.

[5] K. Beumelburg. Fähigkeitsorientierte montageablaufplanung in der direkten mensch-roboter-kooperation. 2005.

[6] O. Brock and O. Khatib. Elastic strips: A framework for motion generation in human environments. *International Journal of Robotics Research*, 21(12):1031–1052, December 2002.

[7] M. Buss, M. Beetz, and D. Wollherr. Cotesys - cognition for technical systems. In *Proceedings of the 4th COE Workshop on Human Adaptive Mechatronics (HAM)*, 2007.

[8] A. Edsinger. *Robot Manipulation in Human Environments*. PhD thesis, Massachusetts Institute of Technology, Department of Electrical Engineering and Computer Science, 2007.

[9] N. J. Emery. The eyes have it: the neuroethology, function and evolution of social gaze. pages 581–604, August 2000.

[10] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395, June 1981.

[11] Mary Ellen Foster, Tomas By, Markus Rickert, and Alois Knoll. Human-robot dialogue for joint construction tasks. In *ICMI '06: Proceedings of the 8th international conference on Multimodal interfaces*, pages 68–71, Banff, Alberta, November 2006.

[12] R. Hanek, T. Schmitt, S. Buck, and M. Beetz. Fast image-based object localization in natural scenes. In *Intelligent Robots and System, 2002. IEEE/RSJ International Conference on*, volume 1, pages 116–122vol.1, 30 Sept.-5 Oct. 2002.

[13] Gunther Heidemann, Robert Rae, Holger Bekel, Ingo Bax, and Helge Ritter. Integrating context-free and context-dependent attentional mechanisms for gestural object reference. *Mach. Vision Appl.*, 16(1):64–73, 2004.

[14] Laurent Itti, Christof Koch, and Ernst Niebur. A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11):1254–1259, 1998.

[15] H. Kato. Artoolkit. http://www.hitl.washington.edu/artoolkit/.

[16] J. Kvarnström and P. Doherty. Talplanner: A temporal logic based forward chaining planner. *Annals of Mathematics and Articial Intelligence*, 30:119–169, 2001.

[17] U. Lindemann. *Methodische Entwicklung technischer Produkte*. Springer, Berlin, 2. auflage edition, 2007.

[18] D.G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, (60):91–110, 2 2004.

[19] D. Nau, T.-C. Au, O. Ilghami, U. Kuter, J. W. Murdock, D. Wu, and F. Yaman. Shop2: An htn planning system. *Journal of Artificial Intelligence Research*, 20:379–404, December 2003.

[20] V. Navalpakkam and L. Itti. Modeling the influence of task on attention. *Vision Research.*, 45(2):205–231, 2005.

[21] G. Panin and A. Knoll. Fully automatic real-time 3d object tracking using active contour and appearance models. *Journal of Multimedia 2006*, 1(7):62–70, 2006.

[22] G. Panin and A. Knoll. Mutual information-based 3d object tracking. *International Journal of Computer Vision (IJCV)*, accepted, 2007.

[23] G. Panin, A. Ladikos, and A. Knoll. An efficient and robust real-time contour tracking system. page 44, 2006.

[24] G. Panin, C. Lenz, M. Wojtczyk, S. Nair, E. Roth, T. Friedlhuber, and A. Knoll. A unifying software architecture for model-based visual tracking (forthcoming). Technical report, Technical University of Munich, Department of Informatics, 2007.

[25] S. Reifinger, F. Wallhoff, M. Ablassmeier, T. Poitschke, and G. Rigoll. Static and dynamic hand-gesture recognition for augmented reality applications. In Julie A. Jacko, editor, *Human-Computer Interaction*, volume 4552 of *LNCS*, pages 728–737. Springer, 2007.

[26] Stefan Reifinger, Frank Wallhoff, Markus Ablaßmeier, Tony Poitschke, and Gerhard Rigoll. Static and dynamic hand-gesture recognition for augmented reality applications. In C. Stephanidis, editor, *Proceedings of the International Conference on Human-Computer Interaction*, Beijing, July 2007. Springer.

[27] G. Reinhart, W. Vogl, W. Rösel, F. Wallhoff, and C. Lenz. Jahir - joint action for humans and industrial robots. In *Intelligente Sensorik - Robotik und Automation*, Augsburg, 2007. Bayern Innovativ - Gesellschaft für Innovation und Wissenstransfer mbH.

[28] G. Reinhart, W. Vogl, W. Tekouo, W. Rösel, and M. Wiesbeck. Robotergestützte kognitive montagesysteme - montagesysteme der zukunft. *wt Werkstattstechnik online*, 97 (9), 2007.

[29] M. Rickert, M.E. Foster, M. Giuliani, T. By, G. Panin, and A. Knoll. Integrating language, vision and action for human robot dialog systems. In C. Stephanidis, editor, *Proceedings of the International Conference on Human-Computer Interaction*, pages 987—-995, Beijing, July 2007. Springer.

[30] R. D. Schraft. Die montage - das tor zum kunden. pages 2–11, Stuttgart, 2002. Fraunhofer-Institut für Produktionstechnik und Automatisierung (IPA).

[31] Natalie Sebanz, Harold Bekkering, and Günther Knoblich. Joint action: bodies and minds moving together. *Trends Cogn Sciences*, 10(2):70–76, Feb 2006.

[32] A. Tate, B. Drabble, and R. Kirby. O-plan2: An architecture for command, planning and control. 1994.

[33] S. Thiemermann. team@work - mensch-roboter-kooperation in der montage. pages 149–159, Stuttgart, 2003. Fraunhofer-Institut für Produktionstechnik und Automatisierung (IPA), FpF - Verein zur Förderung produktionstechnischer Forschung.

[34] Computer Science IX TUM and Institute of Ergonomics TUM. Memoman, methods for real-time accurate model-based measurement of human motion. http://vision.cs.tum.edu/projects/memoman/.

[35] F. Wallhoff, M. Russ, G. Rigoll, J. Göbel, and H. Diehl. Surveillance and activity recognition with depth information. In *Proceedings Intern. Conference on Multimedia and Expo (ICME)*, Beijing, 2007.

[36] Frank Wallhoff, Martin Russ, Gerhard Rigoll, Johann Goebel, and Hermann Diehl. Improved image segmentation using photonic mixer devices. In *Proceedings of the International Conference on Image Processing (ICIP)*, September 2007.

[37] G. Welch and G. Bishop. An introduction to the kalman filter. Technical Report 95-041, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA, 1995.

[38] C. Wu. Siftgpu. http://cs.unc.edu/~ccwu/.

[39] Y. Yang and O. Brock. Elastic roadmaps: Globally task-consistent motion for autonomous mobile manipulation. In *Proceedings of Robotics: Science and Systems*, August 2006.

[40] Zeroc. Internet communications engine. http://www.zeroc.com.

[41] M. Zäh, W. Vogl, C. Lau, M. Wiesbeck, and M. Ostgathe. Towards the cognitive factory. In *2nd International Conference on Changeable, Agile, Reconfigurable and Virtual Production (CARV)*, Toronto, 2007.